J Mol Evol (1990) 31:511–523

Journal of Molecular Evolution

A Maximum Likelihood Approach to the Detection of Selection from a Phylogeny

Brian Golding[1] and Joe Felsenstein[2]

[1] Department of Biology, York University, 4700 Keele Street, North York, Ontario M3J 1P3, Canada
[2] Department of Genetics, University of Washington, Seattle, Washington 98195, USA

Summary. A large amount of information is contained within the phylogenetic relationships between species. In addition to their branching patterns it is also possible to examine other aspects of the biology of the species. The influence that deleterious selection might have is determined here. The likelihood of different phylogenies in the presence of selection is explored to determine the properties of such a likelihood surface. The calculation of likelihoods for a phylogeny in the presence and absence of selection, permits the application of a likelihood ratio test to search for selection. It is shown that even a single selected site can have a strong effect on the likelihood. The method is illustrated with an example from *Drosophila melanogaster* and suggests that deleterious selection may be acting on transposable elements.

Key words: Maximum likelihood—Natural selection—Phylogeny

Introduction

The neutral theory of molecular evolution (Kimura 1983) has had a major impact on our concepts about evolution. The theory suggests that most molecular polymorphisms are influenced more by random drift than by selection. Simple expectations can be constructed to predict the pattern of molecular genetic variation that should be observed when selection is absent. Many of these predictions of the neutral theory have been found to be reasonably accurate, whereas others are still controversial.

The distribution of genetic variation within populations has been used to support the neutral theory (Nei 1987). Several features of these data such as the excess of rare alleles led Ohta (1973) to suggest that the neutral theory should be modified to incorporate the presence of small levels of deleterious selection acting on allozyme polymorphisms. The changes proposed by Ohta (e.g., 1976) brought the theoretical results into much closer agreement with the observed data.

Another prediction of the original neutral theory is that if neutral mutation rates are constant over time then the mean rate of molecular evolution should also be constant. This has been found to be approximately correct for the coding sequences of many genes (Wilson et al. 1977). However, the rates of evolution of some genes are not constant and a closer examination of those that do show constant mean rates reveals that they have much larger variances than expected (Ohta and Kimura 1971; Langley and Fitch 1974; Hudson 1983). Rates appear to be partly influenced by generation time (Maeda et al. 1988; Koop et al. 1986) but this effect cannot explain the higher than expected variance (Gillespie 1989).

As more molecular sequences are determined, the rates with which sequences change are becoming known for many genes. Gillespie (1984a, 1986) has demonstrated that the rates of evolution for these genes may follow an episodic pattern. He suggests that the best way to explain this pattern of evolution is to invoke selection acting on these sequence alterations.

Thus, both within-population variation and between-population variation suggest that at least some

level of selection is a required consideration for studies of molecular evolution. However, very little progress has been made in attempts to extend Ohta's advanced theories because of the difficulties associated with incorporating selection into theoretical models. Gillespie (1984b) has been able to include the effects of strong selection in his models of change between species using an asymptotic analysis assuming weak mutation and strong selection.

A major difficulty with studies of natural selection is the immense power of this natural force. Even selection coefficients that are of the same size as mutation rates can have significant effects when the geological time periods over which species have diverged are considered. Estimating selection coefficients of this size and distinguishing them from mutation rates is difficult. The best method to attempt estimation is experimental verification but this is almost completely ruled out by the potentially small size of the coefficients. It is possible, however, to examine sequences that have been influenced by selection for millions of years and to determine if their changes might suggest the action of selection. This was the approach of Sawyer et al. (1987) when they inferred selection coefficients on the order of 10^{-7} from the distribution of segregating amino acid polymorphisms for the 6-phosphogluconate dehydrogenase gene in *Escherichia coli*. The results of Hartl et al. (1985) suggest that such small selection coefficients may in fact be very common and are a result of the saturation kinetics of enzymes with flux that is a concave function of activity. They suggest that many enzymes have these kinetics and that evolution therefore leads to very small selective differences between most segregating alleles.

Another major problem associated with studies of natural selection is the lack of an accurate concept of what the effects of selection should be. For example, Hill and Hastie (1987) found a gene where the first and second codon positions changed more rapidly than the third position in a group of serine protease inhibitors. This unusual result was interpreted as being due to positive selection. However, Graur and Li (1989) suggest that the frequency distribution of amino acids was unusually skewed and that this accounts for the peculiar pattern of substitutions. An impartial method for detecting the effects of selection is clearly desirable. Furthermore, in addition to simply stating whether or not selection has influenced the genes under study, there should be some method to estimate the actual values of the selection coefficients and to test their statistical significance.

The large amounts of sequence data that are accumulating for many genes from many species permit the reconstruction of phylogenies that relate the ancestries of these sequences. The patterns of sequence change in these phylogenies are significantly altered when some of the characters in these sequences are influenced by selection (Golding et al. 1986). These differences can be used in tests to detect the presence of selection.

One way to use these differences is to examine the most recent distinguishable ancestor of each extant sequence (Golding et al. 1986). This however was done with a deterministic model that does not include the effects of random drift (see Iizuka 1989) and ignored the problems of inference from a phylogeny. Monte Carlo simulations indicate that extensive sequence data are required to detect selection with a reasonable degree of certainty (Golding 1987). Generally, sampling must be sufficient to determine 30–40 distinct haplotypes. Because the collection of sequence data is often difficult and expensive, it is desirable to have a method that extracts more information from the data.

A maximum likelihood approach is developed here in an attempt to gain greater statistical sensitivity. This approach has the advantage that it is based on a fully specified model with all assumptions apparent. There are also standard statistical tests that have been developed making use of alternate likelihoods. It is thus possible with this type of an approach to detect even very weak selection with statistical reliability. Furthermore, the shape of the likelihood surface contains useful information about the variance of estimates.

Method

Consider sequences of sites that can have only one of two possible allelic states. This is appropriate for the presence/absence of restriction sites or the presence/absence of any other feature such as a deletion or an insertion. Consider these sites to be selectively neutral and let the characters spontaneously mutate at a rate υ per gamete per generation. For simplicity, let the mutation rates to and from each state be equal. Generations are assumed to be discrete and the evolution of each site is assumed to be independent of all other sites.

The probability that a site initially in state i will change to state j within time t is P^t_{ij}. The value of P^t_{ij} can be found easily if the site is selectively neutral. It is given by

$$\begin{aligned} P^t_{ij} &= \tfrac{1}{2} + \tfrac{1}{2}e^{-2\upsilon t} && \text{if } i = j \\ &= \tfrac{1}{2} - \tfrac{1}{2}e^{-2\upsilon t} && \text{if } i \neq j \end{aligned}$$

Here υ and t are measured in the same units (usually years or generations) and it is assumed that mutations occur according to a Poisson process.

The likelihood of part of an evolutionary tree subtended by the kth node for a site in state i is designated as $L^{(k)}_i$. This likelihood can be calculated in a recursive fashion. As an example, consider the tree given in Fig. 1. The tree is assumed to be strictly bifurcating with the node or taxa 3 of the tree bifurcating to give descendant nodes or taxa 1 and 2. The time between node 3 and node 1 is t and between nodes 3 and 2 is t'. With these definitions the likelihood of an evolutionary tree can be found recursively from

$$L^{(3)}_i = \left[\sum_{j=1}^{2} P^t_{ij} L^{(1)}_j\right]\left[\sum_{k=1}^{2} P^{t'}_{ik} L^{(2)}_k\right]$$

(Felsenstein 1981). The terms $L^{(1)}_j$, $L^{(2)}_k$ designate the likelihoods of part of the evolutionary tree subtended by nodes 1 and 2, for sites with states j and k. If nodes 1 and 2 designate extant species then these likelihoods are known explicitly. The likelihoods are either 1 or 0 depending on whether the extant species does or does not have that state (allele) at that particular site. This information can be used to determine $L^{(3)}_i$. In more complicated phylogenies with more than two species the likelihoods of interior nodes can be calculated in a similar fashion. In this case identify node 3 with a more ancient bifurcation and nodes 1 and 2 with bifurcations that in turn give rise to more species. Begin at the tips of the phylogeny and move down the tree one node at a time. Each successive step uses the likelihoods just calculated such that the value determined for $L^{(3)}_i$ is used to find the likelihood of the next node. The likelihood of every subtree of every state is calculated for every node using those likelihoods calculated for the previous nodes. This continues until the root of the tree is reached and then the overall likelihood is found by summing the products of the root likelihoods with the prior probabilities of each state. Without any further information, the prior probabilities of each state are usually taken to be their equilibrium frequencies.

Because each site evolves independently, the likelihood of a phylogeny can be calculated separately for each site. The product of the likelihoods for each site provides the overall likelihood of the observed data.

Transition Probabilities with Selection. We would like to modify this method and to find values for P^t_{ij} when selection influences sites within a sequence. This cannot be done in complete generality and instead will be approached using two approximations. First the overall selection model is described, then a method to calculate these probabilities for a single selected site is given assuming that selection is weak, then a method to calculate probabilities is given, which assumes that mutation is weak relative to selection.

This type of algorithm requires that the individual taxa evolve independently after speciation. When selection occurs this is true only of trees consisting of different species. It is not true of individual, selected haplotypes within a species because selection will alter the a priori probabilities of some phylogenies. We will calculate likelihoods assuming that the structure of the phylogeny itself does not contain information about the strength of selection. We require and will assume that the individual lineages have evolved independently after speciation. This implies that a fork in the tree is coincidental with speciation and hence that the effective population size is much less than the divergence time ($N_e \ll t$) so that lineages coalesce rapidly when they are traced back within an ancestor. If polymorphism exists at the time of speciation, daughter species are formed from a randomly sampled haplotype. Note that the descendants must be contemporary in this model (assuming that selection has had equal opportunity to act in each taxa) and hence a molecular clock is assumed.

In addition to the neutral sites previously described, a second class of sites may also be present in the sequences and these sites may be influenced by selection. Assume that these sites can have one of two states, which are designated *A* and *a*. The frequency of allele *A* is designated by x. When individuals carry a copy of the deleterious allele, *a*, their fitness is reduced by a proportion 1 − s. Genic selection is assumed and so the results for either a haploid or a diploid model are equal. Thus the fitnesses are either

Haploid	*A*		*a*
	1		1 − s
Diploid	*AA*	*Aa*	*aa*
	1	1 − s	$(1 - s)^2$

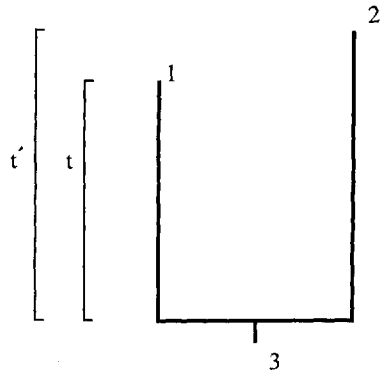

Fig. 1. A single taxon (3) bifurcates (speciates) to yield two new taxa (1, 2). The taxa 1 and 3 are separated by t generations, whereas taxa 2 and 3 are separated by t' generations.

Each character has the same selection coefficient and fitnesses are multiplicative. Note that no restriction is placed on the sign of s and so advantageous alleles can also be modelled with this method. These sites spontaneously mutate at a rate μ per gamete per generation.

To calculate P^t_{ij} consider the expected allele frequencies in a finite population. Given some initial conditions, the expected frequencies can be calculated for different time periods of evolution. The changes in these expectations are then identified with the corresponding probabilities of transition.

A formula for these expectations when selection is acting in a finite population was first illustrated by Kimura (1955) and again by Avery (1978). Consider a population that has only small changes in its allele frequency x from one generation to the next. If this small change is designated as δx, then the expected nth power of the allele frequency in the next generation is

$$E[(x + \delta x)^n] = E[x^n] + nE[x^{n-1}E_g(\delta x | x)] + n(n-1)/2E[x^{n-2}E_g(\delta x^2 | x)] + \ldots,$$

The expectations E_g are over conceptually replicate populations each with a given value of x. For the above model,

$$E_g(\delta x | x) \simeq \mu(1 - 2x) + sx(1 - x)$$
$$E_g(\delta x^2 | x) \simeq x(1 - x)/2N$$

For n = 1 and ignoring higher-order terms

$$E(x + \delta x) \simeq E(x) + \mu[1 - 2E(x)] + sE(x) - sE(x^2)$$

A similar equation for $E(x^2)$ with selection will require still another equation for $E(x^3)$, and so on. Unless s = 0, this leads to an infinite system of equations describing all of the allele frequency moments. However, it is possible to determine a series solution for E(x) in powers of (Ns). For example, consider $E(x^n)$ expanded into

$$E(x^n) = a + b(Ns) + c(Ns)^2 + d(Ns)^3 + \ldots$$

where $a = E(x^n)\Big|_{Ns=0}$ and the other terms are yet to be determined. To calculate E(x) to a first order approximation the above equation gives

$$E(x + \delta x) \simeq E(x) + \mu[1 - 2E(x)] + sE(x)\Big|_{s=0} - sE(x^2)\Big|_{s=0}$$

This formula does not lead to an infinite system but rather to a

three-dimensional system of equations that determines E(x) but ignores terms of the order of $0(Ns)^2$. This system gives the expected change in allele frequency in one generation. The expression for t generations is found by iterating these equations. The solution to this system of equations is

$$E^t(x) = e^{-2\mu t}E^0(x) + \frac{1}{2}(1 - e^{-2\mu t}) + \frac{2Ns}{1+\theta}(e^{-2\mu t} - e^{-4\mu t - t/2N})\left[E^0(x)\Big|_{Ns=0} - E^0(x^2)\Big|_{Ns=0}\right] + \frac{Ns}{1+\theta}\left[1 - e^{-2\mu t} - \frac{\theta}{1+2\theta}(1 - e^{-4\mu t - t/2N})\right]$$

where $\theta = 4N\mu$. Note that $E^t(x) = \frac{1}{2}(1 - e^{-2\mu t}) + e^{-2\mu t}E^0(x)$ when $s = 0$ and that $E^\infty(x) = \frac{1}{2}[1 + 2Ns/(1 + 2\theta)]$ as expected.

The probability of change from allele *A* to the alternate allele *a* in time t can be calculated from this formula as $P^t_{Aa} = 1 - E^t(x)$ with

$$E^0(x) = 1,\ E^0(x)\Big|_{Ns=0} = 1 \quad \text{and} \quad E^0(x^2)\Big|_{Ns=0} = 1$$

Similarly $P^t_{AA} = E^t(x)$ with the same initial conditions and $P^t_{aA} = E^t(x)$ and $P^t_{aa} = 1 - E^t(x)$ with initial conditions

$$E^0(x) = 0,\ E^0(x)\Big|_{Ns=0} = 0 \quad \text{and} \quad E^0(x^2)\Big|_{Ns=0} = 0$$

This is a first order approximation. For the algorithm that is actually implemented, a solution to E(x) is used, which is a second order approximation (the derivation of this formula is given in Appendix 1).

This approximation is only valid when Ns is small. For large Ns another approximation is used. This approximation assumes that mutation is weak relative to selection. The previous requirement that $N_e \ll t$ ensures that the time required for fixation of alternate alleles is very short relative to the length of time between fixations. Advantage can be taken of this condition by letting the fixation time approach zero relative to the time between fixations. The probability of eventual fixation for a particular advantageous allele is

$$U(s, Ns) = (1 - e^{-2s})/(1 - e^{-4Ns})$$

(Kimura 1962). Similarly the probability of eventual fixation for a deleterious allele is $U(-s, -Ns)$. There are $2N\mu$ new mutations of these alleles each generation. Using the approximation, the probability of a change in the allelic state per unit time is $a = 2N\mu\, U(s, Ns)$ or $b = 2N\mu\, U(-s, -Ns)$ depending on the direction of change. Therefore the probability of a change from *A* to *a* in t generations is

$$P^t_{Aa} = a/(a + b)[1 - e^{-(a+b)t}]$$

and

$$P^t_{aA} = b/(a + b)[1 - e^{-(a+b)t}]$$

whereas $P^t_{aa} = 1 - P^t_{aA}$ and $P^t_{AA} = 1 - P^t_{Aa}$.

The two approximations to the transitions probabilities are quite different. The series approximation is only valid with weak selection and includes the chance of polymorphism. This approximation is used when $4Ns < 0.1$. The weak mutation approximation considers the relative proportion of species fixed for alternate alleles and because mutation is weak, does not permit extended polymorphism. This approximation is more appropriate when selection is strong and is used when $4Ns > 0.1$. Both approximations require that $N_e \ll t$.

Multiple Sites within a Sequence. The next step is to combine the answers for a single neutral site with those for a single selected site and to consider the probabilities of transition between states for the sequence as a whole.

The algorithm used by Felsenstein (1981) and other maximum likelihood programs assume that each individual site within a sequence is evolving independently of the other sites. We again assume here that mutation proceeds independently at each site. However, we must be concerned that selection or random drift may generate linkage disequilibrium between each of the sites within a sequence. When all sites are neutral, it is known that random drift will generate linkage disequilibrium in finite populations. Although some disequilibrium is created in any particular population the expected value of this linkage disequilibrium is zero when these expectations are taken across conceptually replicate populations. Thus, the disequilibrium will not alter the expected allele frequency nor the expected probability of transition.

Consider a single selected site that is completely linked to a group of neutral sites. Again random drift can generate linkage disequilibrium in finite populations, but the expected value of linkage disequilibrium will be zero across replicate populations. Birky and Walsh (1988) also found via simulations and a mathematical argument that the substitution of neutral mutations is not affected by complete linkage to a selected site.

If more than one site is influenced by selection, the possibilities for the generation of linkage disequilibrium become greater. It was demonstrated by Felsenstein (1965) that even some selection parameters that do not lead to stable linkage disequilibrium may cause transient associations between alleles. He showed that this was due to epistasis between gametes. In this paper, only genic selection with strict multiplicativeness of selection coefficients will be considered. In this case neither transient nor equilibrium linkage disequilibrium is generated by selection. Thus, neither selection nor random drift generates a non-zero expected disequilibrium.

To double-check these results, and to ensure that no disequilibrium is generated, a series of equations can be derived to describe the evolution of two completely linked loci, each with two alleles. Haplotype frequencies are designated as x_{00}, x_{01}, x_{10}, and x_{11} and fitnesses of the four haplotypes are assumed to be 1, $1 - r$, $1 - s$ and $(1 - s)(1 - r)$, respectively. The first site mutates at a rate μ and the second at a rate υ. Following the same method as given above recursion equations can be derived and are given in Appendix 2. Again all parameters are considered to be small (of the order of 1/2N) and terms in $(Ns)^2$, $(Nr)^2$, and $(Ns)(Nr)$ are ignored.

The equilibrium solution of these equations is

$$x_{00} = \frac{1}{4}\left(1 + 2Nr\frac{1}{1+8N\upsilon} + 2Ns\frac{1}{1+8N\mu}\right)$$

$$x_{01} = \frac{1}{4}\left(1 - 2Nr\frac{1}{1+8N\upsilon} + 2Ns\frac{1}{1+8N\mu}\right)$$

$$x_{10} = \frac{1}{4}\left(1 + 2Nr\frac{1}{1+8N\upsilon} - 2Ns\frac{1}{1+8N\mu}\right)$$

$$x_{11} = \frac{1}{4}\left(1 - 2Nr\frac{1}{1+8N\upsilon} - 2Ns\frac{1}{1+8N\mu}\right)$$

The extension to *n* arbitrary loci is obvious. These solutions are the perfect square of the individual allele frequencies when the level of the approximation is taken into account. As expected, no stable linkage disequilibrium is generated.

To determine if transient linkage disequilibrium is generated, these equations were iterated for $t/2N = 10^{-3}$–10^2 with $2N = 10^4$. The mutation rates were chosen to be smaller than the selection coefficients so that the effects of selection would predominate. The results of this iteration are shown in Table 1 for three different initial conditions. The table demonstrates that again, no transient

Table 1. Linkage disequilibrium generated in finite populations

Generations	Initial condition		
	$x_{00} = 1$	$x_{01} = 1$	$x_{11} = 1$
$t = 2N/1000$	-1.25×10^{-10}	1.25×10^{-10}	-1.25×10^{-10}
$t = 2N/100$	-1.25×10^{-9}	1.25×10^{-9}	-1.25×10^{-9}
$t = 2N/10$	-1.23×10^{-8}	1.21×10^{-8}	-1.26×10^{-8}
$t = 2N$	-8.35×10^{-7}	-6.33×10^{-7}	-8.55×10^{-7}
$t = 10$ (2N)	-2.07×10^{-4}	-2.06×10^{-4}	-2.07×10^{-4}
$t = 100$ (2N)	-9.40×10^{-4}	-9.40×10^{-4}	-9.40×10^{-4}
$t = \infty$	-9.47×10^{-4}	-9.47×10^{-4}	-9.47×10^{-4}

The linkage disequilibrium (as measured by $D = x_{00} - [(x_{00} + x_{01})(x_{00} + x_{10})]$ that is generated by random drift in a finite population for two completely linked loci each with two alleles in the presence of deleterious genic selection and mutation. Fitnesses are assumed to be multiplicative. The population size is $2N = 10^4$ with $4N\nu = 0.05$, $4N\mu = 0.1$, $4Nr = 0.1$, and $4Ns = 0.2$

linkage disequilibrium is generated. The values in this table are not exactly zero because higher order selection terms are ignored.

These results indicate that the transition probabilities can be calculated treating each site independently (as if it were in linkage equilibrium). In the presence of selection any of the approximations can be used to calculate the probabilities of transitions for individual sites. The overall likelihood is then the product of likelihoods from each site.

Maximizing the Likelihood. The algorithm given in Felsenstein (1981) can be used to calculate the maximum likelihood with the transition probabilities altered to those given here. For the purposes of illustration however, the topology of the trees is assumed to be given. It is not usually necessary to determine the likelihood of alternative topologies because the null hypothesis in any test for selection is that $Ns = 0.0$. For this situation many excellent algorithms are available (reviewed in Felsenstein 1988).

Given a tree topology, the individual branch lengths are altered to maximize the overall likelihood of the tree with and without selection. The phylogenies must be rooted, as each lineage must have had equal time periods for selection and mutation to occur. After each branch length has been individually altered, the difference is used as an approximation of the gradient and an attempt is made to improve the likelihood by simultaneously altering all branch lengths in the direction of the gradient. After this, all branch lengths are again individually maximized and a new gradient is determined. This is continued until no further improvement can be made in the likelihood (or until a preset number of iterations have been attempted). Note that the maximum likelihood found is only a local maximum and higher maxima may exist elsewhere on the surface [however, Fukami and Tateno (1989) have proved that for a simple model only one stationary point exists].

Results

Examples

If species are very distantly related then this algorithm is equivalent to determining the probability of selection simply on the basis of the frequency of deleterious characters. Unrelated species will contain deleterious or neutral characters according to their equilibrium frequencies. Departures from these equilibrium frequencies can be used to look for selection. As a character becomes rare, stronger selection against that character is suggested. This algorithm was applied to distantly related species to determine the strength of this effect. It was found that if 1 of just 8 unrelated species carries a deleterious character then a likelihood ratio test would suggest that there is evidence for deleterious selection. If 2 species carry the deleterious character then the number of species must be 11 or more for selection to be supported.

Table 2. Example sequences that have the phylogeny given in Fig. 2

Species	Sequence	Species	Sequence
A	0000001110	K	1100000001
B	0001001100	L	1100001000
C	0111001100	M	0010111100
D	1111000100	N	0010001100
E	0000010000	O	0000000100
F	0001010010	P	0010010100
G	0000010000	Q	0100111100
H	0001000000	R	0001111100
I	0100000000	S	1000011000
J	0100001001	T	1000010100

A simple example of a typical phylogeny has been chosen to illustrate the nature of the likelihood surface. This hypothetical example has been constructed using random data. Data are generated for 20 species each with sequences consisting of 10 sites. The tree is assumed to be given, to be strictly bifurcating, and each branch is of a constant length (except for the three branches leading to the first four species, which are of double length to insure that all species are contemporary). A random number generator is used to randomly alter a single site along each branch (two along the double length branches). The resulting sequences are given in Table 2.

The branch lengths for this phylogeny were adjusted so as to maximize the likelihood for all 20 sequences (Fig. 2). This was done assuming that all sites were selectively neutral and that the mutation

Table 3. The $\log_e$ likelihood of the example phylogeny given in Fig. 2

	4Ns							
	Weak selection				Weak mutation			
$4N\mu$	10^{-4}	10^{-3}	10^{-2}	10^{-1}	10^{-1}	10^{0}	10^{+1}	10^{+2}
a) A single deleterious character carried by a deeply branched species (J)								
10^{-4}	−7.1678	−7.1677	−7.1674	−7.1658	−7.1673	−7.3109	−14.1341	−101.8894
10^{-3}	−5.2729	−5.2727	−5.2707	−5.2522	−5.2529	−5.2515	−11.8920	−100.0931
10^{-2}	−6.6957	−6.6938	−6.6749	−6.4902	−6.4807	−5.0456	−10.0978	−99.9950
10^{-1}	−13.7457	−13.7391	−13.6732	−13.0298	−12.8835	−7.2377	−10.0004	−99.9950
10^{0}	−13.8626	−13.8599	−13.8330	−13.5657	−12.9880	−7.2655	−10.0004	−99.9950
b) A single deleterious character carried by a recently diverged species (K)								
10^{-4}	−7.6916	−7.6916	−7.6912	−7.6885	−7.6911	−7.8341	−14.6490	−102.3336
10^{-3}	−5.7909	−5.7907	−5.7885	−5.7684	−5.7703	−5.7627	−12.3360	−100.2110
10^{-2}	−7.1618	−7.1598	−7.1404	−6.9504	−6.9420	−5.4645	−10.2152	−99.9950
10^{-1}	−13.7863	−13.7797	−13.7136	−13.0676	−12.9221	−7.2584	−10.0004	−99.9950
10^{0}	−13.8626	−13.8599	−13.8330	−13.5657	−12.9880	−7.2655	−10.0004	−99.9940

Columns on the left ($10^{-4} < 4Ns < 10^{-1}$) are calculated using an approximation that assumes selection is weak, whereas the columns on the right ($10^{-1} < 4Ns < 10^{+2}$) are calculated using an approximation that assumes mutation is weak compared to selection

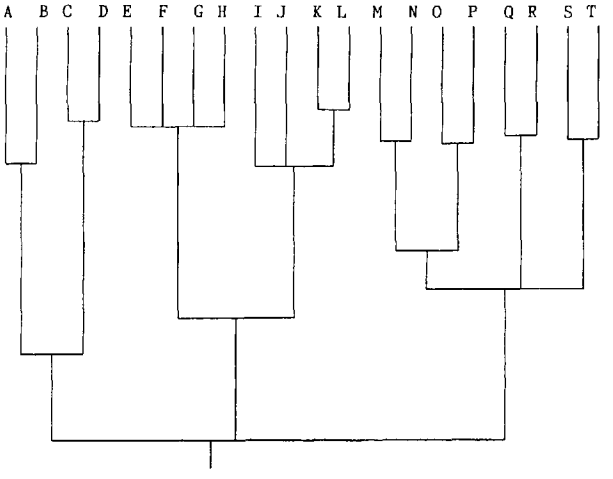

Fig. 2. An example phylogeny with 20 species. The sequences for each species are given in Table 2. The branch lengths were optimized to give the maximum likelihood for the given phylogeny (see text).

rate for these neutral characters is $4N\mu = 10^{-2}$ per site per generation with $N = 10^4$.

Ideally, data should be analyzed using all sites (including those potentially under selection). Trees should be constructed by adding one species at a time and testing all possible rearrangements of individual species to maximize the likelihood. Computational considerations have forced a less ambitious approach. The tree is assumed to be given by the pattern of neutral characters (except for the branch lengths) and the changes in the likelihood are considered as selection is introduced.

The tree in Fig. 2 was used to investigate what would happen if there were an 11th site that was under selection imbedded within this phylogeny. If only one species carried the deleterious character at an 11th site, then this species may be chosen in 20 different ways. The resulting likelihoods for the single selected site are shown in Table 3a and b if this single species is chosen to be the species with the largest (species J) or shortest branch length (species K), respectively. In both cases the likelihood slowly increases with increased levels of weak selection. As selection increases the likelihood continues to increase up to a maximum value. Further increases in selection intensity cause the likelihood to decrease again. This is because the deleterious selection has become so strong that observing even a single deleterious character is unexpected. The difference between the maximum likelihood with selection and the likelihood in the absence of selection is large. A likelihood ratio test yields a ratio of LR = 670 and 684 (for Table 3a and b for $4N\mu = 10^{-1}$) and LR = 733 and 733 (for Table 3a and b for $4N\mu = 10^{0}$). Twice the negative logarithm of the likelihood ratio should be asymptotically distributed as a chi square with 1 degree of freedom. By this test the chi squares are 13.02, 13.06, 13.19, and 13.19, respectively. Thus in all these situations there is evidence that deleterious selection would be operating. With smaller mutation rates the evidence is suggestive but not significant.

Two species may carry deleterious characters. For this case the likelihood was calculated assuming that species I and J (adjacent species in the phylogeny) carried the deleterious character (Table 4a), that species J and K (closely related species) carried the deleterious character (Table 4b), or that species J and M (distantly related species) carried the deleterious character (Table 4c). Table 4 again indicates that selection is warranted if the mutation rate of the selected character is large but the level of support is much less than that present in Table 3. The differences between Table 4a, b, and c suggest there is

Table 4. The $\log_e$ likelihood of the example phylogeny given in Fig. 2

	4Ns							
	Weak selection				Weak mutation			
$4N\mu$	10^{-4}	10^{-3}	10^{-2}	10^{-1}	10^{-1}	10^{0}	10^{+1}	10^{+2}
a) Two adjacent species with deleterious characters (I, J)								
10^{-4}	−13.2268	−13.2270	−13.2284	−13.2435	−13.2448	−13.4865	−19.0730	−104.6357
10^{-3}	−9.0207	−9.0207	−9.0204	−9.0191	−9.0197	−9.1219	−14.6363	−103.6225
10^{-2}	−8.0650	−8.0634	−8.0469	−7.8869	−7.8779	−6.6598	−13.6322	−132.1971
10^{-1}	−13.7488	−13.7429	−13.6845	−13.1155	−12.9860	−8.2367	−19.9999	−199.9900
10^{0}	−13.8627	−13.8603	−13.8363	−13.5990	−13.0880	−8.2654	−19.9999	−199.9900
b) Two closely related species with deleterious characters (J, K)								
10^{-4}	−14.1189	−14.1193	−14.1232	−14.1643	−14.1685	−14.7976	−22.8486	−106.2993
10^{-3}	−9.8996	−9.8998	−9.9018	−9.9240	−9.9274	−10.3668	−16.2977	−103.9431
10^{-2}	−8.8088	−8.8072	−8.7912	−8.6351	−8.6281	−7.4576	−13.9516	−132.1972
10^{-1}	−13.7896	−13.7837	−13.7251	−13.1535	−13.0249	−8.2577	−19.9999	−199.9900
10^{0}	−13.8627	−13.8603	−13.8363	−13.5990	−13.0880	−8.2654	−19.9999	−199.9900
c) Two distantly related species with deleterious characters (J, M)								
10^{-4}	−13.7955	−13.7959	−13.7999	−13.8419	−13.8453	−14.4798	−28.4542	−116.6893
10^{-3}	−9.5866	−9.5868	−9.5891	−9.6143	−9.6170	−10.1061	−23.8176	−110.8945
10^{-2}	−8.6019	−8.6004	−8.5860	−8.4465	−8.4399	−7.5177	−19.7914	−195.4306
10^{-1}	−13.7560	−13.7502	−13.6918	−13.1224	−12.9933	−8.2422	−19.9999	−199.9900
10^{0}	−13.8627	−13.8603	−13.8363	−13.5990	−13.0880	−8.2654	−19.9999	−199.9900

Columns on the left ($10^{-4} < 4Ns < 10^{-1}$) are calculated using an approximation that assumes selection is weak, whereas the columns on the right ($10^{-1} < 4Ns < 10^{+2}$) are calculated using an approximation that assumes mutation is weak compared to selection

Table 5. The $\log_e$ likelihood for 20 species when 1 (M) carries a deleterious character

	4Ns							
	Weak selection				Weak mutation			
$4N\mu$	10^{-4}	10^{-3}	10^{-2}	10^{-1}	10^{-1}	10^{0}	10^{+1}	10^{+2}
10^{-4}	−12.1419	−12.1414	−12.1374	−12.0981	−12.1435	−12.3031	−19.1455	−106.8404
10^{-3}	−9.8432	−9.8428	−9.8388	−9.7995	−9.8447	−10.0028	−16.8427	−104.5351
10^{-2}	−7.5802	−7.5798	−7.5758	−7.5367	−7.5797	−7.7230	−14.5374	−102.1540
10^{-1}	−5.6602	−5.6598	−5.6559	−5.6181	−5.6400	−5.6352	−12.1562	−100.1451
10^{0}	−5.7638	−5.7634	−5.7594	−5.7202	−5.5824	−4.4437	−10.1493	−99.9950

The phylogeny was determined by the coalescent process with $N = 10^4$. Columns on the left ($10^{-4} < 4Ns < 10^{-1}$) are calculated using an approximation that assumes selection is weak, whereas the columns on the right ($10^{-1} < 4Ns < 10^{+2}$) are calculated using an approximation that assumes mutation is weak compared to selection

also an influence of the phylogenetic relationships between the deleterious characters when the mutation rate is small.

The branch lengths in this example phylogeny are rather long and it is of interest to determine what would happen if the branch lengths were much shorter. The shortest branch lengths that could possibly be expected are those that would be appropriate not for 20 species but for 20 sequences chosen from within a single population. The algorithm is not strictly appropriate for intraspecific data but the branch lengths for such data should provide a lower limit to those expected between species. To this end, the coalescent process was used to simulate a phylogeny for 20 sequences with branch lengths appropriate for samples from a single population with effective size $N_e = 10^4$ (Fig. 3). Given this phylogeny the likelihoods are given in Table 5 when 1 species of the 20 carries a deleterious character. In this case, there is little support for deleterious selection.

Restriction Sites in Drosophila

It is desirable to be able to analyze the consequences of selection when it acts on more than one site and with as many species as possible. As an example some variation revealed in *Drosophila* by restriction enzymes will be used. Although this is a between-species method and does not strictly apply to within-species variation it may be used as an illustration.

Restriction sites have been mapped around the alcohol dehydrogenase (*Adh*) gene in *Drosophila melanogaster* (Aquadro et al. 1986). A total of 48 chromosomes were isolated and analyzed by eight enzymes. During the course of the mapping, several restriction fragment length polymorphisms were de-

Table 6. The $\log_e$ likelihood of *Drosophila melanogaster* haplotypes carrying transposable elements

	4Ns							
	Weak selection				Weak mutation			
$4N\mu$	10^{-4}	10^{-3}	10^{-2}	10^{-1}	10^{-1}	10^{0}	10^{+1}	10^{+2}
a) Branch lengths are optimized for each row to maximize the likelihood. Characters other than transposable elements have a mutation rate of 10^{-7}, $N_e = 10^4$								
10^{-4}	−75.3260	−75.3259	−75.3248	−75.3249	−75.3832	−77.1037	−138.9890	−863.6617
10^{-3}	−56.3724	−56.3717	−56.3658	−56.3178	−56.3729	−57.6796	−119.4580	−828.4886
10^{-2}	−47.4171	−47.4134	−47.3766	−47.0246	−47.0409	−45.6042	−106.6875	−872.0732
10^{-1}	−54.5345	−54.5276	−54.4586	−53.7831	−53.4142	−46.6322	−107.4376	−993.2115
10^{0}	−123.9903	−123.9718	−123.7867	−121.9535	−117.8246	−79.4223	−185.8273	NA
b) Branch lengths are optimized for each row to maximize the likelihood. Characters other than transposable elements have a mutation rate of 10^{-8}, $N_e = 10^4$								
10^{-4}	−55.9817	−55.9815	−55.9798	−55.9764	−55.9819	−57.2863	−119.0439	−828.0234
10^{-3}	−47.4523	−47.4487	−47.4129	−47.0741	−47.0758	−45.6370	−106.6996	−871.7206
10^{-2}	−54.5343	−54.5243	−54.4251	−53.4629	−53.4143	−46.6322	−107.4376	−993.2148
10^{-1}	−120.7628	−120.7176	−120.2658	−115.8584	−114.8262	−77.6976	−177.6275	NA
10^{0}	−123.9911	−123.9725	−123.7871	−121.9511	−117.8191	−79.3750	−185.8838	NA
c) Branch lengths are optimized for each row to maximize the likelihood. Characters other than transposable elements have a mutation rate of 10^{-9}, $N_e = 10^4$								
10^{-4}	−47.4522	−47.4486	−47.4129	−47.0755	−47.0757	−45.6371	−106.6997	−871.7145
10^{-3}	−54.5312	−54.5208	−54.4174	−53.4164	−53.4115	−46.6311	−107.4360	−993.1702
10^{-2}	−61.4378	−61.4230	−61.2758	−59.8449	−59.7818	−39.3608	−110.6055	−1051.5039
10^{-1}	−123.9879	−123.9410	−123.4725	−118.9032	−117.8191	−79.3749	−185.8839	NA
10^{0}	−123.9911	−123.9725	−123.7871	−121.9510	−117.8184	−79.3701	−185.8901	NA

The likelihood is given only for the eight sites with transposable elements. The phylogeny was determined using a total of 47 sites (including the transposable elements). Columns on the left ($10^{-4} < 4Ns < 10^{-1}$) are calculated using an approximation that assumes selection is weak, whereas the columns on the right ($10^{-1} < 4Ns < 10^{+2}$) are calculated using an approximation that assumes notation is weak compared to selection. NA indicates that the likelihood is very small

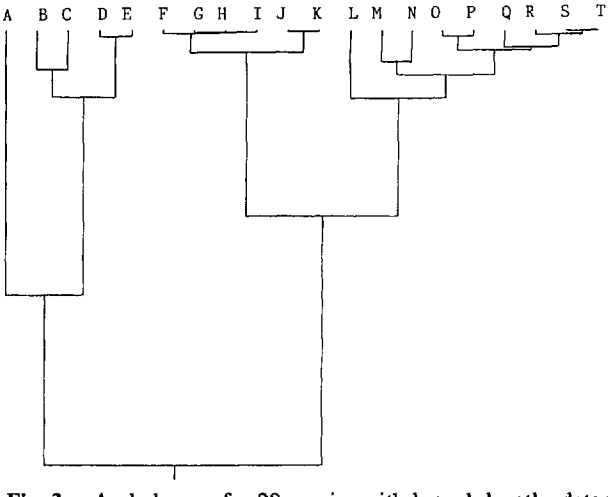

Fig. 3. A phylogeny for 20 species with branch lengths determined by the coalescent process with $N = 10^4$

tected and these were shown to be due to either insertions, deletions, or the presence/absence of transposable elements. These data provide suitable information to examine the selection coefficients that may influence transposable elements.

Of the 29 distinct haplotypes discovered, 4 appear to be potential recombinants. For the purposes of this analysis these recombinants are ignored. A phylogeny was reconstructed by maximum parsimony and rooted at the longest internal branch. The transposable elements are considered to be potentially selectively deleterious and the fast/slow polymorphism, the restriction sites, and the deletions/insertions are considered to be neutral characters. (Note that this is probably a false assumption but is adequate for the purposes of illustration.)

The likelihood surface for the haplotypes is given in Table 6. In this case, the branch lengths were maximized assuming that the neutral characters have a mutation rate of either 10^{-9}, 10^{-8}, or 10^{-7} (these were chosen to give $4N\nu = 0.004$–0.00004 with $N = 10^4$, values that would be appropriate at the molecular level). For each of these, the mutation rates of the transposable elements were assumed to range between $4N\mu = 10^{-4}$ to 10^{0} ($\mu = 2.5 \times 10^{-9}$ to 2.5×10^{-5}). For every combination of mutation rates, the branch lengths were maximized but the topology was again fixed as that determined by maximum parsimony. The resulting trees were then used to analyze what happens with selection assuming that both the topology and branch lengths are constant for given mutation rates.

For each of the three neutral mutation rates in Table 6, the results are similar. There is a strong suggestion of the presence of deleterious selection against transposable elements in this data set if their mutation rate is large. This table gives the natural

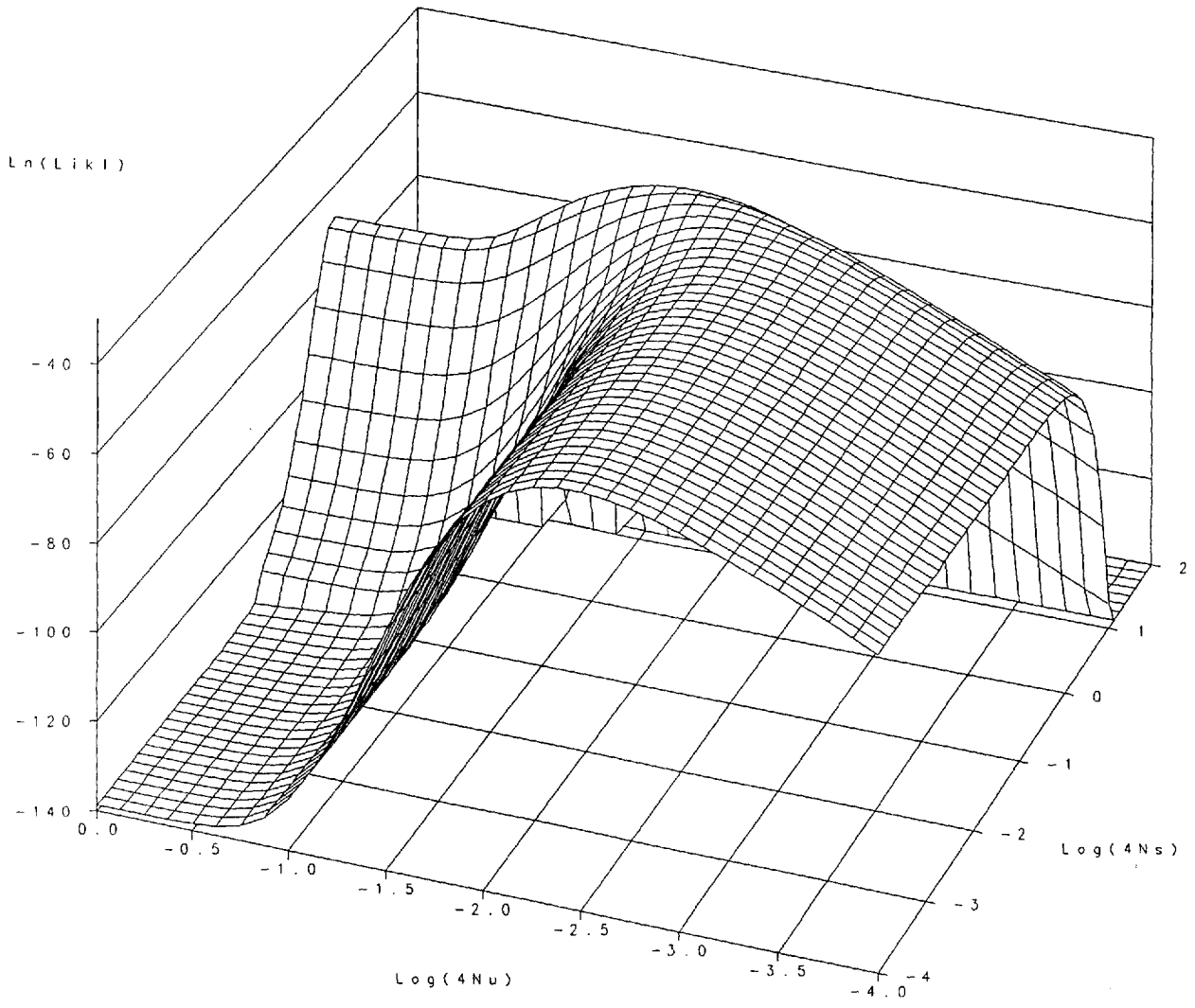

Fig. 4. The $\log_e$likelihood surface for transposable elements in *Drosophila*. The branch lengths are fixed and optimize the likelihood when $N_e = 10^4$, $4N\mu = 10^{-2}$ and $v = 10^{-8}$.

logarithm of the product of the likelihood for the eight transposable elements observed. Each of the different transposable elements has different patterns and these are reflected differently in their effect on the likelihood surface. One advantage of the likelihood method is that the effect of each individual character on the likelihood can be examined. The p element is shared by roughly half of the haplotypes. If this element is analyzed in isolation there is no evidence that selection is operating on it. The likelihood decreases with increasing levels of selection, so that the maximum likelihood occurs in the absence of selection. Only one other transposable element is shared by more than one haplotype and it is only shared by two. The remaining transposable elements are all present only once. In each of these cases, there is strong evidence that the individual transposable elements may be selectively deleterious.

To gain a better concept of the likelihood surface, the tree with $4N\mu = 10^{-2}$ and $4Nv = 10^{-8}$ was chosen as an example. Figure 4 gives the likelihood surface for this fixed tree as the mutation rate and selection coefficient of the deleterious characters changes [plotted as a function of $\log(4N\mu)$ and $\log(4Ns)$]. Figure 4 indicates that the surface is continuous and smooth except for a slight discontinuity at the junction between the two approximations (as is to be expected). Note that the likelihood surface is very broad and indicates that a wide range of parameter values is compatible with the data. This feature has been noted previously and seems to be a common feature of phylogenetic reconstructions (Felsenstein 1981). For any reconstructed tree of this nature, maximum likelihood estimates of the mutation rates or other parameters are extremely variable. However, there is significant curvature in the surface indicating that the likelihood does have large changes and that useful maximum likelihood estimates can be made for these parameters.

The overall maximum likelihoods for Table 6 and Figure 4 all suggest the presence of selection if the mutation rate of transposable elements is large. It is interesting that Kaplan and Brookfield (1983) have found high values of mutation from a dynamic model of transposition and deletion. They used the distribution of three *copia*-like elements (Montgomery and Langley 1983) and found estimates of $\theta = 4N\mu$ larger than 16 (their μ measures only the rate of spontaneous deletions and not insertions; hence the corresponding value of $4N\mu$ might be even larger). This would place the transposable elements into a region where deleterious selection would be very strongly supported. The magnitude of the MLE estimate of $4N\mu$ from these tables is somewhat smaller, but this estimate assumes that mutation occurs equally to each state. If this assumption is strongly violated, this maximum likelihood model will not be valid.

Discussion

Many different kinds of mutational events can occur within a small region of a DNA sequence. We have focused on the information that can be obtained from a reconstructed evolutionary history to follow not only the frequency of mutational events, but also the origins of sequence types. The origin of sequence types, their branching pattern, and branch lengths combined with other mutational events permit an inference of evolutionary process.

One way to extract a large portion of the information contained within a tree is to examine the likelihood. This is done here for several simple trees. The tables present examples of what the likelihood surface looks like. This is done without the added information that is available from within population data. This information can be extracted using the coalescent process but the addition of selection to this process appears to be technically difficult. The algorithm here uses only that information available assuming that each taxon represents a different species. The likelihood surface for this case is generally broad emphasizing the variable nature of populations. There is however good curvature to the surface both as the selection coefficient changes and as the mutation rate changes. This implies that MLE estimates of these parameters are possible and would be reliable.

Using standard likelihood ratio tests it is possible to test for the presence of selection. In some cases these tests have a great deal of power for demonstrating the effects of deleterious selection. The examples given here show that deleterious selection may help to explain the distribution of some character states. In all of the cases presented, the mutation rate for the deleterious characters had to be large to support the presence of selection. This is partly because only one or a few deleterious characters are considered. In DNA sequence data, the potential number of sites that might have deleterious states is very large. For each of these sites the effect on the likelihood would accumulate and could be quite substantial if many sites are considered. In this way, even very small amounts of selection could be detected.

The results shown here do not explicitly rule out other more complicated hypotheses. These results are based on a simple model and other explanations invoking more complicated phenomena should not be discounted. Maximum likelihood phylogenies could be calculated for most of these phenomena. However, adding such things as unusual mutation patterns would be specific to each case, and although they can be incorporated into the algorithm, they would require a specific model of their effects to be presented. At present most of these phenomena are not sufficiently understood to permit specific models capable of accurately describing their dynamics.

The possibility that selection can significantly alter the maximum likelihood would caution against overinterpretation of trees based upon many sites that may be under very strong selection such as in RNA. The sequence changes that have occurred in RNAs are almost certainly under very strong selection pressures as evidenced by the high frequency of complementary mutations.

There are several potential sources of error in this analysis. The average age of molecular polymorphisms is a quantity that has yet to be determined with accuracy. The age of these polymorphisms could be very large and during these long time periods it is conceivable that mutation rates and selection coefficients could change. The model used here assumes that all parameters are constant. However, this will not be a serious omission whenever the use of an average value for these parameters gives an accurate result.

This method assumes that all haplotypes are independent after branching. Although this is probably true for separate species it is only an approximation for haplotypes chosen from within a single species. In this case, the evolution of each haplotype is not independent and the a priori probabilities of different trees will be affected by selection. This alteration of the tree by selection is a potential source of information that is not being exploited here. Another way in which the haplotypes may not be independent is through the overdominance between the fast and slow alleles of *Adh* that has been suggested by several authors. Due to this selection these haplotypes have not had independent evolution since their creation via mutation. Hence this is not a de-

finitive analysis of the transposable elements in *Drosophila* but rather has been used to illustrate the methodology.

Other potential sources of error are the accuracy of the reconstructed tree and the accuracy of the location of the root of the tree. More than one tree is possible and several different trees should be analyzed. This can simply be done by considering the different trees that are presented from bootstrap samples of the sequences data.

As more detailed molecular information becomes available for population studies, more powerful analyses can be made. The pattern of transitions in a phylogenetic reconstruction contains a large store of this information. Although the likelihood surfaces of the example phylogenies are broad, there do appear to be general patterns to the likelihood, and maximum likelihood estimates of parameters are possible. These methods also permit standard statistical tests to be applied in the search for the effects of selection. Perhaps advantage can be taken of these in future, more extensive studies.

Acknowledgments. We would like to thank E. Zuckerkandl, A.M. Dean, B.S. Weir, S. Williams, and anonymous reviewers for their very helpful comments on this study. This work was supported by the Natural Sciences and Engineering Research Council of Canada grant number U0336 to B.G. and NIH grant GM41716 to J.F. Additional support was provided by the Canadian Institute for Advanced Research.

References

Aquadro CF, Desse SF, Bland MM, Langley CH, Laurie-Ahlberg CC (1986) Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. Genetics 114:1165–1190

Avery PJ (1978) Selection effects in a model of two intermigrating colonies of finite size. Theor Pop Biol 13:24–39

Birky, GW, Walsh JB (1988) Effects of linkage on rates of molecular evolution. Proc Natl Acad Sci USA 85:6414–6418

Felsenstein J (1965) The effect of linkage on directional selection. Genetics 52:349–363

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376

Felsenstein J (1988) Phylogenies from molecular sequences: inference and reliability. Annu Rev Genet 22:521–565

Fukami K, Tateno Y (1989) On the maximum likelihood method for estimating molecular trees: uniqueness of the likelihood point. J Mol Evol 28:460–464

Gillespie JH (1984a) The molecular clock may be an episodic clock. Proc Natl Acad Sci USA 81:8009–8013

Gillespie JH (1984b) Molecular evolution over the mutational landscape. Evolution 38(5):1116–1129

Gillespie JH (1986) Variability of evolutionary rates of DNA. Genetics 113:1077–1091

Gillespie JH (1989) Lineage effects and the index of dispersion of molecular evolution. Mol Biol Evol 6:636–648

Golding GB (1987) The detection of deleterious selection using ancestors inferred from a phylogenetic history. Genet Res Camb 49:71–82

Golding GB, Aquadro CF, Langley CH (1986) Sequence evolution within populations under multiple types of mutation. Proc Natl Acad Sci USA 83:427–431

Graur D, Li WH (1989) Evolution of protein inhibitors of serine proteinases—positive Darwinian selection or compositional effects. J Mol Evol 28:131–135

Hartl DL, Dykhuizen DE, Dean AM (1985) Limits of adaptation: the evolution of selective neutrality. Genetics 111: 655–674

Hill RE, Hastie ND (1987) Accelerated evolution in the reactive centre regions of serine protease inhibitors. Nature 326:96–99

Hudson RR (1983) Testing the constant rate neutral allele model with protein sequence data. Evolution 37:203–217

Iizuka M (1989) A population genetical model for sequence evolution under multiple types of mutation. Genet Res 54: 231–237

Kaplan NL, Brookfield JFY (1983) Transposable elements in Mendelian populations III. Statistical results. Genetics 104: 485–495

Kimura M (1955) Solution of a process of random genetic drift with a continuous model. Proc Natl Acad Sci USA 41:144–150

Kimura M (1962) On the probability of fixation of mutant genes in a population. Genetics 4:713–719

Kimura M (1979) Model of effectively neutral mutations in which selective constraint is incorporated. Proc Natl Acad Sci USA 76:3440–3444

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, New York

Koop BF, Goodman M, Xer P, Chan K, Slightom JL (1986) Primate z-globin DNA sequences and man's place among the great apes. Nature 319:234–238

Langley CH, Fitch WM (1974) An examination of the constancy of the rate of molecular evolution. J Mol Evol 3:161–177

Maeda N, Wu C, Bliska J, Reneke J (1988) Molecular evolution of intergenic DNA in higher primates: pattern of DNA changes, molecular clock, and evolution of repetitive sequences. Mol Biol Evol 5(1):1–20

Montgomery EA, Langley CH (1983) Transposable elements in Mendelian populations II. Distribution of three copia-like elements in a natural population of *Drosophila melanogaster*. Genetics 104:473–483

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Ohta T (1973) Slightly deleterious mutant substitutions in evolution. Nature 246:96–98

Ohta T (1976) Role of very slightly deleterious mutations in molecular evolution and polymorphism. Theor Pop Biol 10: 254–275

Ohta T, Kimura M (1971) On the constancy of the evolutionary rate of cistrons. J Mol Evol 1:18–25

Sawyer SA, Dykhuizen DE, Hartl DL (1987) Confidence interval for the number of selectively neutral amino acid polymorphisms. Proc Natl Acad Sci USA 84:6225–6228

Wilson AC, Carlson SS, White TJ (1977) Biochemical evolution. Annu Rev Biochem 46:573–639

Received August 20, 1989/Revised June 20, 1990/Accepted July 1, 1990

Appendix 1

Using the same method as described in the text, an approximation to $E^t(x)$ can be derived that is accurate to an order of $(Ns)^2$. In this case designate $E^t(x)$, $E^t(x^2)$, and $E^t(x^3)$ all evaluated at $Ns = 0$ as E_1^t, E_2^t, and E_3^t, respectively. Designate the (Ns) order approximations of $E^t(x)$ and $E^t(x^2)$ by W_1^t and W_2^t and the $(Ns)^2$ approximation of $E^t(x)$ by Z_1^t. The system of recursion equations for these is

$$\mathbf{Z}^{t+1} = \mathbf{M}\mathbf{Z}^t + \mathbf{V}$$

where

$$\mathbf{Z}^t = [E_1^t, E_2^t, E_3^t, W_1^t, W_2^t, Z_1^t]^T$$
$$\mathbf{V} = [\mu, \ 0, \ 0, \ \mu, \ 0, \ \mu \ ]^T$$

and

$$\mathbf{M} = \begin{bmatrix} 1-2\mu, & 0, & 0, & 0, & 0, & 0 \\ 2\mu + 1/2N, & 1-4\mu-1/2N, & 0, & 0, & 0, & 0 \\ 0, & 3\mu + 3/2N, & 1-6\mu-3/2N, & 0, & 0, & 0 \\ s, & -s, & 0, & 1-2\mu, & 0, & 0 \\ 0, & 2s, & -2s, & 2\mu + 1/2N, & 1-4\mu-1/2N, & 0 \\ 0, & 0, & 0, & s, & -s, & 1-2\mu \end{bmatrix}$$

To solve this system note that $\mathbf{M}$ can be decomposed into $\mathbf{PDP}^{-1}$. Where

$$\mathbf{D} = \begin{bmatrix} 1-2\mu, & 0, & 0, & 0, & 0, & 0 \\ 0, & 1-4\mu-1/2N, & 0, & 0, & 0, & 0 \\ 0, & 0, & 1-6\mu-3/2N, & 0, & 0, & 0 \\ 0, & 0, & 0, & 1-2\mu, & 0, & 0 \\ 0, & 0, & 0, & 0, & 1-4\mu-1/2N, & 0 \\ s^2, & 0, & 0, & 0, & 0, & 1-2\mu \end{bmatrix}$$

$$\mathbf{P} = \begin{bmatrix} -(1+\theta)(3+2\theta)/2N\theta, & 0, & 0, & 0, & 0, & 0 \\ -(1+\theta)(3+2\theta)/2N\theta, & 0, & 0, & 0, & 1, & 0 \\ -3(1+\theta)(2+\theta)/4N\theta, & 0, & 1, & 0, & 3/2, & 0 \\ s, & 0, & 0, & 1, & 2Ns/(1+\theta), & 0 \\ 0, & 1, & 4Ns/(2+\theta), & 1, & 0, & 0 \\ s^2, & 2Ns/(1+\theta), & 8N^2s^2/(2+\theta)(3+2\theta), & 0, & -4N^2s^2/(1+\theta)^2, & 1 \end{bmatrix}$$

$$\mathbf{P}^{-1} = \begin{bmatrix} \dfrac{-2N\theta}{(1+\theta)(3+2\theta)}, & 0, & 0, & 0, & 0, & 0 \\ \dfrac{-6Ns}{(2+\theta)}, & \dfrac{2Ns(5+4\theta)}{(1+\theta)(2+\theta)}, & \dfrac{-4Ns}{(2+\theta)}, & -1, & 1, & 0 \\ \dfrac{3(1+\theta)}{2(3+2\theta)}, & -3/2, & 1, & 0, & 0, & 0 \\ \dfrac{+6Ns}{(3+2\theta)}, & \dfrac{-2Ns}{(1+\theta)}, & 0, & 1, & 0, & 0 \\ -1, & 1, & 0, & 0, & 0, & 0 \\ \dfrac{4N^2s^2(5\theta^2+9\theta+3)}{(1+\theta)^2(3+2\theta)^2} + \dfrac{\theta(4N^2s^2)}{2N(1+\theta)(3+2\theta)}, & \dfrac{-12N^2s^2}{(1+\theta)(3+2\theta)}, & \dfrac{8N^2s^2}{(1+\theta)(3+2\theta)}, & \dfrac{2Ns}{(1+\theta)}, & \dfrac{-2Ns}{(1+\theta)}, & 1 \end{bmatrix}$$

The transient solution is then given by

$$\mathbf{Z}^t = \mathbf{PD}^t\mathbf{P}^{-1}\mathbf{Z}^0 + \sum_{i=0}^{t-1} \mathbf{PD}^i\mathbf{P}^{-1}\mathbf{V}$$

and the sum is easily solved due to the nature of $\mathbf{D}$.

Appendix 2

Two completely linked loci, each with two alleles can also be analyzed in the presence of selection. Designate the haplotype frequencies as x_{00}, x_{01}, x_{10}, and x_{11}. Fitnesses of the four haplotypes are assumed to be 1, 1 − r, 1 − s, and (1 − s)(1 − r), respectively. The first site mutates at a rate μ and the second site at a rate ν. The method described in the text yields recursion equations for these frequencies as,

$$\begin{aligned}
E(x_{00} + \delta x_{00}) &= \mu E(x_{10}) + \nu E(x_{01}) + (1 - \mu - \nu)E(x_{00}) \\
&\quad + sE(x_{00}x_{10}) + sE(x_{00}x_{11}) + rE(x_{00}x_{01}) \\
&\quad + rE(x_{00}x_{11}) \\
E(x_{01} + \delta x_{01}) &= \mu E(x_{11}) + \nu E(x_{00}) + (1 - \mu - \nu - r)E(x_{01}) \\
&\quad + sE(x_{01}x_{10}) + sE(x_{01}x_{11}) + rE(x_{01}x_{01}) \\
&\quad + rE(x_{01}x_{11}) \\
E(x_{10} + \delta x_{10}) &= \mu E(x_{00}) + \nu E(x_{11}) + (1 - \mu - \nu - s)E(x_{10}) \\
&\quad + sE(x_{10}x_{10}) + sE(x_{10}x_{11}) + rE(x_{01}x_{10}) \\
&\quad + rE(x_{10}x_{11}) \\
E(x_{11} + \delta x_{11}) &= \mu E(x_{01}) + \nu E(x_{10}) + (1 - \mu - \nu - s - r)E(x_{11}) \\
&\quad + sE(x_{10}x_{11}) + sE(x_{11}x_{11}) + rE(x_{01}x_{11}) \\
&\quad + rE(x_{11}x_{11})
\end{aligned}$$

with

$$\begin{aligned}
E(\delta x_{00}|x) &= -(\mu + \nu)x_{00} + \mu x_{10} + \nu x_{01} \\
E(\delta x_{01}|x) &= -(\mu + \nu)x_{01} + \mu x_{11} + \nu x_{00} \\
E(\delta x_{10}|x) &= -(\mu + \nu)x_{10} + \mu x_{00} + \nu x_{11} \\
E(\delta x_{11}|x) &= -(\mu + \nu)x_{11} + \mu x_{01} + \nu x_{10}
\end{aligned}$$

due to mutation,

$$\begin{aligned}
E(\delta x_{00}|x) &= x_{00}[s(x_{10} + x_{11}) + r(x_{01} + x_{11})] \\
E(\delta x_{01}|x) &= x_{01}[-r + s(x_{10} + x_{11}) + r(x_{01} + x_{11})] \\
E(\delta x_{10}|x) &= x_{10}[-s + s(x_{10} + x_{11}) + r(x_{01} + x_{11})] \\
E(\delta x_{11}|x) &= x_{11}[-s - r + s(x_{10} + x_{11}) + r(x_{01} + x_{11})]
\end{aligned}$$

due to selection and,

$$\begin{aligned}
E[(x_{00} + \delta x_{00})^2]\Big|_{s=0} &= E(x_{00})/2N \\
&\quad + (1 - 2\mu - 2\nu - 1/2N)E(x_{00}^2) \\
&\quad + 2\mu E(x_{00}x_{10}) + 2\nu E(x_{00}x_{01}) \\
E[(x_{01} + \delta x_{01})^2]\Big|_{s=0} &= E(x_{01})/2N \\
&\quad + (1 - 2\mu - 2\nu - 1/2N)E(x_{01}^2) \\
&\quad + 2\mu E(x_{01}x_{11}) + 2\nu E(x_{00}x_{01}) \\
E[(x_{10} + \delta x_{10})^2]\Big|_{s=0} &= E(x_{10})/2N \\
&\quad + (1 - 2\mu - 2\nu - 1/2N)E(x_{10}^2) \\
&\quad + 2\mu E(x_{00}x_{10}) + 2\nu E(x_{10}x_{11}) \\
E[(x_{11} + \delta x_{11})^2]\Big|_{s=0} &= E(x_{11})/2N \\
&\quad + (1 - 2\mu - 2\nu - 1/2N)E(x_{11}^2) \\
&\quad + 2\mu E(x_{01}x_{11}) + 2\nu E(x_{10}x_{11}) \\
E[(x_{00} + \delta x_{00})(x_{01} + \delta x_{01})]\Big|_{s=0} &= (1 - 2\mu - 2\nu - 1/2N)E(x_{00}x_{01}) \\
&\quad + \mu E(x_{00}x_{11}) + \mu E(x_{01}x_{10}) \\
&\quad + \nu E(x_{00}x_{00}) + \nu E(x_{01}x_{01}) \\
E[(x_{00} + \delta x_{00})(x_{10} + \delta x_{10})]\Big|_{s=0} &= (1 - 2\mu - 2\nu - 1/2N)E(x_{00}x_{10}) \\
&\quad + \mu E(x_{00}x_{00}) + \mu E(x_{10}x_{10}) \\
&\quad + \nu E(x_{01}x_{10}) + \nu E(x_{00}x_{11}) \\
E[(x_{00} + \delta x_{00})(x_{11} + \delta x_{11})]\Big|_{s=0} &= (1 - 2\mu - 2\nu - 1/2N)E(x_{00}x_{11}) \\
&\quad + \mu E(x_{00}x_{01}) + \mu E(x_{10}x_{11}) \\
&\quad + \nu E(x_{01}x_{11}) + \nu E(x_{00}x_{10}) \\
E[(x_{01} + \delta x_{01})(x_{10} + \delta x_{10})]\Big|_{s=0} &= (1 - 2\mu - 2\nu - 1/2N)E(x_{01}x_{10}) \\
&\quad + \mu E(x_{10}x_{11}) + \mu E(x_{00}x_{01}) \\
&\quad + \nu E(x_{00}x_{10}) + \nu E(x_{01}x_{11}) \\
E[(x_{01} + \delta x_{01})(x_{11} + \delta x_{11})]\Big|_{s=0} &= (1 - 2\mu - 2\nu - 1/2N)E(x_{01}x_{11}) \\
&\quad + \mu E(x_{11}x_{11}) + \mu E(x_{01}x_{01}) \\
&\quad + \nu E(x_{00}x_{11}) + \nu E(x_{01}x_{10}) \\
E[(x_{10} + \delta x_{10})(x_{11} + \delta x_{11})]\Big|_{s=0} &= (1 - 2\mu - 2\nu - 1/2N)E(x_{10}x_{11}) \\
&\quad + \mu E(x_{00}x_{11}) + \mu E(x_{01}x_{10}) \\
&\quad + \nu E(x_{11}x_{11}) + \nu E(x_{10}x_{10})
\end{aligned}$$

Again all parameters are considered to be small (on the order of 1/2N) and terms in $(Ns)^2$, $(Nr)^2$, and (Ns)(Nr) are ignored.

These equations were iterated to yield Table 1.