The Effect of Selection on Genealogies

N.H. Barton
A.M. Etheridge


Supported by the Austrian Federal Ministry of Education, Science and Culture
Available via http://www.esi.ac.at
The effect of selection on genealogies

N. H. Barton * and A.M. Etheridge †

*Institute of Cell, Animal and Population Biology
University of Edinburgh
EH9 3JT U.K.

† Department of Statistics
University of Oxford
U.K.
ABSTRACT

The coalescent process can only describe the effects of selection at linked loci if selection is so strong that genotype frequencies evolve deterministically. Here, we develop methods proposed by Kaplan, Darden and Hudson to find the effects of weak selection. We show that the overall effect is given by an extension to Price’s equation: the change in properties such as moments of coalescence times is equal to the covariance between those properties and the fitness of the sample of genes. The distribution of coalescence times differs substantially between allelic classes, even in the absence of selection. However, the average coalescence time between randomly chosen genes is insensitive to the current allele frequency, and is only affected significantly by purifying selection if deleterious mutations are common, and selection is strong (Ns > 3). Balancing selection increases mean coalescence times, but the effect only becomes large when mutation rates between allelic classes are low, and when selection is extremely strong. Our analysis supports previous simulations which show that selection has surprisingly little effect on genealogies. Moreover, small fluctuations in allele frequency due to random drift can greatly reduce any such effects. This will make it difficult to detect the action of selection from neutral variation alone.
INTRODUCTION

Kingman (1982) introduced the coalescent process as a simple description of the genealogical relationships amongst a set of neutral genes. Although the theory of the coalescent has developed largely independently, it is closely related to the classical concept of identity by descent (Nagylaki, 1989). The coalescent extends naturally to describe structured populations, in which genes may be found in different places, or embedded in different genetic backgrounds. The effects of selection can easily be included, provided that it is so strong relative to random drift that the frequencies of different genetic backgrounds can be approximated as changing deterministically (i.e. $Ns >> 1$; Kaplan et al., 1988; Hudson, 1990).

In some cases, assuming that the genetic or spatial structure of a population changes deterministically is a good approximation. For example, when a single favourable mutation arises and spreads, it carries with it any linked variants. The effects of such ‘selective sweeps’ on genetic variability can be accurately described by assuming that the new allele increases exponentially, even though it is subject to strong random fluctuations in the early generations, when it is present in few copies (Maynard Smith & Haigh, 1974; Kaplan et al., 1988; Barton 1998). However, the deterministic approximation plainly fails when selection is weak or absent. Consider the relationships between genes which can be of two allelic types. Even if these alleles are neutral, and so represent an arbitrary labelling of the genes, two genes of the same allelic type are likely to be substantially more closely related than are two genes of different type. Moreover, the average relationship between randomly chosen genes depends on the allele frequency, since an allele which happens to have increased by chance will cause a ‘selective sweep’ just as if it had increased by selection. Although relationships averaged over the distribution of allele frequencies, and over allelic classes, must be unaffected by the labelling of neutral alleles, relationships do depend on allelic class, and on allele frequency (e.g. Slatkin, 1996).

We usually do not know which alleles are selected, and so can only observe relationships amongst randomly chosen genes, in populations with random genotype frequencies. However, with selection, even the average relationships are distorted, and can only be calculated using the structured coalescent when selection is much stronger than drift. Weak selection $hNs ~ 1$ spread across many loci can have significant cumulative effects (McVean & Charlesworth, 2000). Even when $Ns$ is large, fluctuations may still be important. For example, Barton and Navarro (2002) showed that the effects of balancing selection at multiple loci are strongly affected by drift even when $Ns$ is extremely large, because each particular genetic background is present in few (if any) copies.

When selection and drift are of comparable strength, a purely coalescent-based approach becomes complicated. Neuhauser and Krone (1997) and Krone and Neuhauser (1997) have shown that certain kinds of selection can be represented by ‘ancestral graphs’, which are constructed by allowing branching as well as coalescence as one moves back in time, followed by a culling of potential ancestors to generate the genealogy. This method is computationally demanding, especially with strong selection, because of the proliferation of ancestral lineages. Slade (2000a, b, 2001) and Fearnhead (2001) have introduced modifications which make calculations feasible for stronger selection. Nevertheless, this method does not seem likely to lead to a deeper analytical understanding, which would extend to more general kinds of selection, and more complex genetics.

Kaplan et al. (1988) introduced a more direct approach to the problem, by following relationships between genes within and between allelic classes, conditional on the frequencies of those classes in the population. However, their approach has not been taken up. This may be partly because Kaplan et al. (1988) did not specify boundary conditions for their equations, so that they could not be solved using standard software; however, Darden et al. (1989) do provide an alternative numerical algorithm. Barton et al. (submitted) give a rigorous justification for Kaplan et al.’s (1988) diffusion equations, including the necessary boundary conditions. In this paper, we show that in the absence of selection, the overall average relationship between random genes in a random population is the same as under simple neutrality. This must of course be the case, since labelling a pair of neutral alleles cannot affect the distribution of genealogies. However, this result extends to give a general expression for the effect of selection on the genealogy, which can be seen as an extension of Price’s (1970) equation.
The analysis is of a neutral locus which is linked to a single selected locus that carries two alternative alleles. (Setting the recombination rate to zero allows us to follow genealogies at a single selected locus.) Essentially the same equations apply to probabilities of identity in state, assuming infinite-allele mutation at the neutral locus; the mean and higher moments of coalescence time; and the full distribution of coalescence times. The equivalence between these can be seen by noting that the identity in state is the generating function for the distribution of coalescence times. Our numerical results are mainly for the distribution of pairwise coalescence times, but in the last part of the paper, we consider the distribution of the total length of a large genealogy.

We begin by setting out the diffusion approximation for identities in allelic state; the equations for coalescence times are essentially the same. We then change variables to work with the average over randomly chosen pairs of genes; differences associated with one or other allele; and differences within versus between classes. This change of variables leads to a simple formula for the expectation over the stationary distribution, and to approximations for strong mixing between classes, and for strong selection. Throughout this first part of the paper, two simple examples are used to illustrate the derivations (two genes sampled from either a neutral locus, or a locus under balancing selection). In the later sections, a wider range of parameters is explored.

THE MODEL

Consider selection on a single locus which carries two alleles, labelled $P, Q$. This is linked to a second neutral locus, with recombination rate $r$. Allele frequencies at the selected locus are $p, q$ at the beginning of the generation. For simplicity, assume that selection acts on haploids; however, detailed assumptions about the life cycle do not affect the diffusion approximation. Numerical examples assume purifying selection with fitnesses of $Q, P$ of $1: 1 + s$; balancing selection is modelled by assuming frequency-dependence such that $s$ is replaced by $s(1 - p_0 - p)$. This is close to a model of overdominance with diploid fitnesses $1 - s p_0 : 1: 1 - s q_0$ (with $p_0 + q_0 = 1$). Mutation then occurs at a rate $np$ from $Q$ alleles to $P$ alleles, and $nq$ in the opposite direction. (In terms of the actual mutation rates $m = nQ; p + nQ; q, p = nQ, P, m$; the equilibrium under mutation alone is $p$.)

Where we consider identity in allelic state, mutation to novel alleles occurs at a rate $s$ at the linked neutral locus. Diploid zygotes are formed by random union, and undergo meiosis. Finally, $2N$ gametes are sampled to find the next generation. (We keep the convention that population size is $2N$ genomes, corresponding to $N$ diploid individuals). This discrete time model is defined in more detail by Barton et al. (2002), who also set out the analogous continuous time Moran model. Our notation is summarised in Table 1.

In the limit where selection, drift and mutation are weak, we can take a diffusion approximation to this model. We scale selection, mutation and recombination relative to $N$, and time relative to $2N$, so that $T = 2N$, $S = NS$, $R = Nr$, $U = Nm$, and $V = Nn$. (Note that Barton et al. (2002) scale time relative to $N$ generations). Suppose that we sample $n$ genes. We need to follow the probability $f_{jk}$ that $j$ genes of type $Q$, and $k$ genes of type $P$, are identical in state at the neutral locus. We assume this neutral locus to be subject to infinite-alleles mutation at rate $n$. $f_{jk}$ is also the generating function for the distribution of total length of the genealogy, and so can be used to find the distribution of the number of segregating sites.

Kaplan et al., 1988, Eq. 20, set out the diffusion equations for $f_{jk}$:

$$\frac{\partial f_{jk}}{\partial T} = -2V(j + k) f_{jk} + \left( \frac{j(j - 1)}{2} \frac{(f_{j-1,k} - f_{j,k})}{q} + \frac{k(k - 1)}{p} \frac{(f_{j,k-1} - f_{j,k})}{q} \right) + 2 \left( U \frac{p + R}{q} (f_{j-1,k+1} - f_{j,k}) + kq \left( U \frac{p + R}{p} \right) (f_{j+1,k-1} - f_{j,k}) \right) + 2 \left( U \frac{p + R}{p} \right) \frac{f_{j,k}}{pq} + \frac{p}{2} \frac{\partial^2 f_{jk}}{\partial p^2}$$

with $f_{0,1} = f_{0,1} = 1$ by convention. After a long time $T >> N$ or $T >> 1$, the identities will approach an equilibrium (i.e. $f_{j,k} \approx 0$). Barton et al. (2002) give a rigorous derivation for this stationary version, for $n = 2$. 


The terms involving $V = N_R$ represent the steady decay in identity due to mutation at the neutral locus. The positive terms $\frac{q_{j,k-1} - f_{j,k}}{p}$ represent the increase in identity due to coalescence within allelic classes. The terms involving differences in identity $Hf_{j-1,k+1} - f_{j,k}$ represent the movement of genes between allelic classes by mutation of allelic classes, and by recombination. Note that mutation from allele $Q$ to allele $P$ dominates recombination when allele $P$ is rare (term $U_p$ in Eq. 1), because most copies of $P$ will in that case have arisen as recent mutations from $Q$. Finally, the last two terms represent the proportion of populations currently at allele frequency $p$ which derive from populations with a different frequency; this process is approximated as a backwards diffusion.

In principle, the full distribution of genealogies can be recovered by assigning a notional mutation rate to each node, $\alpha_i$, $\nu_j$, $\tau_{jk}$, $\ldots$ say, and by following identities among sets of genes which are either in background $Q$ or $P$ (say). Then, terms such as $\frac{q_{j-1,k-1} - f_{j,k}}{p}$ in Eq. 1 separate out into distinct terms, corresponding to different permutations of loci over backgrounds. This gives a set of equations in a very large number of variables, and worse, an extremely large number of $f$’s: all possible partitions of the lineages must be tracked separately. So, numerical solution is difficult for even three genes. This approach might be useful, however, for deriving simpler equations which describe particular features of the distribution of genealogies.

As detailed in Barton et al. (2002), the probabilities of identity are the minimal positive solutions to the equilibrium version of Eqs. 1. This implicitly specifies the boundary conditions for the system, but in order to obtain numerical solutions, we require them explicitly. Consider first $p$ small. The right hand side of Eqs. 1 is dominated by terms in $\frac{1}{q}$. Solving for these terms leads to:

\[
\begin{align*}
    f_{j,k} &= \frac{Hk - 1L f_{j,k-1} + 4 N m p f_{j+1,k-1}}{Hk - 1L + 4 N m p} \quad \text{for } Hk > 0L \\
    f_{j,0} &= \frac{j Hj - 1L f_{j-1,0} + 4 N n p \sum_p f_{j,0}}{j Hj - 1L + 4 N j n}
\end{align*}
\]

Similarly, as $p$ tends to one, the terms in $\frac{1}{q}$ dominate and we obtain analogous boundary conditions to Eqs. 2.

**NUMERICAL METHODS**

Darden et al. (1989) described a numerical method for solving Eq. 1, and gave some results for the case of two genes. Because they did not specify boundary conditions, they could not use standard algorithms. Instead, they approximated the differential equations on a discrete grid, and thus obtained a set of linear equations which could be solved using matrix methods. This method is similar to solving the exact Wright-Fisher model for a finite population of 2N genes (see Fig. 1 below).

Since we have specified the boundary conditions (Eqs. 2), we can solve the equilibrium version of Eq. 1 using the built-in algorithms in Mathematica (Wolfram, 1996), at least for up to 5 or so genes. We proceed iteratively. Initially, all identities are set to zero. At the $n$’th stage of the iteration we solve the stationary version of Eq. 1 subject to the boundary conditions Eqs. 2 using a ‘shooting method’: the boundary conditions and the contribution from mutation and recombination involve identities that are provided by our trial solution. Thus, for example, to solve for $f_{0,2}^n$, we set the boundary condition $f_{0,2}^n(0)$ to the value given by Eqs. 2 with $f_{0,1}^{n-1}$ and choose two different (negative) values of $\sum_p f_{0,2}^{n-1}/p$, $a$ and $b$ say. This gives solutions $f_{0,2,1}$ and $f_{0,2,2}$. Write $z_{\Delta \lambda}$ for the corresponding values of $H + 4 V + f_{0,1}^{n-1} H + 2 U q \sum_p f_{0,2}^{n-1} L$. Since the correct choice of $\sum_p f_{0,2}^{n-1}/p$, $a$ say, must give $|z_{\Delta \lambda}| = 1$, and since we are dealing with a linear equation, the solution that we seek is given by setting $g = \frac{1 - z_{\Delta \lambda} \Delta \lambda}{z_{\Delta \lambda}} f_{0,2,1} + \frac{z_{\Delta \lambda} \Delta \lambda}{z_{\Delta \lambda}} f_{0,2,2}$. The iteration is repeated until the mean square change in the estimates is less than some small threshold.

For a sample of two genes, this procedure is successful after only a few iterations. For larger samples, the differential equations are less stable, and so it is necessary to integrate over a series of separate intervals. The algorithm we used is as follows. Starting at some small $e << 1$, integrate forwards as described above. By making
a small change in the initial condition, obtain a second solution which is close to the first for small $p$, but which may diverge as $p$ becomes large. Choose a point $p_1$ at which these two trial solutions are close to each other; we thus have an accurate solution for $e < p < p_1$. Repeat the procedure starting from the opposite boundary $\bar{p} = 1 - e$, obtaining a solution valid for $p_2 < p < 1 - e$. If $p_2 < p_1$, splice the two solutions together at the point where they differ least. Otherwise, repeat the procedure, working in from $p_1$ to $p_3$, from $p_2$ to $p_4$, and so on until the solutions starting from left and right overlap.

With more than 5 or so genes, the differential equations become extremely sensitive to the values near the boundaries, and so ’shooting methods’ fail. This instability arises because the terms due to coalescence grow quadratically with the number of genes, and so become extremely large relative to the diffusion terms which smooth the solution. This suggests an alternative approximation. We begin by calculating the identities amongst up to (say) 5 genes using the methods described above. To calculate identities amongst $n=6$ genes, we first discard the diffusion terms (i.e. the last two terms in Eq. 1). This is equivalent to assuming that allele frequency fluctuations are negligible over the short timescale set by coalescence among the $n$ lineages. This gives a set of linear equations: $0 = f_5 + B_f f_0$ where $B_f$ is a matrix, and $f_5$ is the vector of identities amongst sets of 5 genes, which we have already calculated. A correction $f_4$ to this solution $f_0$ can now be calculated by including the diffusion terms: $0 = B_f f_4 + \mathcal{L}_{pp} f_0$ where $\mathcal{L} = 2i/t + \mu - \mu L + S p q L + \frac{f_0}{2} q p$. By making repeated corrections in this way, we obtain an asymptotic series $f_0, f_1, f_2, \ldots$. In numerical calculations, the first few terms converge towards the correct solution (calculated as above), but further corrections lead to divergence. A satisfactory approximation can be found by taking just the first correction, $f_1$. We have checked that this is close to the solution calculated using the method described above, at least for small $n$. Once a solution for $n$ genes is found, the procedure can be repeated for $n+1$ genes, and soon. The method is fast, and allows calculations for up to a hundred or so genes. One further modification is required. Slight inaccuracies in solutions for a few 10-50 genes tend to accumulate as fine-scaled fluctuations in solutions for more genes. We therefore force smooth solutions by fitting an eighth-order polynomial solution for each $n$.

We begin by considering pairwise measures. Figure 1 gives a check on our numerical methods for two genes by comparing three methods for calculating identities within and between neutral allelic classes. First, identities can be calculated conditional on allele frequencies, simulated over 50,000 generations, for a population of $2N=1000$ (dots in Fig. 1). Allele frequencies were simulated over 50,000 generations, using the exact backwards transition matrix calculated for the Wright-Fisher model. Second, the identities can be calculated by solving the discrete equivalent of Eq. 1, which gives exact results for the Wright-Fisher model. This involves linear equations for a set of three vectors for $f_{02}, f_{11}, f_{20}$, each of length $2N+1$. The results fit closely with those estimated by simulation, and are indistinguishable in Fig. 1. Finally, the diffusion approximation (Eq. 1) was used. This fits closely over most of the range (compare solid curves with dots). However, it underestimates identities between genes in rare allelic classes e.g. $f_{02}$ for $p = 0$; upper left of Fig. 1). The approximation is expected to fail for $p \sim \frac{1}{N}$, since only a small number of copies are involved. A similar comparison for $2N = 1000$ shows that the discrepancy is then restricted to a narrower region, as expected.

There are substantial differences between identities involving different allelic classes: identities between classes are much lower than those within (compare the lower curve for $f_{11}$ with the upper curves for $f_{02}, f_{20}$). Identities also vary substantially with allele frequency. Within-class identity decreases from 1 when the allele is present in a few copies, down to $1 - 4v_{1,1}$ (Eq. 2) when it is frequent enough for the diffusion approximation to hold $i p \gg \frac{1}{N}$, and then down to a value somewhat greater than the neutral expectation of $1 - 4v$ when the allele nears fixation. The between class identity necessarily approaches that within the commonest class near fixation, and decreases inbetween, because there is then a rapid influx into the rarer class by mutation from the commoner class.
Note that there is considerable variation in the identities, and hence in the distribution of coalescence times, in any particular generation of a simulated population (Fig. 2, top panel). This arises from the random history of allele frequencies (Fig. 2, middle panel), and is an extra source of variation, over and above the intrinsic variation in the actual coalescence time. The latter is a sample from a distribution which itself fluctuates with allele frequencies. The smooth curves shown in Fig. 1 are the identities conditional on current allele frequency, and are an average over the distribution of past fluctuations in allele frequency. Calculation of the variance in pairwise identity would require consideration of associations among sets of four genes.

Figure 3 shows the same comparison as in Fig. 1, but with balancing selection of strength $S_b = Ns_b = 4$ towards an equilibrium point $p_0 = 0.7$. Again, the three methods agree to high accuracy, except for identities within rare allelic classes. Even though selection is much stronger than mutation and drift, it has surprisingly little effect in reducing the identities $f_{2,0}, f_{1,1}, f_{0,2}$. The stationary distribution is now concentrated around $p_0 = 0.7$; because identities depend strongly on allele frequency, this might be expected to alter the identity between random pairs of genes, averaged over the stationary distribution. However, balancing selection only reduces this average to $E[f] = 0.4977$, relative to the neutral value of $\frac{1}{1+4V} = 0.5$. This is because the average identity $f = q^2 f_{2,0} + 2pq f_{1,1} + p^2 f_{0,2}$ is almost independent of allele frequency (e.g. bottom panel of Fig. 2). We consider this issue in more detail below.

### The distribution of coalescence times

Similar equations can be derived for the probability density of total length of the genealogy, $F_{j,k}$. This is a function of the total length, $J$, and current allele frequency, $p$. After a long time, the density approaches a steady state which satisfies:

$$
\frac{\partial F_{j,k}}{\partial t} = \frac{J}{2} H_j - 1L H_{j-1,k} - F_{j,k}L + k H_k - 1L H_{j,k-1} - F_{j,k}L + N + 2Jj)pUq + RN H_{j-1,k+1} - F_{j,k}L + kqJU + p + RN H_{j+1,k-1} - F_{j,k}LN + 2HUP - pL + S pqL \frac{F_{j,k}}{p} + pqf^2_{j,k} + \frac{2F_{j,k}}{p^2}
$$

At $J=0$, we have:

$$
F_{j,0} = \frac{jH_j - 1L}{2q}
$$

$$
F_{j,k} = 0 \text{ for } j \neq k
$$

$$
F_{0,k} = \frac{kH_k - 1L}{2p}
$$

The boundary conditions at $J=0$ set the rate of coalescence within allelic classes as being inversely proportional to the size of the class. (Recall that time has been scaled relative to $2N$). The partial differential equations themselves have essentially the same form as for the identities, and represent the movement of genes between allelic classes, and the diffusion of populations between allele frequency states. (Note that Eqs. 23 in Barton et al. (2002) give the cumulative distribution of coalescence time, rather than the density).

Figure 4 shows the solution to Eqs. 3, for the same parameters as Fig. 3. The rate of coalescence between allelic classes, $F_{1,1}$, is necessarily zero, and so the distribution $F_{1,1}$ passes through the origin (thick lines in Fig. 4). However, when one or other allele is rare, genes in the rarer class are likely to be descended from genes in the common class relatively recently. Hence, $F_{1,1}$ rapidly approaches $F_{2,0}$ when $P$ is rare (top left and bottom right panels of Fig. 4). Coalescence times within a rare allelic class are likely to be very short, unless the two genes derive from the common class via recent mutation. Because the mutational flux from common to rare is high, these two possibilities have comparable probability (see two terms $1/p$ in Eq. 1). Thus, for small $p$ the distribution $F_{0,2}$ is a mixture of a singularity at zero, and a component proportional to $F_{2,0}$ (lower curve in top left and
bottom right panels of Fig. 4). At intermediate frequency, both within-class distributions are close to the neutral expectation, \( \text{Exp}(-\theta T) \) (middle panels of Fig. 4).

Figure 5 compares the distribution of coalescence times between randomly sampled pairs of genes, \( F = q^2 F_{2,0} + 2pq F_{1,1} + p^2 F_{0,2} \), with the neutral formula \( \text{Exp}(-\theta T) \) for the same example of balancing selection. When allele \( P \) is rare, the coalescence time tends to be shorter, whilst as the frequency approaches the fixation, coalescence times again become slightly longer than the neutral expectation. As \( P \) approaches fixation, coalescence times again become slightly shorter than in the absence of selection and allelic structure. Overall, there is little change: in this example, balancing selection increases mean coalescence time by 13.9%.

The relation between Eqs. 1, 3 can be understood by noting that the identity in state is the generating function for selection. The mean coalescence time between genes in different allelic classes is much greater than that within commoner class. Within-class coalescence times approach the same value as within the commoner class as that class approaches fixation. Within-classes, but approaches the intermediate value where the population is most likely to be found \( |p - p_0 = 0.7L \), coalescence times become slightly longer than the neutral expectation. As \( P \) approaches fixation, coalescence times again become slightly shorter than in the absence of selection and allelic structure.

Overall, there is little change: in this example, balancing selection increases mean coalescence time by 13.9%.

The relation between Eqs. 1, 3 can be understood by noting that the identity in state is the generating function for selection. The mean coalescence time between genes in different allelic classes is much greater than that within commoner class. Within-class coalescence times approach the same value as within the commoner class as that class approaches fixation. Within-classes, but approaches the intermediate value where the population is most likely to be found \( |p - p_0 = 0.7L \), coalescence times become slightly longer than the neutral expectation. As \( P \) approaches fixation, coalescence times again become slightly shorter than in the absence of selection and allelic structure.

Overall, there is little change: in this example, balancing selection increases mean coalescence time by 13.9%.

The relation between Eqs. 1, 3 can be understood by noting that the identity in state is the generating function for selection. The mean coalescence time between genes in different allelic classes is much greater than that within commoner class. Within-class coalescence times approach the same value as within the commoner class as that class approaches fixation. Within-classes, but approaches the intermediate value where the population is most likely to be found \( |p - p_0 = 0.7L \), coalescence times become slightly longer than the neutral expectation. As \( P \) approaches fixation, coalescence times again become slightly shorter than in the absence of selection and allelic structure.

Overall, there is little change: in this example, balancing selection increases mean coalescence time by 13.9%.

The relation between Eqs. 1, 3 can be understood by noting that the identity in state is the generating function for selection. The mean coalescence time between genes in different allelic classes is much greater than that within commoner class. Within-class coalescence times approach the same value as within the commoner class as that class approaches fixation. Within-classes, but approaches the intermediate value where the population is most likely to be found \( |p - p_0 = 0.7L \), coalescence times become slightly longer than the neutral expectation. As \( P \) approaches fixation, coalescence times again become slightly shorter than in the absence of selection and allelic structure.

Overall, there is little change: in this example, balancing selection increases mean coalescence time by 13.9%.

The relation between Eqs. 1, 3 can be understood by noting that the identity in state is the generating function for selection. The mean coalescence time between genes in different allelic classes is much greater than that within commoner class. Within-class coalescence times approach the same value as within the commoner class as that class approaches fixation. Within-classes, but approaches the intermediate value where the population is most likely to be found \( |p - p_0 = 0.7L \), coalescence times become slightly longer than the neutral expectation. As \( P \) approaches fixation, coalescence times again become slightly shorter than in the absence of selection and allelic structure.

Overall, there is little change: in this example, balancing selection increases mean coalescence time by 13.9%.

The relation between Eqs. 1, 3 can be understood by noting that the identity in state is the generating function for selection. The mean coalescence time between genes in different allelic classes is much greater than that within commoner class. Within-class coalescence times approach the same value as within the commoner class as that class approaches fixation. Within-classes, but approaches the intermediate value where the population is most likely to be found \( |p - p_0 = 0.7L \), coalescence times become slightly longer than the neutral expectation. As \( P \) approaches fixation, coalescence times again become slightly shorter than in the absence of selection and allelic structure.

Overall, there is little change: in this example, balancing selection increases mean coalescence time by 13.9%.

The relation between Eqs. 1, 3 can be understood by noting that the identity in state is the generating function for selection. The mean coalescence time between genes in different allelic classes is much greater than that within commoner class. Within-class coalescence times approach the same value as within the commoner class as that class approaches fixation. Within-classes, but approaches the intermediate value where the population is most likely to be found \( |p - p_0 = 0.7L \), coalescence times become slightly longer than the neutral expectation. As \( P \) approaches fixation, coalescence times again become slightly shorter than in the absence of selection and allelic structure.

Overall, there is little change: in this example, balancing selection increases mean coalescence time by 13.9%.

Figure 6 shows how the mean coalescence time changes with allele frequency, with and without balancing selection. The mean coalescence time between genes in different allelic classes is much greater than that within classes, but approaches the same value as within the commoner class as that class approaches fixation. Within-class coalescence times approach \( 4 Up J_{1,1} \) as the allele becomes rare. This value is determined by a balance between the rapid rate of coalescence within rare classes, and the rapid influx of copies by mutation from the commoner class.

\section*{A CHANGE OF VARIABLES}

In order to make some approximations, and to understand average identity, it is helpful to change variables, as follows. First, consider the pairwise case. Let:

\begin{align*}
q &= q^2 f_{2,0} + 2pq f_{1,1} + p^2 f_{0,2} \\
D &= -q f_{2,0} + Hq - pL f_{1,1} + p f_{0,2} \\
q &= f_{2,0} - 2 f_{1,1} + f_{0,2} \\
J &= f - 2pD + p^2 q \\
\eta &= f + Hq - pL D - pq q \\
\gamma &= f + 2qD + q^2 q
\end{align*}

(7)
Similar transformations apply for the expected total length, J, and for the distribution of coalescence times, F. The average identity amongst randomly chosen pairs of genes is \( f : D \) is the difference in identity between a gene associated with \( P \) and a random partner, and a gene associated with \( Q \) and a random partner; and \( q \) is twice the difference in identity between alleles within classes, and alleles between classes. Eqs. 1 transform to:

\[
0 = 1 - H + 4 V L f + L \frac{H q - p L}{p q} H p - p L + R N D + 4 p q S D
\]

\[
0 = -2 J L + 2 V + U + U \frac{H q - p L}{p q} H p - p L + R N D + L \frac{H q - p L}{p q} - 2 q L + H p - q + 2 S p q L q
\]

\[
L J \frac{H q - p L}{p q} q + 2 \frac{H p - q L}{p q} q N + \frac{H p - q L}{p q} H 2 p q L q + L J \frac{H q - p L}{p q} q + 2 \frac{H p - q L}{p q} q N + \frac{H p - q L}{p q} H 2 p q L q + L J \frac{H q - p L}{p q} q + 2 \frac{H p - q L}{p q} q N + \frac{H p - q L}{p q} H 2 p q L q + L J \frac{H q - p L}{p q} q + 2 \frac{H p - q L}{p q} q N + \frac{H p - q L}{p q} H 2 p q L q + 0
\]

where \( L = 12 S p q + 2 U H p - p L + \frac{p q}{2} q M \)

Each variable is augmented by a diffusion term, \( L() \). As is shown below (Eq. 9), the average of this diffusion term over the stationary distribution is zero; thus, it shifts the variable without producing a net change. The first equation shows that average identity is augmented by \( 4 p q S D + p q \), which is the product of the change in allele frequency due to selection, and the difference in identity between genes associated with \( P \) rather than \( Q \). Net identity increases if a higher identity is associated with an allele favoured by selection. The crucial quantity, then, is \( D \). The second equation shows that this is reduced by recombination and mutation, especially near the edges, and augmented by a term proportional to \( q \), which is the difference in identity within and between allelic classes. The last equation, for \( q \), shows that it is augmented by coalescence, especially at the boundaries, in proportion to \( \frac{H q - p L}{p q} q \). It is also augmented by a term proportional to \( D \).

The equations for the distribution of coalescence times are obtained by setting \( V = 0 \), and dropping the terms which do not involve \( f, D \) or \( q \) (i.e. 1 in Eq. 8a, and \( \frac{1}{pq} \) in Eq. 8c). Boundary conditions at \( t = 0 \) are that \( F = 1, D = 0 \), and \( q = \frac{1}{pq} \). The expected total length (which is twice the mean pairwise coalescence time for \( n = 2 \)) given by the same equations as above, but setting \( V = 0 \), and subtracting \( \frac{1}{pq} \) from Eq. 8c.

Figure 7 shows the transformation for mean coalescence time, to the variables \( t, D^\frac{1}{2}, \frac{1}{q} \). The mean coalescence time between randomly chosen pairs of genes varies rather little with allele frequency, and increases only moderately with increasing balancing selection (top panel). This is because the equation for \( t \) is mainly driven by the terms \( 1 - t \), which leads to the standard solution \( t = \frac{1}{2} \). The diffusion terms \( L \frac{H q - p L}{p q} t - 4 D^\frac{1}{2} L \) redistribute \( t \) across the range of allele frequencies, but as we show below, do not alter its average value. The main driving term is small, because under balancing selection both \( S \) and \( D \) are zero somewhere in the interior, and so their product is small when allele frequency is concentrated in the interior.

Figure 8 shows the same comparisons, but for purifying selection. The overall mean \( t \) is affected little by weak selection (\( S \leq 1 \), say). With stronger purifying selection, mean coalescence time is substantially reduced when the favourable allele becomes rare (left of top left panel), but is slightly increased when the favourable allele is common. The average identity is insensitive to selection when selection is weak because then, \( D^\frac{1}{2} \) changes sign in the centre and so its product with \( S p q \) also changes sign; the net effect through the driving term \( 4 S p q D^\frac{1}{2} \) is thus small. As selection becomes stronger, \( D^\frac{1}{2} \) becomes positive over a wider region, implying that the mean coalescence times involving genes in the favoured allelic class become longer. However, the net effect on the pattern of \( t \) is hard to predict, because the diffusion term is strong. The next section sets out a simple result for the mean coalescence time, averaged over the stationary distribution, which does not involve this diffusion term.
THE NET EFFECT OF SELECTION

A remarkably simple result is obtained by taking the expectation of the average identity over the stationary distribution, $E[q]$. We know that the stationary distribution satisfies the forwards diffusion 0=H2 S pq + 2 Hp - p L L y - H2 pq yL. Integrating the first of Eqs. 8 over the stationary distribution $y$, we have:

$$0 = 1 - H1 + 4 V E[H] + y L H^2 p - 4 DL p + 4 E[O] p q D[q] d$$  \hspace{1cm} (9)

Integrating by parts, the third term vanishes, and we have:

$$E[H] = \frac{1 + 4 E[O] p q D[q]}{1 + 4 V}$$  \hspace{1cm} (10)

We immediately see that in the neutral case $E[q]$ is unaffected by the arbitrary labelling. Selection will perturb average identity by a proportion that depends on $E[S pq]$ [p]; this will be small for purifying selection, since $D$ changes sign somewhere in the centre. However, it will be large and negative for balancing selection, if the null point of selection coincides with the null point for $D$. Then, both $D$ and $S$ will change sign near the centre and so $E[S pq]$ is negligible throughout. Thus, balancing selection is expected to reduce average identity.

The mean coalescence time is given by the same equation as Eq. 9, but with $V$ set to zero, and $D^2 = p^2 q^2 - t_1,1 L q^2 - t_2,0 L$:

$$E[H] = 1 + 4 E[O] p q D[q]$$  \hspace{1cm} (11)

Similarly, the distribution of coalescence times, averaged over random pairs of genes and over the distribution of allele frequencies, is:

$$E[H] = \exp(-t) + 4 \sum_{q}^{1} \exp(-t) - t' L E[H] p q D[q] t'$$  \hspace{1cm} (12)

The expected mean coalescence time can be calculated either directly, or by using the right side of Eq. 11. In numerical calculations, the latter is more accurate, both because it gives what is usually a small deviation from the neutral expectation of 1, and because the regions near fixation (where the stationary distribution diverges for $U p$, $U q < 1$) do not contribute significantly to the integral (since $pq$ and $D^2$ tend to zero at the boundaries). The boundaries do not contribute to the deviation from neutrality because in these regions selection is negligible relative to drift, and there is rapid flow between backgrounds.

The relationships of Eq. 9-12 extend to arbitrary numbers of genes in the sample. To be definite, consider the expected total length of the genealogy, $J_{jk}$ (Eq. 5). However, the same argument applies to other properties of the genealogy, such as the identity in state, or the distribution of times to the most recent common ancestor. As mentioned above, one can write down the generating function for the complete density in the same form, and so this result applies for any quantities derived from it.

Summing over the binomial probability density for the number of genes $q$, $k< sampled from each background, we obtain:

$$0 = n + \frac{n H n - 1 L}{2} H_{n-1} - J_{n} L + 4 S p q D_{n} + L H^2 p J_{n} - 4 D_{n} L$$

where $J_{n} = \sum_{j+k=n} q^j p^k J_{j,k}$ and $D_{n} = \sum_{j+k=n} q^j p^k J_{j,k}$

\hspace{1cm} (13)
The first term $n$, represents the increase in tree length at a rate $n$, when $n$ lineages are present. The second term represents coalescence at a rate $\frac{n^{n-1}}{2}$, which is half the regression of average tree length on the number of $P$ alleles in the sample (i.e., $J_{j,k}$). This suggests that the effects of selection will primarily be on the deeper parts of the genealogy (i.e., small $j$). Numerical calculations confirm our intuition that the effects of selection decrease rapidly as $j$ increases. Figure 9 shows the successive perturbations due to selection for up to $n = 50$ genes, compared with the successive contributions under neutrality. The $D_j$ decrease for $j > 10$, and so the net perturbation, which is further reduced by the factor $\frac{2}{j^{j-1}L}$, quickly become negligible. The sum of the perturbations is -0.29, compared with a neutral expectation of $E[D] = 8.96$. Thus, purifying selection only reduces total tree length by 3.3% in this example, and more than half of that effect arises during the time when there are only 2 or 3 lineages extant. (As explained under "Numerical methods", we use the insensitivity of the genealogy to allele frequency fluctuations in our calculations of terms involving large numbers of genes).

Equation 14 can be interpreted as an instance of Robertson’s (1966) “Secondary Theorem of Natural Selection”, which states that the increase in any quantity caused by selection is equal to the covariance between that quantity and fitness. (This was independently developed into a general representation of selection by Price, 1970). In terms of the original variables, $E[D] = \frac{2}{j^{j-1}L}$, which is equal to the covariance between the fitness of the sample of $n$ genes, and the expected total length of the genealogy which relates these genes. (Here, relative fitness of a sample with $k$ copies of allele $P$ is $sk$, or $2Ns = 2Sk$ when rescaled). Equation 14 shows that this covariance is exactly equal to the increase in the expected total length caused by selection. Because essentially the same equations apply to the identity $f$, which can be viewed as the generating function of the genealogy, we can see that relations similar to Eq. 14 apply to any quantities which are linear functions of the probability density of genealogies (for example, the $k$’th moments of coalescence).

Approximations

There does not appear to be an explicit solution to Eqs. 1 or 8, even in the absence of selection. We therefore explore several approximations, in the hope that these might extend to more complex situations which are difficult to investigate numerically. We consider in turn the extremes of no mixing between allelic classes, assuming that change in allele frequency is negligible, and assuming that there is rapid mixing between backgrounds by mutation and recombination.
No movement between backgrounds

If there is no recombination or mutation between alleles $U, R = 0 \text{ then the three identities given by Eqs. 1 change independently. The between-class identity } f_{ij,1} \text{ clearly tends to zero, whilst the within class identities are those of two separate populations of size } p, q \text{, which fluctuate according to a diffusion process. (Strong frequency-dependence is required to prevent loss of variation at the selected locus in the absence of mutation: near the boundaries: } S \text{ ff } p^2 \text{ as } p \text{ ff } 0, a > 1 \text{). However, even this uncoupling into a set of single-variable equations does not lead to an explicit solution. The identities could be found by a change of timescale in the standard coalescent, but this does not lead to closed form solutions (Donnelly and Kurtz, 1999b).}

Weak random drift

If random drift is weak relative to mutation, recombination and selection, then allele frequencies will be close to a deterministic limit. Almost all treatments of the effects of selection on genealogies assume this limit, and apply the standard structured coalescent (e.g. Kaplan et al., 1988, 1989; Hudson and Kaplan, 1995; Wakeley, 2001). However, strong underdominance has little effect, because the population is then near fixation. For the model presented here, this limit corresponds to dropping the diffusion terms (i.e. those terms involving $q_p$ or $q_{p,p}$ from Eq. 1), thereby assuming that the population has always had the same allele frequency. This approximation is independent of the strength of selection, but does depend on both mutation and recombination rates, which determine the rate of mixing between allelic classes. Under this approximation, the mean pairwise coalescence time is:

$$ t = \frac{p^2 q^2 R + 3 p^2 q^2 U + 4 H_U C_2 + p q R L^2 + U H c_2 + 8 U R c_2 p q + 4 p^2 q^2 R^2}{p q R + p^2 q^2 R + 4 H_U C_2 + p q R L^2 + U H c_2 + 3 p^2 q^2 L} $$

where $c_1 = H p q^2 + q p^2 L$

A similar expression can be obtained for the identity in allelic state, which allows calculation of higher moments of the distribution of coalescence times. The overall mean, $E=t$ can be estimated either by integrating over the stationary distribution, or by fixing $p$ at its deterministic equilibrium value (Neuhauser and Krone, 1997).

Figure 10 shows the mean coalescence time as a function of the strength of balancing selection. Here, $U=N_m=0.5$ or 0.25 and so effects are not large: balancing selection only has a substantial effect on genealogies when mutation is low enough that genes rarely move between backgrounds. The mean coalescence time does approach the prediction from Eq. 15 for large $S$ (right of figure), but balancing selection must be extremely strong for the deterministic limit to be accurate. That is, weak random fluctuations in allele frequency can substantially reduce coalescence times. Note that weak disruptive selection (e.g. underdominance) reduces coalescence times slightly, because allele frequencies tend to sweep back and forth between alternative alleles (see left of graph). However, strong underdominance has little effect, because the population is then near fixation.

Figure 11 shows an example in which fluctuations significantly reduce mean coalescence time despite strong selection. With $S = N_s = 32$, allele frequencies cluster around the equilibrium of $p_0 = 0.7$ (bell-shaped curve). The coalescence time is almost independent of allele frequency, simply because populations away from equilibrium are recently derived from populations close to equilibrium (thin solid curve). In contrast, the deterministic prediction (dashed line) ignores the diffusion of populations between different allele frequencies, and so depends more strongly on allele frequency. Taking the value of this approximation at $p_0 = 0.7$ gives a substantial overestimate of mean coalescence time, even under such strong selection.

Figures 12 and 13 show similar results, but for mutation/selection balance rather than for balancing selection. The lines to the right of Fig. 12 show the prediction for the deterministic limit (Eq. 15); as $S \text{ ff } 1$, mean coalescence time is reduced by a factor $1 - U_S + O(U_S^2)$ (Charlesworth et al., 1993). This effect of "background selection" against deleterious alleles is reduced substantially by random drift for $S = N_s$ less than about 3. Figure 13 shows the stationary distribution at $S=3$, $L = \frac{1}{S} = \frac{1}{3}$; as for balancing selection, a moderate degree of fluctuation around the deterministic expectation substantially reduces the effect of selection in reducing mean coalescence time (here, from $1 - U_S = 0.875$ at $S \text{ ff } 1$ to 0.908 at $S = 3$).
Rapid mixing

If mutation and/or recombination are strong relative to drift (U or R >> 1) then there will be little divergence between allelic classes: D and q will be small, and f close to $1_{1+4V}$. First, consider the case with U, S ~1 and R>>1. From Eqs. 8, we see that $G$ is augmented by the term $\lambda L pq$ and dissipated by recombination, and so is $O U R_1 L$. D is generated by $g_q$, and dissipated by recombination and also mutation at the boundaries $p q$. Hence, $D \sim 1 R^2$. The perturbation to $f$ is therefore only of order $1 R^2$. Letting $f = \frac{1}{1+4V} + f$ the leading terms are:

$$0 = -H_1 + 4 V L f + I 2 S p q + 2 H_p - p L U + \frac{p q}{2} f_{p M H p} f - 4 D L + 4 p q S D$$

$$0 = -2 J U H_q - p L H_p - p L + R N D +$$

$$H H p - q L q - 2 S p q - 4 H_p - p L U q - p q f_{p q}$$

$$0 = -4 R q + H_1 - f L p q$$

Hence:

$$q = \frac{V}{R p q H_1 + 4 V L 2}$$

$$D = \frac{V H S p q + 2 H_p - p L U L}{\frac{1}{1+4V} - 4 V H_1 + 4 V L^2 E A S p q H S p q + 2 H_p - p L U L H U p - p L H p - p L + R p q L R}$$

Note that the term $U p - p L H q - p L$ in the denominator of the expression for D is negligible except near the boundaries, when $R p q \sim 1$. The perturbation $f$ to average identity is not given explicitly. However, its expected value, integrated over the stationary distribution, is given by Eq. 10. Hence:

$$E g_f D = \frac{1}{1+4V} + \frac{4}{1+4V} E g S p q D g D =$$

$$1 + 4 V - \frac{4 V H_1 + 4 V L^2 E A S p q H S p q + 2 H_p - p L U L H U p - p L H p - p L + R p q L R}{E}$$

Because the stationary distribution is known explicitly, this perturbation can readily be calculated. Moreover, it has a simple form when seen as a function of $V$. This allows us to take the inverse Laplace Transform, which shows that the distribution of coalescence times is:

$$E g f 0 \times D D =$$

$$\text{Exp} g_0 - t D J L + H t - 1 L E A S p q H S p q + 2 H_p - p L U L H U p - p L H p - p L + R p q L R \text{ EN and}$$

$$E g 0 t D = 1 + E A S p q H S p q + 2 H_p - p L U L H U p - p L H p - p L + R p q L R \text{ E}$$

Note that the key term $E[\cdot]$ in Eq. 19 is negative for purifying selection, and positive for balancing selection.

If balancing selection is strong enough that the stationary density near the boundaries is negligible, and if $R>>U$, then the term involving mutation in the denominator is negligible, and:

$$E g F 0 \times D D - \text{Exp} g_0 - t D J L + H t - 1 L E A S^2 p q \frac{R^2}{EN}$$

$$E g 0 t D - 1 + E A S^2 p q \frac{R^2}{E}$$

Thus, when linkage is loose the effect of selection is to increase mean coalescence times by an amount proportional to the additive variance in fitness $2 S^2 p q$. This can be understood as an inflation in the rate of random genetic drift due to inherited variation in fitness (Hill and Robertson, 1966; Santiago & Caballero, 1995). With
balancing selection, the marginal selection coefficient $S = S_k \frac{p_0 - p_L}{p_0}$ is zero at $p = p_0$. However, allele frequency fluctuations around this expectation with variance $\text{var}(p_L) \approx \frac{S_k}{4 L S_0}$ for large $S_k$. Hence, $\mathbb{E}(S^T_{pq} R^U) \approx S_0 \frac{p_0 q_0}{4 R U}$. Note that this is not the same as the limit of Eq. 15 as $R \rightarrow \infty$, which is $\frac{1}{4 R}$. Allele frequency fluctuations have a significant effect which cannot be neglected even for strong selection.

We examine the accuracy of these approximations by considering the effect of increasing recombination. Figure 14 shows the increase in mean coalescence time caused by balancing selection, plotted against recombination rate. Mutation is set to a small positive value $\Upsilon = 0.05$, to ensure that a stationary distribution exists. However, mutation has negligible effect on the results unless linkage is tight $\Upsilon \ll U$. The increase over the neutral value, $\mathbb{E}(S^T_{pq} R^U - 1)$, is plotted on a log scale, because when $R$ is large, small effects must be discerned. As selection increases ($S_k = 4, 8, 16, 32$, bottom to top), mean coalescence time converges to the deterministic limit (thick line; Eq. 15). However, convergence is slow for large $R$. There, the large $R$ approximation of Eq. 19 is accurate (dashed lines to right). However, the approximation is only good for $R > 10$, in which case mean coalescence time is increased by at most 2.5%. The large $R$ approximation is not helpful for parameters that give a large

Figure 15 shows a similar plot for the decrease in mean coalescence time caused by purifying selection. Now, mutation is set to a moderately high value $\Upsilon = 0.5$ with weak mutation, the population would almost always be fixed, and effects on linked variation would be negligible. The deterministic limit of Eq. 15 (upper thick line) now performs poorly. This is because it is based on the assumption that allele frequency is close to the deterministic equilibrium of $1 - \frac{U}{S}$, which is never the case for $U = 0.5$, even when selection is strong. This approximation is only expected to be accurate for $U >> 1$. The large $R$ approximation of Eq. 19 is more accurate (dashed lines), but systematically underestimates the effect by ~18%. Examination of the differences in mean coalescence time between backgrounds, $\Delta^S_{pq}$, shows that the approximation of Eqs. 17 breaks down near the margins $\mathbb{E}(S^T_{pq} R^U) \approx \frac{1}{4 R} L$. Since the stationary density is appreciable in this region for $R = 10$, much larger recombination rates would be needed for accuracy to be improved. As for balancing selection, this approximation is only accurate in cases where the effect on coalescence time is small.

For tight linkage, the effect of purifying selection at first increases with selection, but then decreases (see left of Fig. 15). This can be seen more clearly in Fig. 16, which shows the mean coalescence time as a function of $S = Ns$, with complete linkage. As selection becomes very strong, the rare allele is driven out of the population, and so the genealogy returns towards its form under the neutral coalescent.

**Large genealogies**

We have concentrated on numerical examples for pairwise coalescence time partly because computations are then much faster, but also because the effects of fluctuations in allele frequency, and hence of selection, are primarily on the deeper parts of the genealogy, when there are just a few ancestral lineages. Here, we use Eq. 5 to consider the effect of purifying selection on the expected total length of a larger genealogy. Figure 17 shows how the expected length depends on allele frequency and on the composition of the sample. If 5 genes of type P are sampled, then the genealogy is much shorter when that allele is rare (line rising steeply from left to right); similarly, if 5 copies of type Q are sampled, the genealogy is shorter when Q is rare, because coalescence is much more rapid within the rarer class. However, the relationship with allele frequency is quite weak for mixed samples, because coalescence occurs more slowly within each allelic class, and so lineages are likely to move between classes by mutation. In this example, there is purifying selection $S = Ns = 2$. However, the patterns for a neutral locus, or for other kinds of selection, are similar.

Figure 18 shows the net effect of purifying selection, $S = Ns$, on the expected total length, for up to $n = 50$ genes. Mutation rate is $\Upsilon = N_m = 0.5$, and there is no recombination. Equation 14 shows that the net effect of selection can be separated exactly into a sum, the $j$’th term being due to the time when $j$ lineages are present. Thus, Fig. 18 shows the total effect on genealogies with 50 genes (upper points), and the component effects on genealogies with up to 2, 3, 4, 5… genes (lower series of points). As explained above, most of the perturbation due to selection accrues when there are just a few lineages present. Overall, the effects are small relative to the expected total tree length under neutrality $L^S = 8.96$ for $n = 50$ genes. The curves on the right give the simple
DISCUSSION

We have used the equations set out by Kaplan et al. (1988) to find the effects of weak selection \( \mu N s \ll 1 \) on genealogies at a linked neutral locus. These genealogies are produced by the structured coalescent process, in which genes move between the genetic backgrounds defined by the selected locus as a result of both recombination and mutation of the selected alleles. Most previous studies have assumed that allele frequencies evolve deterministically, and so are restricted to strong selection \( \mu N s \gg 1 \). Here, we allow allele frequencies at the selected locus to evolve as a diffusion process. This leads to sets of equations which are the sum of two terms: recursions which describe the structured coalescent at given allele frequency, and diffusion terms which allow for random changes in allele frequency through time. We find that quite small stochastic fluctuations in the frequencies of alternative genetic backgrounds substantially reduce the effects of selection.

In this paper, we have only considered selection on a single biallelic locus, and have for the most part sampled two genes from the linked neutral locus. Moreover, we have assumed that the selected locus has reached a stationary state, so that properties of the genealogy such as mean coalescence time can be taken to be functions of allele frequency only. In principle, it is straightforward to extend the method. Non-stationary processes can be described by taking the variables to be functions of time as well as allele frequency, and following them using the same backwards diffusion as in Eqs. 1. For example, one might ask about the genealogy immediately after a substitution has occurred by chance: is the genealogy shortened in the same way as with a 'selective sweep'?

Allowing more loci or more alleles under selection requires that the variables be functions of genotype frequencies, and so greatly slows down numerical calculations. Also, more variables need to be followed, because genes can find themselves in many more genetic backgrounds. Extension to larger samples of genes at the neutral locus is simpler (Kaplan et al., 1988). Now, we follow the relationship between a set of \( j \) genes in background \( Q \), and \( k \) genes in background \( P \); for a sample of \( n \) genes, this requires \( \mu n - 1 \mu n + 4 L \) 2 variables. This does not raise substantial difficulties, because one can break up the calculation into sets of equations involving 2, 3... \( n \) genes at a time. Direct solution of the differential equations (Eq. 5) are feasible up to 5 or so genes, and much larger numbers can be approximated by assuming that coalescence from \( n \) genes down to \( \sim 5 \) genes occurs quickly, relative to the timescale of the diffusion process (Fig. 9).

Alternative methods for dealing with the effects of weak selection on genealogical structure are largely based on simulation (or more precisely, on Monte Carlo methods). The traditional approach is to use forwards simulations of the whole population. This can be an efficient representation of even large populations, because a rescaling based on the diffusion approximation allows small populations to represent much larger populations (e.g. Hill and Robertson, 1966). For example, McVean and Charlesworth (1999) found that simulation of mutation and selection on several thousand loci in only 100 individuals was a good approximation to weaker forces acting on much larger populations. Williamson and Orive (2002) also found that simulations of 100 individuals to be adequate. Nevertheless, very large numbers of replicates are needed to obtain accurate estimates of sample statistics.

Neuhauser and Krone (1997) introduced a method which has the advantage that it focuses on the sample, rather than on the whole population. Tracing back from the sample, lineages coalesce as a result of random drift. Selection is included by allowing branching events which generate "virtual ancestors". This leads to an "ancestral selection graph" which eventually traces back to one ultimate ancestor. If the type of this ultimate ancestor is specified (for example, as a draw from the stationary distribution) then the actual genealogy can be resolved by pruning out virtual ancestors using a rule which is biased towards the favourable allele. This method has been extended by Neuhauser (1999) and by Donnelly and Kurtz (1999a) to cover certain kinds of frequency dependent selection. The algorithm has been refined by Slade (2000a,b), such that the graph only needs to be traced back to the most recent common ancestor, and by Fearnhead (2001), who shows how the simulation results can be guaranteed to reach stationarity. Donnelly and Kurtz (1999a) and Slade (2001) have also shown
how recombination can be included with selection in the algorithm. However, despite these various advances, the method is still computationally intensive for strong selection: for example, with overdominance, only Ns < 10 or so can be simulated (Slade, 2000a). Moreover, it is limited to certain kinds of selection: linear frequency dependence or selection on diploids requires branching into three potential ancestors instead of two, and more generally, a k'th order polynomial dependence of haploid fitness on allele frequency requires branching into \( k + 2 \) virtual ancestors. In practice, anything beyond the simplest kind of epistasis of frequency dependence is ruled out.

Other simulation techniques are closer to the analytical framework presented here, in that they follow the state of the whole population backwards through time. Donnelly et al. (2001) discuss methods for importance sampling, which start with the well-understood neutral process, and apply a bias which represents the action of selection. This is most efficient when selection is weak. Slatkin (2001) begins by reversing the selective process, which should allow stronger selection to be represented accurately. Because the diffusion is reversible with additive selection, the procedure is exact in this case. With non-additive selection, Slatkin (2001) uses a procedure which approximates the correct backwards diffusion. Finally, one can follow the entire path of allele frequencies back through time, and use the Metropolis-Hastings algorithm to sample the appropriate distribution of paths. Under the diffusion approximation, the probability of any particular path is given by a Gaussian distribution of velocities around their deterministic expectation, which approximates the product of Markov transition matrices (Schulman, 1981; Rouhani and Barton, 1987).

We see the main advantage of our method as giving the possibility of analytical approximations which will allow progress in understanding more complex cases. Most existing results assume that allele frequencies evolve deterministically (e.g. Kaplan et al., 1988, 1989; Hudson and Kaplan, 1995; Wakeley, 2001). Our numerical results show that this deterministic approximation is only accurate for quite strong selection: moderate fluctuations, reflected in dispersion of the stationary distribution, is sufficient to reduce effects substantially below those predicted. An obvious extension for the future is to make an expansion in powers of 1/Ns, so as to improve on the deterministic approximation. The opposite approach is to examine the effects of weak selection. Krone and Neuhauser (1997; Theorem 4.26) show that the effect of purifying selection and symmetric mutation on coalescence times is \( \mathcal{O}(N_s)^{1/2} \) (see Fig. 12). This result was obtained by showing that various terms, each corresponding to an alternative topologies of the ancestral graph, cancel. Under our approach, we see immediately from Eq. 19 that the first order contribution is zero: to leading order, \( \mathcal{D} \) is an odd function centred on \( p=0.5 \), and so \( E[\mathcal{D}] = 0 \) for a symmetrical stationary distribution. (It is not possible for us to compare with Krone and Neuhauser’s Theorem 4.19, because this gives the probability that two individuals are identical in the sense that they have experienced no mutations that alter their allelic class since they shared a common ancestor. This is not a property of the genealogy alone, and is not the same as the classical "identity by descent").

We have shown that the expected change in certain properties of a genealogy caused by selection is equal to the covariance between those properties, and fitness. Applying Robertson’s (1966) "Secondary Theorem" (or Price’s, 1970, equation) in this way is unusual in several respects. First, the argument applies to the expected value of a randomly distributed variable. Second, the argument is applied to samples of genes, rather than to individuals. Third, the phenotype of the sample of genes is taken to be a measure of their joint ancestry - in this instance, the expected length of the genealogy. However, one can see that Price’s (1970) arguments do apply with these extensions: a sample of genes can be connected with their offspring in the next generation, and the fitness of the whole sample can be seen as covarying with its genealogical properties. By including the transmission terms in Price’s (1970) equation, one can derive the whole of Eq. 14: the expected state of the offspring sample changes as a result of the extra time step (first term) and as a result of coalescence (second term). It may be fruitful to apply this stochastic extension of Price’s equations to other problems.

Recent simulations suggest that the effects of selection at one site on the genealogy is surprisingly weak (Golding, 1997, Neuhauser and Krone, 1997, Przeworski et al. 1999, Williamson and Orive, 2002). Of course, the frequencies of selected alleles are strongly affected: purifying eliminates allelic variation, and balancing selection maintains it. Thus, observations on biased codon usage can give direct estimates of \( N_s \) (McVean and Charlesworth, 2000), because the frequencies of alternative bases at the third position are themselves under selection through their effects on translation. However, the structure of the genealogy is typically affected very little by selection, and thus, observations of this genealogy, or of closely linked neutral variation, tells us little about the
action of selection. This is a serious practical limitation, because we often do not know the actual nucleotides which cause fitness differences, and so must base our inferences on synonymous or non-coding variation that we assume to be neutral.

Przeworski et al. (1999) describe the "competition of alleles on a genealogy" whose structure is largely independent of the distribution of the selected alleles. This description is somewhat misleading, because the genealogy does depend strongly on the allelic state of the sample, even when no selection is acting (e.g. Fig. 1): genes in the same allelic state are more likely to share a recent common ancestor. However, when genealogies are averaged over the possible allelic configurations, and over the stationary distribution of allele frequencies, the result is close to the neutral coalescent. Since we typically do not know which alleles influence fitness, we can usually observe only this average.

To examine the claim that selection has small effects on the genealogy, even at the selected locus, we must distinguish balancing from purifying selection. In the former case, there can be very strong effects provided that mutation rates are extremely low ($U<<1$), since this allows time for the two genetic backgrounds to diverge considerably. Such divergence can of course be observed directly, for example at self-incompatibility loci in plants, or the major histocompatibility locus in mammals (Hughes, 1999). Extension of this argument to multiple balanced polymorphisms shows that in a sufficiently large population, and with sufficiently stable selection, coalescence times can become extremely long. However, random drift in even large populations greatly reduces this effect (Barton and Navarro, 2002), because balancing selection cannot maintain every combination of selected alleles at constant frequency. Similarly, our single-locus results (Fig. 10) show that even with $N_{sp} = 20$, the effect on neutral variability can be more than halved. In reality, fluctuations in selection are likely to further reduce the effect of balancing selection below the ideal case of an extremely large population under constant conditions.

Recent discussions have concentrated on the effects of selection against deleterious mutations on genealogical structure. Purifying selection is of most general importance, because it acts on all functional sequences, and because its effects can be substantial in aggregate, at least for organisms with a high genomic mutation rate and in regions of low recombination (Charlesworth et al., 1993). Purifying selection on a single site has surprisingly weak effects. For example, Williamson and Orive (2002, Table 4) found that with a total mutation rate $U=5$, the expected total length of a genealogy connecting 50 genes is reduced by at most 28%, at $S=N_{sp} = 5$. (We express $U$ in terms of our model of reversible mutation; Williamson and Orive used $4 N_{sp} \mu_i p_i = 4 N_{sp} \mu_i q_i = 4 U p = 10$; their $2N_{sp}$ is equivalent to our $2S$). Przeworski et al. (1999) used a much lower mutation rate ($U=0.1$) and observed a reduction of at most 0.53% at $S=3$. However, because of the computational difficulties of simulations of either the whole population, or of the ancestral selection graph, the view that purifying selection at a single site has small effects on genealogies is based on quite limited numerical results.

We can identify three distinct reasons why purifying selection should have little effect on genealogies. First, even when selection is strong relative to drift, the effect of a single site is small. With complete linkage, and two selected alleles, effective population size is reduced by a factor $1 - \frac{m}{L}$ for large $N_{sp}$ (Charlesworth et al., 1993). This is a consequence of the fact that if back-mutation is negligible, and if the fittest class is to be maintained indefinitely, then only the fraction $1 - \frac{m}{L}$ of the population which is free of deleterious mutations can contribute in the long term. Since deleterious alleles are likely to be rare at any one site, the effect on neutral variability will be small. Nevertheless, the cumulative effects of many sites can be large. Averaging over a genetic map of length $R$, the effective population size is reduced by $\text{Exp} e^{-\mu_i R_{L}}$, a factor which depends on the total mutation rate per map distance (Hudson and Kaplan, 1995).

Second, for a given mutation rate, a significant effect is seen only for intermediate selection strengths. For strong selection, deleterious mutations become negligibly rare, while for weak selection, the effect is only of second order in $S$ (Fig. 12; Neuhauser and Krone, 1997). The maximum possible effect increases with mutation rate (Fig. 16), because the frequency of deleterious mutations is increased. However, this effect is offset by the more frequent movement of genes between the two alternative backgrounds, which reduces the covariance between the allele frequency in the sample, and the structure of the genealogy. For fixed and large $S$, mean coalescence time decreases in proportion to $U$ (right of Fig. 16), but for weak selection, it is almost independent of mutation rate.
Presumably, the effect of increased frequency of deleterious alleles, and of faster mixing between backgrounds, counterbalance when $S$ is small.

Finally, the effect of selection on large genealogies is only appreciable deep in the tree, when just a few lineages segregate. This is because lineages rapidly coalesce down to a small number of ancestors, and so the fluctuations in allele frequency which mediate the effects of selection have little influence during this period. Consider the perturbation to the expected total length of the genealogy, which is plotted against the strength of purifying selection in Fig. 18, for $U=0.5$. For 50 genes, the greatest reduction below the neutral value is by 4.2%, when $S = Ns \sim 3$. About half of this reduction is due to the effects of selection during the time when there are fewer than 6 lineages present. Overall, the proportionate reduction in tree length is about the same for small and large samples. (In this example, pairwise coalescence time is reduced by at most 5.5%). Statistics such as the length of external branches are expected to be much less sensitive to the effects of selection, an argument supported by the simulations of Przeworski et al. (1999) and Williamson and Orive (2002).

The methods which we have developed in this paper suggest several avenues for future research. First, to what extent does selection distort the shape of the genealogy, rather than just changing its depth? An understanding of such distortions is necessary if we are to be able to distinguish the effects of different kinds of selection from their effects on neutral variability. Second, analytical approximations can be developed for strong selection and for large genealogies, which may allow a better understanding of the joint effects of multiple loci on genetic variability. Finally, it may be possible to develop the idea that selection can act on samples of genes, and their genealogical relationships, in the same way that it acts on individual genes and their phenotypes. This suggest an intriguing link between the separate literatures on the stochastic evolution of genealogical relationships, and on the deterministic evolution of groups of related individuals.

ACKNOWLEDGEMENTS

We are grateful to the Erwin Schrödinger International Institute for Mathematical Physics in Vienna for support during winter 2002/2003, where this manuscript was completed. We also acknowledge support from the EPSRC and BBSRC (grant MMI09726).

LITERATURE CITED


# TABLES

Table 1. Summary of notation.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>effective population size</td>
</tr>
<tr>
<td>m</td>
<td>sum of mutation rates, $m_{Q+P} + m_{P+Q}$</td>
</tr>
<tr>
<td>p</td>
<td>equilibrium under mutation, $m_{Q+P} m_{P+Q}$</td>
</tr>
<tr>
<td>s</td>
<td>selection favouring allele P</td>
</tr>
<tr>
<td>s_b</td>
<td>strength of balancing selection</td>
</tr>
<tr>
<td>p_0</td>
<td>rate of mutation to new neutral alleles</td>
</tr>
<tr>
<td>r</td>
<td>recombination rate</td>
</tr>
<tr>
<td>f_{j,k}</td>
<td>identity in allelic state amongst j genes of type Q, k of type P</td>
</tr>
<tr>
<td>J_{j,k}</td>
<td>expected total length of a genealogy relating j, k genes of type Q, P</td>
</tr>
<tr>
<td>F_{0,2}, F_{1,1}, F_{2,0}</td>
<td>distribution of pairwise coalescence times</td>
</tr>
<tr>
<td>t_{2,0}, t_{1,1}, t_{0,2}</td>
<td>mean pairwise coalescence time</td>
</tr>
<tr>
<td>E@D</td>
<td>expectation over the stationary density</td>
</tr>
<tr>
<td>n</td>
<td>number of genes in the sample</td>
</tr>
<tr>
<td>L</td>
<td>differential operator, $2iUtp - pL^t + SpqL^t - \frac{pq}{2} \frac{q}{p}$</td>
</tr>
</tbody>
</table>

- T: scaled time, $t = 2N$
- U: scaled mutation rate, $Nm$
- S: scaled purifying selection, $Ns$
- S_b: scaled balancing selection, $Ns_b$
- V: scaled neutral mutation rate, $N_n$
- R: scaled recombination rate, $Nr$
- f: average identity, $\frac{1}{j} \sum_{i=0}^{j-k} q_{j-k}^i p_{j-k}^i j + k N_{j,k}$
- J: expected total length, averaged over samples
- J: total length; expected value is J
- t: mean pairwise coalescence time; $2 \tau = J$
- F: distribution, averaged over samples
- t: mean pairwise coalescence time; $2 \tau = J$
- D_n: half the regression of identity on allele frequency in the sample
- D_n^t: the same, but for total length
- q_i: difference in identity within vs between classes, $f_{PP} - 2f_{PQ} + f_{QQ}$
FIGURE CAPTIONS

Figure 1. Comparison between three methods for calculating identities as functions of allele frequency. The dots show identities calculated by simulation of a neutral allele over 50,000 generations, and by solution of a matrix recursion; the values are barely distinguishable on this scale. The thin solid lines show the identities calculated using the diffusion approximation, Eqs. 1. The thick curve shows the stationary distribution. (This distribution is divided by 4 for clarity; the probability of being in the range 0.1<p<0.9 is 0.59). 2N = 100, n = m = 0.005, p = 0.5; thus, U = V = 0.25.

Figure 2. An example of the time-course of identities (top panel), allele frequencies (middle panel) and average identity (bottom panel), from simulations of a neutral allele (2N = 100, n = m = 0.005, p = 0.5, as in Fig. 1). In the top panel, the lower thick line shows the identity f_{1,1} between allelic classes, over generations 2000 to 4000. The upper two lines show the within-class identities f_{2,0}, f_{0,2}. Around generation 3000, allele P is lost (middle panel); then the identity f_{0,2} is set to 1, and the identities f_{1,1}, f_{2,0} become equal. The converse pattern is seen when allele Q is lost, around generations 2000 and 4000. Although the identities within and between classes fluctuate greatly with allele frequency, the average identity stays close to the expected value f = \frac{1}{1+4Nn} = 0.5.

Figure 3. Comparison between three methods for calculating identities as functions of allele frequency, under balancing selection. Parameters are as in Fig. 1, except that there is balancing selection with coefficient s = p_0 - p_L, s = 0.08, p_0 = 0.7. The stationary distribution is now concentrated around p_0; the probability of 0.1<p<0.9 is increased to 0.74.

Figure 4. The distribution of coalescence times, calculated from Eqs. 3, for a locus under balancing selection. The three panels show distributions at p = 0.01, 0.1, 0.5 (left to right). In each, the thick line shows the between-class distribution F_{1,1}, the dashed curve shows F_{2,0}, and the thin solid curve shows F_{0,2} as in Fig. 3). Time is scaled relative to 2N generations.

Figure 5. The distribution of coalescence times between randomly chosen pairs of genes (thick line) compared with the neutral expectation \text{Exp}[\frac{T}{2}] (thin line). Parameters are for a locus under balancing selection, as for Figs. 3, 4.

Figure 6. Mean coalescence time within and between allelic classes, plotted against allele frequency. The thin lines are for neutral alleles, whilst the thick lines are for balancing selection S = 4, p_0 = 0.7; U=0.25 as before. The upper pair of curves are for genes in different allelic classes \text{Exp}t, the lower two pairs are for mean coalescence time within classes (dashed curves, \text{Exp}t; solid curves \text{Exp}t). Mean coalescence time between two P genes, \text{Exp}t, decreases to \frac{4Uq}{1+4Uq} as p \rightarrow 0; conversely, \text{Exp}t increases to \frac{4Uq}{1+4Uq} as p \rightarrow 1.

Figure 7. Transformed representation of the mean coalescence time, for increasing strengths of balancing selection (U = 0.25, p_0 = 0.7). Top left: \text{Exp}t, the average over randomly chosen pairs of genes. Bottom left: \text{Exp}t, the effect on mean coalescence time of association with P rather than Q. Top right: \text{Exp}t, twice the difference in mean coalescence time within relative to between classes. Bottom right: y[p], the stationary distribution. The thick curves are for neutral alleles, and the successive thin curves are for S=1, 2, 4, 8, 16, 32.

Figure 8. Transformed representation of the mean coalescence time, for increasing strengths of purifying selection; parameters and notation are as for Fig. 7. The thick curves are for neutral alleles, and the successive thin curves are for S = 0.25, 0.5, 1, 2, 4, 8.

Figure 9. The perturbations to total expected tree length caused by purifying selection, plotted against the number of extant genealogies, j (lower series of points). These are calculated from \text{Exp}t p \text{Dj} (Eq. 16). The upper series of points shows the contribution to total expected tree length expected under neutrality, \text{Exp}t. Mutation rate is U = 0.5; p_0 = 0.5, and purifying selection S = 2; there is no recombination.

Figure 10. The effect of balancing selection on the mean coalescence time \text{Exp}t. The strength of balancing selection, S_b, increases to the right; negative values correspond to disruptive selection; p_0 = 0.7. The two
curves are for mutation rate $U = 0.5$ (thick curve) and 0.25 (thin curve). There is no recombination ($R = 0$). The straight lines are the predictions assuming that allele frequency is fixed at $p_0$.

Figure 11. The mean coalescence time, $\tau$, as a function of allele frequency; $U = 0.5$, $p = 0.5$, $R = 0$, $S_b = 32$, $p_0 = 0.7$. The thin solid curve gives the exact solution from Eqs. 5. This is compared with the prediction from Eq. 15 (dashed line), which assumes that allele frequency is fixed; this prediction is independent of selection. The thick curve shows the stationary distribution.

Figure 12. The effect of purifying selection on the mean coalescence time $E[\tau]$. The two curves are for equilibrium frequency $U = \frac{1}{8}$ (thick) and $\frac{1}{4}$ (thin). There is no recombination ($R = 0$). The lines on the right are the predictions assuming that allele frequency is fixed at $U = \frac{1}{S}$.

Figure 13. The mean coalescence time, $\tau$, as a function of allele frequency, at mutation-selection balance; $U = \frac{3}{8}$, $p = 0.001$, $R = 0$, $S = 3$. The thin solid curve gives the exact solution from Eqs. 5. This is compared with the prediction from Eq. 15, which assumes that allele frequency is fixed; this prediction is independent of selection. The thick curve shows the stationary distribution.

Figure 14. The effect of balancing selection on mean coalescence time, plotted against recombination rate, $R$; $U = 0.05$. The vertical axis shows increases over the neutral value, $1 - E[\tau]$, on a logarithmic scale. The upper thick curve shows the deterministic limit, in which allele frequency is assumed to be fixed at $p_0 = 0.7$ (Eq. 15). The thin solid curves are for $S_b = 4, 8, 16, 32$ (bottom to top), calculated using Eqs. 1 and Eq. 11. The dashed curves show the high recombination limit (Eq. 19).

Figure 15. The effect of purifying selection on mean coalescence time, plotted against recombination rate, $R$; $U = 0.5$, $p = 0.5$. The vertical axis shows decreases from the neutral value, $1 - E[\tau]$, on a logarithmic scale. The upper thick curve shows the deterministic limit for $S = 8$, in which allele frequency is assumed to be fixed at $p_0 = 1 - \frac{U}{S}$ (Eq. 15). The thin solid curves are for $S = 0.5, 1, 2, 4, 8$ (bottom to top at right of figure), calculated using Eqs. 1 and Eq. 11. The dashed curves show the high recombination limit (Eq. 19).

Figure 16. The effect of purifying selection on mean coalescence time, plotted against selection, $S$, for $U = 0.25$, 0.5, 1; $p = 0.5$. There is complete linkage $R = 0$.

Figure 17. The expected total length of the genealogy which connects a sample of five genes, plotted as a function of allele frequency; the six cases $J_{0.5}$ to $J_{5,0}$ are shown; $J_{0.5}$ increases most steeply from left to right, $J_{5,0}$ shows the opposite gradient, and the other four variables interpolate between these. There is purifying selection against $Q$ alleles of strength $S = 2$; mutation is at rate $U = 0.5$ with $p = 0.5$, and there is no recombination.

Figure 18. The effect of purifying selection, $S$, on the expected total length of a genealogy, $E[\ell]$. The graph shows the reduction in $E[\ell]$ below the neutral value, $\sum_{j=1}^{n-1} \frac{1}{j}$. The upper large dots are for $n = 50$ genes; the lower series of smaller dots are for $n = 2, 3, 4, 5$ genes. For comparison, $E[\ell] = 8.96$ for $n = 50$ genes, and 2, 3, 3.67, 4.17... for 2, 3, 4, 5... genes. The curves on the right are based on a reduction in effective population size by $(1-U/S)$; this prediction applies when $S$ is large. Mutation rate is $U = 0.5$, with $p = 0.5$. 

Stochastic hitchhiking
FIGURES

Figure 1
Figure 2

Stochastic hitchhiking
| 3000 | 4000 |
Stochastic hitchhiking
Figure 4

\[
F_{\sigma T_D} \quad p = 0.01
\]

\[
F_{\sigma T_D} \quad p = 0.3
\]

\[
F_{\sigma T_D} \quad p = 0.9
\]

\[
F_{\sigma T_D} \quad p = 0.1
\]

\[
F_{\sigma T_D} \quad p = 0.7
\]

\[
F_{\sigma T_D} \quad p = 0.99
\]
Figure 5

Figure 5
Figure 6
Figure 7

Stochastic hitchhiking
Figure 8
Figure 9

Stochastic hitchhiking
Figure 10
Figure 11

Stochastic hitchhiking
Figure 12

Selection in fluctuating backgrounds
Figure 13

Stochastic hitchhiking
Figure 14

The figure shows a graph with the y-axis labeled $E_{tD-1}$ and the x-axis labeled $R$. The graph plots different curves, each corresponding to a different value of $t$ or $D$. The y-axis ranges from 0.0001 to 0.1, and the x-axis ranges from 2 to 16.

The y-axis labels are 0.0001, 0.001, 0.01, 0.1, and the x-axis labels are 2, 4, 6, 8, 10, 12, 14, 16.
Figure 15

![Graph 1](image1)

Figure 16

![Graph 2](image2)

Stochastic hitchhiking