Journal of Theoretical Biology 431 (2017) 79–86

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

Statistical aspects of evolution under natural selection, with implications for the advantage of sexual reproduction

Daniel J.M. Crouch

Department of Oncology, University of Oxford, Oxford, OX3 7DQ, UK

ARTICLE INFO

Article history:
Received 27 August 2016
Revised 6 June 2017
Accepted 21 July 2017
Available online 2 August 2017

Keywords:
Sexual reproduction
Linkage disequilibrium
Hill–Robertson effect

ABSTRACT

The prevalence of sexual reproduction remains mysterious, as it poses clear evolutionary drawbacks compared to reproducing asexually. Several possible explanations exist, with one of the most likely being that finite population size causes linkage disequilibria to randomly generate and impede the progress of natural selection, and that these are eroded by recombination via sexual reproduction. Previous investigations have either analysed this phenomenon in detail for small numbers of loci, or performed population simulations for many loci. Here we present a quantitative genetic model for fitness, based on the Price Equation, in order to examine the theoretical consequences of randomly generated linkage disequilibria when there are many loci. In addition, most previous work has been concerned with the long-term consequences of deleterious linkage disequilibria for population fitness. The expected change in mean fitness between consecutive generations, a measure of short-term evolutionary success, is shown under random environmental influences to be related to the autocovariance in mean fitness between the generations, capturing the effects of stochastic forces such as genetic drift. Interaction between genetic drift and natural selection, due to randomly generated linkage disequilibria, is demonstrated to be one possible source of mean fitness autocovariance. This suggests a possible role for sexual reproduction in reducing the negative effects of genetic drift, thereby improving the short-term efficacy of natural selection.

1. Introduction

Sexual reproduction is by far the most prevalent mating system among the animals and plants, despite appearing to confer substantial disadvantages. For instance, when males contribute genetic material but no economic resources e.g. food or protection, individuals that clone themselves asexually ought to have 2^n as many descendants, after n generations, as those reproducing sexually. In the absence of any intrinsic benefit to reproducing sexually, a population of asexuals producing on average two offspring each will double in size per generation, while a sexually reproducing variety would stay at a constant size, as half of each female's resources are spent on male offspring that only reproduce by utilising female resources. This observation is referred to as the two-fold cost of sex (Maynard Smith, 1978). The effect is diminished when males invest economic resources in their offspring, but persists to some extent so long as they invest less than females, as is typically the case.

The preponderance of sexual reproduction is thus perplexing on the face of it, and various theories aspire to explain why it persists (Barton, 2010; Otto, 2009; Otto and Gerstein, 2006). Popular ideas are centred on the capacity of sexual reproduction to clear away the otherwise inexorable accumulation of deleterious mutations in finite-sized populations of asexual organisms (Muller, 1932), or similarly to combine favourable mutations within individuals more efficiently than under asexuality; referred to as the Fisher–Muller model (Fisher, 1930; Muller, 1932). When some degree of recombination is present, the Hill–Robertson effect (Hill and Robertson, 1966) operates via a related mechanism, whereby, under selection, loci that are linked are on average more susceptible to the effects genetic drift than unlinked loci, making recombination favourable. This is because linkage disequilibrium (LD) that is negative between beneficial alleles, though just as likely to occur through genetic drift as positive LD, persists for longer by reducing the pace of natural selection. It has been argued that the Fisher–Muller and Hill–Robertson arguments are fundamentally equivalent, as both are consequences of finite population size (Felsenstein, 1974). In an effectively infinitely sized population there is still a possible advantage to recombination, as selection can generate negative LD between beneficial loci so long as epistasis is common, weak and producing fitnesses lower than those expected based on the marginal effects of alleles (Barton, 1995; Kondrashov, 1988) (known as negative epistasis). Another theory is based on interactions between species, and emphasises the role of host-parasite dynamics. When parasites evolve to exploit particular

E-mail address: daniel.crouch@oncology.ox.ac.uk

http://dx.doi.org/10.1016/j.jtbi.2017.07.021
0022-5193/

genotypes, the capacity of the host population to rapidly bring together rare or unique combinations of previously disparate alleles, via sexual recombination, may benefit its overall health (Hamilton, 1980). Similar arguments apply to random fluctuations in the abiotic environment (Charlesworth, 1976).

Although these are all productive theories, none are yet universally accepted (Otto, 2009). Empirical investigations mostly suggest that the preponderance of negative epistasis required to drive selection for sex is unlikely (de Visser and Elena, 2007), though this has been recently challenged (Sohail et al., 2017). The parasite avoidance theory is persuasive, but requires relatively strict constraints on the nature of host-parasite interactions (Iles et al., 2003; Otto, 2009; Otto and Gerstein, 2006), for instance that the parasites must have very strong selective effects on their hosts. The Fisher–Muller and Hill–Robertson mechanisms, predicated on certain consequences of finite population sizes, enjoy considerable theoretical support (Barton and Otto, 2005; Iles et al., 2003; Keightley and Otto, 2006), but do rely on genetic drift being a significant force.

Here, a quantitative genetic model with an arbitrary number of loci is analysed in order to examine the interference of selection by randomly generated LD (a consequence of genetic drift), when acting upon standing genetic variation. This provides an understanding of how reduction of LD provides short-term gains in fitness to sexual populations. Previously, simulations of many loci have shown that recombination confers long-term advantages even for large populations, in which genetic drift is expected to be less powerful, provided that the number of loci under selection is sufficiently large (Iles et al., 2003). Detailed analysis of theoretical models has demonstrated that recombination is most effective at increasing mean fitness in small populations (Bodmer, 1970; Felsenstein, 1974), but has not considered cases where there are many loci under selection. It also remains unclear what precise mechanisms provide advantages to sex in the short term, i.e. the increase in mean fitness from one generation to the next, under this theory, as the majority of work focuses on long-term consequences of randomly generated LD.

The most widely accepted explanations for sex are centred on its capacity to break down deleterious LD (i.e. negative LD between beneficial alleles), increasing the additive genetic variance in fitness among individuals in the population. Fisher's Fundamental Theorem of Natural Selection (Fisher, 1930) states that this variance is proportional to the rate of increase in mean fitness due to natural selection. As part of this work, we derive an expanded version of the Fundamental Theorem. When there is genetic drift, the expected rate of increase in mean fitness is affected by randomly-induced correlations, more formally the autocovariance, between mean fitnesses of consecutive generations. The expanded theorem is then shown to encapsulate certain finite-population based advantages of sexual reproduction, as randomly generated LD induces autocovariance of mean fitness between generations by interfering systematically with natural selection. In contrast with the classic formulation of the Hill-Robertson effect, which describes the long-term detriment to fitness caused by the persistence of deleterious LD, this is shown to slow the rate of change of mean fitness in the short term, immediately after LD is generated.

2. The model

We assume that the fitness w_i for individual i is the linear combination of their genotype scores; $x_{ij} \in \{0, 1, 2\}$ for 0, 1 or 2 major alleles respectively (where j is a locus identifier and all loci are biallelic), and the *average effects* on fitness $\hat{\beta}_j$, plus an intercept $\hat{\beta}_0$, and a residual ϵ_i that captures variation in reproductive success that cannot be attributed to the genotypes:

$$w_i = \hat{\beta}_0 + \sum_{j=1}^{M} \hat{\beta}_j x_{ij} + \epsilon_i. \tag{1}$$

The hats on the average effects indicate that they are fitted least squares regression coefficients, treating all individuals in the population as the 'data'. There are M polymorphic loci in total. The breeding value for fitness, $\hat{g}_i = \hat{\beta}_0 + \sum_{j=1}^{M} \hat{\beta}_j x_{ij}$, represents its heritable component, and is the main subject of this analysis. The mean breeding value is written $\bar{g}$, and is equal to the mean fitness, $\bar{w} = \sum_i^N w_i/N$, where N is the number of individuals. This is because, due to properties of least squares regression, $\sum_{i=1}^{N} \epsilon_i = 0$. In a deterministic system, i.e. where all the variables in Eq. 1 are known, the change in mean breeding value between parental and offspring generations can be written using the Price Equation (Frank, 1998; Price, 1970; Robertson, 1966) as

$$\Delta\bar{g} = \frac{var_i[\hat{g}_i] + E_i[w_i \Delta\hat{g}_i]}{\bar{w}}, \tag{2}$$

where $\hat{g}_i$ represents the breeding value for a given individual and the subscript i on the variance and expectation operators ($var_i[]$ and $E_i[]$) indicates that these are taken with respect to the individuals within the population (i.e. summing the expression in parentheses over i and dividing by N). In the terminology of quantitative genetics, $var_i[\hat{g}_i]$ is the additive genetic variance in fitness. The mean difference in breeding value between a given parent and their offspring is represented by $\Delta\hat{g}_i$, and can be influenced both by environmental effects and by genetic differences between the parent and offspring that are due to sexual recombination. *De novo* mutations, though also a cause of parent-offspring differences, are sufficiently rare to be unimportant, and are thus excluded from our analysis for simplicity. Eq. 2 partitions the causes of breeding value evolution into two components: the first term represents the action of natural selection, and the second term the transition between parents and offspring (Frank, 2012). Extending to incorporate the stochastic effects of the environment, we can write:

$$E_e[\bar{w}\Delta\bar{g}] = E_e[var_i[\hat{g}_i]] + E_e[E_i[w_i \Delta\hat{g}_i]], \tag{3}$$

where the newly added expectations are taken with respect to the random environmental effects, represented by e. Here, random environmental factors are considered to be any that affect the distribution of genetic (x_{ij}) or non-genetic (ϵ_i) variables, or those that are affected by both e.g. the 'fitted' average effects $\hat{\beta}_j$. Genetic variables are considered random variables with respect to the environment, as random environmental forces can affect the distribution of genotypes in subsequent generations, e.g. through genetic drift. The subscript e therefore refers to quite general random phenomena and is intended mainly to differentiate from expectation, variance and covariance operations over i, which describe statistical relationships between the individuals within a given population. Eq. 3 is a formulation similar to that of Grafen (Grafen, 2000), who refer to the random environmental effects as "states of nature". In the following section, the model will be rearranged so as to partition the causes of breeding value evolution into separate statistical processes.

3. Results

Individual variance in fitness

First, we separately derive expressions for the two terms on the right hand side of Eq. 3 that permit their further analysis in terms of population genetic parameters. For a given environmental state initially, the first term, which is the between-individual variance in breeding values, can be expressed in terms of LD coefficients and