Theoretical Population Biology 114 (2017) 128–135

Contents lists available at ScienceDirect

Theoretical Population Biology

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

Sex versus asex: An analysis of the role of variance conversion☆

Andrew E.M. Lewis-Pye [a,*], Antonio Montalbán [b]

[a] Department of Mathematics, London School of Economics and Political Science, London, WC2A 2AE, UK
[b] Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA

ARTICLE INFO

Article history:
Received 30 November 2016
Available online 19 January 2017

Keywords:
Evolution
Recombination
Linkage disequilibrium
Epistasis
Variance
Sexual reproduction

ABSTRACT

The question as to why most complex organisms reproduce sexually remains a very active research area in evolutionary biology. Theories dating back to Weismann have suggested that the key may lie in the creation of increased variability in offspring, causing enhanced response to selection. Under appropriate conditions, selection is known to result in the generation of negative linkage disequilibrium, with the effect of recombination then being to increase genetic variance by reducing these negative associations between alleles. It has therefore been a matter of significant interest to understand precisely those conditions resulting in negative linkage disequilibrium, and to recognise also the conditions in which the corresponding increase in genetic variation will be advantageous. Here, we prove rigorous results for the multi-locus case, detailing the build up of negative linkage disequilibrium, and describing the long term effect on population fitness for models with and without bounds on fitness contributions from individual alleles. Under the assumption of large but finite bounds on fitness contributions from alleles, the non-linear nature of the effect of recombination on a population presents serious obstacles in finding the genetic composition of populations at equilibrium, and in establishing convergence to those equilibria. We describe techniques for analysing the long term behaviour of sexual and asexual populations for such models, and use these techniques to establish conditions resulting in higher fitnesses for sexually reproducing populations.

1. Introduction

Sexual propagation must certainly confer immense benefits on those populations undergoing it, given that sex involves substantial costs such as the breaking down of favourable gene combinations established by past selection. There are many hypotheses as to the form these advantages take, and they fall naturally into two groups (Felsenstein, 1974; Maynard-Smith, 1978; Kondrashov, 1993). On the one hand, a function of sexual reproduction and meiotic recombination may be in providing immediate and physiological benefits, such as allowing repair of double strand DNA damage (Bernstein and Bernstein, 1991; Michod, 1993). Such mechanisms alone, however, are unlikely to account for the continued prevalence of sexual reproduction (Barton and Charlesworth, 1998; Kondrashov, 1993; Maynard-Smith, 1988), and so, on the other hand, decades of research have seen evolutionary biologists looking to develop explicit theoretical models which explain the advantages of sex in terms of the interaction between variation and selection. Many of these models (Barton, 1995; Otto and Barton, 1997; Hill and Robertson, 1966) focus on ideas originally due to Morgan (1913), Fisher (1930) and Muller (1932) which stress the ability of recombination to place beneficial mutations together on the same chromosome. In a similar vein, one may consider the accumulation of deleterious mutations (Muller, 1964; Felsenstein, 1974). Since the effect of selection is dictated by levels of genetic variability in a population, one may also look more directly to understand the effect of recombination on genetic variance. The key observation here is that under appropriate conditions negative linkage disequilibria will build up, impeding the response of the population to directional selection (Mather, 1943; Felselstein, 1965).

The mechanisms by which negative linkage equilibrium may be created in the first place, may be classified as either deterministic or stochastic. A key finding for deterministic models (Barton, 1995) is that recombination may be favoured when weak negative epistasis (measured relative to the multiplicative contribution of individual gene fitnesses) exists between favourable alleles. There is also strong evidence that stochastic effects (Hill and Robertson, 1966; Barton and Otto, 2005) may be substantial in the realistic

☆ This paper has no principal author. The ordering is alphabetical. Both authors contributed equally to the construction of proofs and simulations.

* Corresponding author.
E-mail address: a.lewis7@lse.ac.uk (A.E.M. Lewis-Pye).

http://dx.doi.org/10.1016/j.tpb.2017.01.002
0040-5809/

setting of finite populations. The basic mechanism in this case may be seen as follows. In the rare event that particularly beneficial alleles at distinct loci combine in a single genome, selection acts quickly to achieve fixation for the coupled beneficial alleles, meaning that the associated positive disequilibrium disappears quickly. In the case of a strongly beneficial allele which initially appears on a genome with weaker alleles at other loci, however, selection is slowed down (when recombination is weak or non-existent), meaning that the negative disequilibrium persists for a much longer period of time. Any variance in disequilibrium thus ultimately leads to negative disequilibrium on average.

Here, we shall consider a deterministic setting in which sex is seen to robustly outperform asex across a broad spectrum of models, and in which the fitness contributions of genes which can be attained via mutation may be bounded or unbounded. We shall make certain simplifying assumptions. It will be convenient to carry out most of our analysis, for example, relative to models in which individual genes contribute *additively* to the fitness of the genome, and relative to this assumption of additive contributions from individual genes we shall assume zero epistasis. It should be noted that seen relative to models in which genes contribute to fitness multiplicatively, our model therefore assumes negative epistasis, and so may be expected to display benefits to recombination (e.g. Barton, 1995). We shall also assume that loci are unlinked, so that they either correspond to loci on distinct chromosomes (one may consider that we are choosing a 'representative' from each chromosome), or else lie at sufficient distances when they share a chromosome. As well as facilitating the mathematical analysis, these simplifications allow us to establish the most basic conditions under which certain mechanisms of variance conversion (described in detail in later sections) will operate with substantial effect. If a phenomenon is already observed in such a model, it is because no extra hypotheses are necessary to make it true — that a cause is already present within the few features of the simple model. Moreover, analysing our proofs, we can extract key ideas that surely carry over to more general models. An added benefit of working with these simplified models is also a dramatic reduction in the computational complexity of running large simulations. Even before providing mathematical proofs of our results, we are able to run simulations modelling populations with many more loci and more alleles than would otherwise be possible. Simulations for these vast fitness landscapes robustly show sexual populations achieving more rapid increases in mean fitness. Fig. 1 shows a small cross-section of the results of simulations for models with finite or infinite haploid populations and where fitness contributions from individual genes may be combined additively or multiplicatively (further examples are given in Figures 6–10 Appendix E). It is worth noting a fact first observed by Maynard-Smith (1968) and illustrated in (e) of Fig. 1, that in the multiplicative model with zero epistasis and infinite populations beginning in linkage equilibrium, the sexual and asexual populations remain identical. This holds because selection then preserves linkage equilibrium.

We then concentrate our mathematical analysis on the infinite populations additive model, since dealing with this case allows us to avoid some of the complexities inherent in the finite population models while illustrating basic principles which carry through to the finite population additive model. We are able to give a rigorous mathematical analysis of the manner in which, during the process of asexual propagation, a negative linkage disequilibrium will be created and maintained, meaning that an occurrence of recombination at any stage of the process will cause an immediate increase in fitness variance and a corresponding increase in the rate of growth in mean fitness. For contexts where there is a large but finite bound on allele fitnesses, it is not surprising that the long term behaviour differs qualitatively from the case where there is no a priori bound of the fitnesses of genes resulting from mutation. In this case, a standard application of the Perron–Frobenius Theorem suffices to establish that the asexual process converges to a fixed point of the corresponding dynamical system, but a deeper analysis is required in order to establish the mean fitness of the population at this fixed point and to relate this to the long term behaviour for sexual populations. We develop techniques which suffice to carry out such an analysis, and establish higher resulting mean fitnesses for sexual populations in these bounded models.

2. The model

We consider haploid populations with non-overlapping generations. In the absence of dominance between alleles at a single locus, our analysis could easily be extended to consider diploid populations. We describe here the additive infinite population variants of the model (other variants are described in Appendix D). We do not assume alleles come from a pre-existent pool, but consider a (form of random walk mutation) model in which alleles are created by mutation as time passes, possibly without any bound on attainable fitness. For certain aspects of the mathematical analysis, it will be convenient to be able to assume that gene fitnesses occur in a discrete range rather than taking any real value. We ensure this by assuming that gene fitnesses take integer values. Appropriate scaling means this entails essentially no loss in generality—in order to simulate a model in which fitnesses take values to d decimal places, we can simply multiply all fitness values by 10^d, apply the model as described here, and then finally divide by 10^d in order to correct fitness values at any stage of the process. One could consider a model in which fitnesses can take any real values, without substantial changes in the behaviour of the model. Most other features of the model, which we now describe in more detail, are essentially standard in the literature.

Each instance of the model is determined by three principal parameters: ℓ, D and μ. First, $\ell \in \mathbb{N}\ (>1)$ specifies the number of loci. With each individual specified by ℓ genes, in the absence of epistasis, we need only be concerned with the fitness contributions corresponding to those genes, and so each individual can be identified with a tuple $\boldsymbol{x} = (x_1, \ldots, x_\ell) \in \mathbb{Z}^\ell$. The *fitness* of $\boldsymbol{x}$ is $F(\boldsymbol{x}) = \sum_{i=1}^{\ell} x_i$. (For the multiplicative model, one would define $F(\boldsymbol{x}) = \prod_{i=1}^{\ell} x_i$ instead.) Second, the *domain* $D \subset \mathbb{Z}^\ell$ determines which individuals are allowed to exist. We will use three types of domains in this paper: The $\mathbb{N}$-*model* uses as domain $D = \mathbb{N}^\ell$, where $\mathbb{N} = \{1, 2, 3, \ldots.\}$; the $\mathbb{Z}$-*model* uses $D = \{\boldsymbol{x} \in \mathbb{Z}^\ell : F(\boldsymbol{x}) > 0\}$; and the *bounded-model* uses $D = \{1, \ldots, N\}^\ell$ for some upper bound $N \in \mathbb{N}$ on gene fitness contributions. In practice, there is almost no difference between the $\mathbb{N}$- and $\mathbb{Z}$-models, but there are situations when it is simpler to consider one or the other. Third, $\mu\colon \mathbb{Z} \to \mathbb{R}^{\geq 0}$, the *mutation probability function*, determines how mutation affects gene fitness contributions: $\mu(k)$ is the probability that the fitness contribution of a gene will increase by k. For the sake of simplicity, we assume this distribution to be identical for all loci. While there is no clear canonical choice for μ, the behaviour of the model is robust to changes in this parameter so long as negative mutations are more likely than positive ones, both being possible. This is because any such choice of μ will approximate a Gaussian distribution over multiple generations. The simplest mutation distributions one may consider are those taking non-zero values only on $\{-1, 0, 1\}$. Unless stated otherwise, it should be assumed that from now on mutations are of this form and that $\mu(0) > \mu(-1) > \mu(1)$ (giving a form of stepwise-mutation model Ohta and Kimura, 1973).

By a *population*, we mean a probability distribution $\boldsymbol{\phi}\colon \mathbb{Z}^\ell \to \mathbb{R}^{\geq 0}$, where $\boldsymbol{\phi}(\boldsymbol{x})$ is the proportion of individuals that have 'genotype' $\boldsymbol{x} \in \mathbb{Z}^\ell$. For a population $\boldsymbol{\phi}$, we shall also use $X = (X_1, \ldots, X_\ell)$, where the X_i's take values in $\mathbb{Z}$, to denote a random variable that picks an individual with gene fitness contributions