Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 427–431, January 1986
Genetics

Sequence evolution within populations under multiple types of mutation

(transposable elements/deleterious selection/phylogenies)

G. B. GOLDING*, C. F. AQUADRO†, AND C. H. LANGLEY

Laboratory of Genetics, National Institute of Environmental Health Sciences, P.O. Box 12233, Research Triangle Park, NC 27709

Communicated by C. Clark Cockerham, September 20, 1985

ABSTRACT **DNA sequence and restriction map data from natural populations can be used to estimate the phylogenetic history of observed sequences. Studies of this sort usually examine large regions of DNA where many evolutionary events have taken place. From such data, detailed phylogenies can be constructed and qualitatively different kinds of mutational and substitutional processes can be studied. The value of investigating more than one mutational type is in the power of comparing relative rates of these distinct substitutional processes. In this paper, we construct a neutral model to describe the frequencies of sequence haplotypes according to the haplotypes from which they arose. This theory of the frequency of haplotypes (incorporating their historical past) is applied to data from the alcohol dehydrogenase gene region of *Drosophila melanogaster*. The observed patterns of change associated with transposable elements around the *Adh* locus are not in accord with a neutral model. Values for the mutation rates cannot be found that will bring the observed data into agreement with a simple neutral model, but with the addition of mildly deleterious selection, the model can explain these patterns of change. These results suggest that transposable elements are deleterious to the organisms carrying them, but at levels only several times their rate of transpositional insertion. Similar analyses of small deletions in the *Adh* region suggest that they may experience mildly deleterious selection.**

Recent technical advances in molecular biology have made possible the examination of genetic variation among individuals and populations at a new level of detail. In particular, restriction mapping, cloning, and rapid DNA sequencing allow the determination of not only the quantity of DNA sequence variation but also the nature of this variation. DNA restriction map and sequence data also contain a wealth of information concerning the origins of different sequence types. It is possible to use the pattern of sequence variants to reconstruct the phylogenetic relationships among the sequences being examined. This history provides an estimate of the order and relative frequency of the various types of sequence-altering events. For example, base substitutions, deletions, and transposable element insertions represent different types of mutational events that can be identified. These different classes of events may occur at different rates and may have different evolutionary consequences. If we focus on a relatively short segment of sequence (say 5–20 kilobases), most of the different substitutions observed in this sequence will have shared a common evolutionary history and have been influenced to a similar extent by the natural forces that have shaped this history. Thus, the analysis of many different substitutional events among a sample of relatively short DNA sequences eliminates many of the problems inherent in an unknown evolutionary history. In this paper, we consider the frequency of events inferred from phylogenies for an intraspecific sample of DNA sequences with special reference to the origins and evolutionary consequences of unique sequence insertions, small deletions, and transposable element insertions.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

METHOD

As a conceptual example, consider the simplest case with only two types of events: nucleotide substitutions and insertions. Each sequence will have a characteristic number of insertions relative to some standard, and base substitutions may further differentiate the sequences. Some sequences may have the same number of insertions and yet differ by base substitutions. Consider a random mating population of infinite size without selection. Let the generations be discrete and nonoverlapping. Assume that, from one generation to the next, only one mutational event can occur per gamete. Let the two different kinds of characters contained within the sequence be denoted by α and β.

A particular sequence containing i of the β characters could have arisen in several ways. The most recent distinguishable ancestor of this sequence could have contained either $i - 1$ or $i + 1$ β characters. This ancestor could also have been a sequence type with i β characters and have given rise to the new sequence type by a change in an α-type character. For example, in this model a sequence that contains insertions has three possible ancestors; the most recent distinguishable ancestor could have had one less insertion, it could differ from the present sequence by the presence of a base substitution, or the ancestor could have one extra insertion. Denote the frequency of sequences in each of these classes of recent ancestry by $g[i - 1, i]$, $g[i, i]$, and $g[i + 1, i]$. The second argument refers to the number of β-type characters in the sequence and the first refers to the number of β-type characters in the most recent distinguishable ancestor.

Let α characters mutate at a rate u_1 per gamete per generation. Let β characters increase at a rate u_2 per gamete per generation. The loss of β characters occurs at a rate linearly dependent on the number already present in the gamete. A sequence with i β characters is assumed to change to a sequence with $i - 1$ β characters at a rate iu_3 per gamete per generation. The α and β characters may represent any two types of defined sequence alteration. For example, they could represent base substitutions and deletions and compare the relative changes occurring to these two types of characters. Alternatively, the α characters may represent changes in the first and second codon positions, whereas β characters represent the gain or spontaneous reversion (loss) of third base substitutions. In this case the relative changes in the

Abbreviation: UPG, unweighted pair group.

*Present address: Department of Biology, York University, Downsview, ON, Canada, M3J 1P3.
†Present address: Section of Genetics and Development, Cornell University, Ithaca, NY 14853.

different codon positions would be compared and one could safely assume that $u_3 << u_2$. Note that when the rates of loss are of the same magnitude as the rates of gain, only a limited amount of β character divergence can accumulate. This is because, although two sequences will accumulate β characters as they diverge, the mutation pressure to revert back to the original configuration increases as more and more β characters occur.

By analyzing the changes that can occur to α- and β-type characters it is possible to build a system of recursion equations that describe the frequencies of sequence types. Consider first that, when no mutations occur from one generation to the next, all sequences of a particular type will remain in that particular class (frequencies will not change). When the number of β characters increases by mutation (u_2), only sequences with $i - 1$ βs need be considered. There are three possible origins for sequences with $i - 1$ β characters. If a change in an α character occurs (u_1), then the number of β characters does not change and the most recent distinguishable ancestor of any sequences with i β characters that sustain such a mutation will also have i β characters. Again there are three possible origins for sequences with i β. If the number of β characters is decreased by mutation (u_3), then those sequences with $i + 1$ β characters contribute to the frequency class $g[i + 1, i]$. The frequency of each sequence type in the next generation (denoted by a prime) is therefore

$$\begin{aligned} g[i-1, i]' &= (1 - u_1 - u_2 - iu_3)g[i-1, i] + u_2 f[i-1] \\ g[i, i]' &= (1 - u_1 - u_2 - iu_3)g[i, i] + u_1 f[i] \\ g[i+1, i]' &= (1 - u_1 - u_2 - iu_3)g[i+1, i] + (i+1)u_3 f[i+1], \end{aligned} \quad [1]$$

where $f[i] = g[i-1, i] + g[i, i] + g[i+1, i]$, with $f[0] = g[0, 0] + g[1, 0]$ and $f[n] = g[n-1, n] + g[n, n]$. Boundary conditions can be added so that

$$\begin{aligned} g[0, 0]' &= (1 - u_2)g[0, 0] + u_1 g[1, 0] \\ g[n, n]' &= (1 - nu_3)g[n, n] + (u_1 + u_2)g[n-1, n]. \end{aligned}$$

With these conditions, the number of β characters per sequence must be between zero and n. Of course, n can be infinitely large.

These equations form a linear system of equations and are a special case of a Poisson process. Solved at equilibrium these equations give

$$\begin{aligned} g[i, i] &= [u_1/(u_1 + u_2 + iu_3)]f[i] \\ g[i+1, i] &= [(i+1)u_3/(u_1 + u_2 + iu_3)]f[i+1] \\ g[i, i+1] &= [u_2/(u_1 + u_2 + (i+1)u_3)]f[i], \end{aligned} \quad [2]$$

where $0 < i < n$. The boundary solutions are

$$\begin{aligned} g[0, 0] &= [u_1/(u_1+u_2)]f[0] \\ g[n, n] &= [(u_1 + u_2)/(u_1 + u_2 + nu_3)]f[n] \end{aligned}$$

with

$$f[i] = (u_2/u_3)^i \left[i! \sum_{j=0}^{n} \frac{1}{j!} (u_2/u_3)^j \right]^{-1}. \quad [3]$$

Two inequalities allow a quick visual check of data to determine if they are in accord with the model. From the solutions (Eqs. 2 and 3) it can be seen that at equilibrium, for $i > 0$, and for all values of u_2 and u_3, that $g[i-1, i] < g[i, i]$, when $iu_3 < u_1$, and that $g[i-1, i] < g[i, i]+g[i+1, i]$, if $iu_3 < u_1+u_2$. The first inequality states that unless β characters are lost at a high rate, the number of sequences whose most recent distinguishable ancestor had one fewer β should be less frequent than those whose most recent distinguishable ancestor had the same number of β characters. The second inequality shows that those sequences with one less β in their most recent distinguishable ancestor should be fewer than all other kinds of sequences with the same number of β characters. Most of the time, the second inequality is easily satisfied because the spontaneous loss rate for many βs will be less than the rate at which they occur.

RESULTS

Data. As an illustration of the application of this theory we have examined the data of Aquadro *et al.* (1). These authors surveyed 49 lines of *Drosophila melanogaster*, each homozygous for a different second chromosome. These chromosomes were isolated from five different locations (2). A 13-kilobase region around the *Adh* gene in each line was examined with seven restriction endonucleases (*Bam*HI, *Ban* II, *Bgl* II, *Eco*RI, *Hin*dIII, *Sal* I, *Xho* I) and for the allozyme polymorphism. In addition to the fast/slow (F/S) allozyme polymorphism and the gain and loss of restriction sites, numerous insertions and deletions were detected. The 49 sequences formed 29 different haplotypes on the basis of the F/S allozyme polymorphism, eight restriction site polymorphisms, seven deletions, two insertions of unique sequence, and seven sizes of transposable element insertions. It is assumed that these characters are close enough that they will reflect the history of the events that have occurred. This does not imply that recombination never occurred but that recombination has not occurred at a rate sufficient to obscure the phylogenetic history of the haplotypes (see ref. 1).

Trees. Phylogenetic trees can be constructed in many different ways, often with slightly different results (3). For our analysis to be useful, the results should not depend strongly on the particular method used to construct the tree. Therefore, several different trees are examined. The first was built according to the unweighted pair group (UPG) method as modified by Li (4) to allow for different rates within each lineage. The second method used the maximum parsimony algorithm supplied by Felsenstein's (3) PHYLIP package. Because this algorithm does not produce a rooted tree and since a direction is implied by the expected frequencies presented above, it was necessary to root the trees. The method should be only mildly sensitive to the location of the root because only the immediate ancestor of a sequence and not more remote ancestors are considered. The parsimony tree was rooted in the general location as suggested by the UPG method (denoted by Max. Pars. ii). Two other locations for the root were also analyzed. In these two cases (denoted by Max. Pars. i and iii), the roots were placed toward opposite ends of the above rooted maximum parsimony tree.

The final tree was constructed using maximum parsimony criteria as well as considering the direction of mutational changes [e.g., convergent loss of a restriction site is more likely than convergent gain (as discussed in ref. 1)] and allowing for the possibility of recombination. Recombination can be approximated as another of the events that occur with a rate u_1 and, like base substitutions, it acts as a marker on the sequence. In this case, the recombinant gametes can be said to have originated from each type of gamete and an equal probability of origin is given to each type. For example, one of the three recombinants in this tree occurred between a sequence with a deletion and one without any deletions. Because this recombinant contained a deletion, this gamete is scored as contributing equally to [0, 1] and [1, 1]. In general, it will not be possible to score all recombinants in this way because the model assumes that changes occur in single steps but recombination can result in apparent multiple changes. Fig. 1 illustrates the third tree.

Analysis. The deletions, insertions, deletions and insertions together, and transposable elements from each of the variously constructed trees were considered in turn to be the β-type character with all others considered as α-type characters. The observed numbers of haplotypes of each gametic

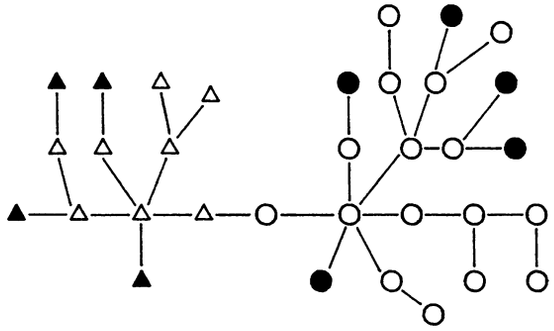

FIG. 1. Diagrammatic phylogeny for *Adh* region haplotypes of *D. melanogaster*. Note that the haplotypes containing transposable elements (solid symbols) are all at the tips of the tree branches. The details of the illustrated phylogeny are published elsewhere (1). Similar results are obtained for all methods of tree construction used (see text). Haplotypes carrying the *Adh* fast allele or slow allele are indicated by triangles or circles, respectively. Branch lengths do not reflect the number of substitutions. Recombinant chromosomes (see ref. 1) are placed according to the origins of their sequences immediately surrounding the *Adh* locus.

type $[i, j]$ were then calculated from the reconstructed trees, as given in Table 1. Because only one event is assumed to occur per generation, when there are two or more differences between the ancestral and extant sequences, each difference is given equal weight as possibly being the most recent. For example, a sequence that differs from its ancestral sequence by a base substitution and the addition of an insertion would contribute equally to the $[i, i]$ and $[i-1, i]$ classes because it is not known whether the base substitution or the insertion was the last event to occur. For computational convenience, the maximum number of deletions, insertions, or transposable elements was set to be $n = 2$. This artificial limit should not strongly affect the results since there were no more than two deletions, insertions, or transposable elements per haplotype and only two cases with three deletions and insertions combined.

Table 1. Observed numbers of sequence types in each mutational class according to the reconstructed phylogenetic tree

	Mutational class*						
	[0, 0]	[1, 0]	[0, 1]	[1, 1]	[2, 1]	[1, 2]	[2, 2]
UPG							
Del.	7.0	1.0	3.5	12.5	0.0	3.0	2.0
Ins.	22.0	0.0	3.0	4.0	0.0	0.0	0.0
TE	20.0	0.0	8.0	1.0	0.0	0.0	0.0
Del. + Ins.	5.0	1.0	5.0	10.0	0.0	4.0	4.0
Max. Pars. i							
Del.	8.0	0.0	5.5	10.5	0.0	3.0	2.0
Ins.	21.0	1.0	1.0	6.0	0.0	0.0	0.0
TE	20.0	0.0	7.83	1.17	0.0	0.0	0.0
Del. + Ins.	5.0	1.0	4.5	10.5	0.0	4.0	4.0
Max. Pars. ii							
Del.	7.0	1.0	4.5	11.5	0.0	3.0	2.0
Ins.	21.0	1.0	1.0	6.0	0.0	0.0	0.0
TE	20.0	0.0	7.83	1.17	0.0	0.0	0.0
Del. + Ins.	4.0	2.0	3.5	11.5	0.0	4.0	4.0
Max. Pars. iii							
Del.	8.0	0.0	4.5	11.5	0.0	3.0	2.0
Ins.	21.0	1.0	1.0	6.0	0.0	0.0	0.0
TE	20.0	0.0	7.83	1.17	0.0	0.0	0.0
Del. + Ins.	5.0	1.0	3.5	11.5	0.0	4.0	4.0
With recombination							
Del.	7.5	0.5	5.5	10.5	0.0	3.0	2.0
Ins.	21.5	0.5	1.5	5.5	0.0	0.0	0.0
TE	20.0	0.0	8.0	1.0	0.0	0.0	0.0
Del. + Ins.	5.0	1.0	5.0	10.0	0.0	4.0	4.0

Phylogenies were reconstructed by UPG, by maximum parsimony with three different roots (Max. Pars. i–iii), and by maximum parsimony with recombination. The observed numbers are for deletions (Del.), insertions (Ins.), transposable elements (TE), and the deletions and insertions combined (Del. + Ins.).

*Mutation class is denoted by $[i, j]$, where j is the number of β characters present in the extant sequence and i is the number in the sequence's most recent distinguishable ancestor.

The frequencies that best matched the observed numbers were calculated. This was done by finding the set of u_1, u_2, and u_3 that minimized the χ^2 between observed and expected numbers of each sequence type. These minima were found by multiple univariate grid searches starting from several different initial values. In some cases, it was not possible to find values for u_1, u_2, and u_3 that minimized the χ^2 deviations below a statistically significant level. For example, an examination of Table 1 indicates that the data for transposable elements violate the inequalities given above, as there is an excess of sequences that apparently arose from sequences with one less transposable element. This unusual distribution is demonstrated diagrammatically in Fig. 1. All sequences containing transposable elements (the solid circles and triangles) are at the tips of the tree and none is in the interior. There are no sequences with a transposable element that definitely gave rise to sequences with two transposable elements or to sequences with an additional base substitution (those present in Table 1 in the class $g[1, 1]$ are due only to uncertainties in the most recent mutational event). One possible explanation would be that the transposable elements are selectively deleterious.

To determine the amount of selection necessary to explain the data, an additive genic selection model is assumed, such that the fitness of sequences with i transposable elements is $(1 - is)$, where $s > 0$. This was incorporated into Eq. **1** and numerical equilibrium solutions were found by using Newton's method. The values of u_2, u_3, and s were found that minimized the χ^2, subject to the constraint that s also be at a minimum whenever the χ^2 deviations were not significant at the 5% level. This value for s determines expected frequencies such that the probability of the differences between observed and expected (based on a χ^2) is 0.05 and any smaller value for s would imply significant differences. In this way, the minimum amount of selection that could possibly explain the data is found. Although these are only local solutions and cannot be shown to be the global minimum, no other solutions were found by the algorithm in runs starting from several different initial values.

Table 2 gives the values of u_2 and u_3 that minimize the χ^2 with the minimum amount of selection possible. The values of the rate of gain of β characters (u_2), the rate of loss of β characters (u_3), and the selection coefficient against β characters (s) are shown as a percentage of the mutation rate of α characters (u_1). This is because the solutions, for the model presented, depend on u_1 only as a constant multiple.

Note that these results are for equilibrium frequencies and are only approximations for nonequilibrium data. Comparison of haplotype diversity of *Adh* fast and slow chromosomes suggests that transposable element and deletion/insertion variation are at or near equilibrium (1).

The deletions can be explained without selection by all but one of the trees. The maximum parsimony tree (*i*) suggests a mild level of deleterious selection against deletions. When deletions and insertions of unique sequences are considered together they can again be explained without selection. This is not surprising because this model examines all deletions as a group and one of the deletions is quite common, labeled "p" by Aquadro *et al.* (1). This deletion occurs in roughly half of the sequences and sequences carrying it give rise to

Table 2. Mutation rates (u_2 and u_3) and selection coefficient (s) for β-type characters (expressed as a percentage of u_1) for the data of Table 1 (the value for s is a minimum)

Method of tree reconstruction		β character			
		Del.	Ins.	TE	Del. + Ins.
UPG	u_2	21.4	30.0	34.6	34.5
	u_3	18.3	70.0	30.8	26.1
	s	0.0	0.0	63.7	0.0
Max. Pars. i	u_2	35.5	18.8	34.7	34.5
	u_3	36.4	34.8	44.7	21.8
	s	1.6	0.0	56.8	0.0
Max. Pars. ii	u_2	25.7	18.8	34.7	46.5
	u_3	20.1	34.8	44.7	23.4
	s	0.0	0.0	56.8	0.0
Max. Pars. iii	u_2	24.5	18.8	34.7	30.3
	u_3	23.6	34.8	44.7	16.2
	s	0.0	0.0	56.8	0.0
With recombination	u_2	31.1	30.0	34.6	34.5
	u_3	30.6	70.0	30.8	26.1
	s	0.0	0.0	63.7	0.0

Phylogenies were reconstructed by UPG, by maximum parsimony with three different roots (Max. Pars. i–iii), and by maximum parsimony with recombination. These numbers are for the observed deletions (Del.), insertions (Ins.), transposable elements (TE), and the deletions and insertions combined (Del. + Ins.).

many other sequences. When this deletion is ignored it is found that most of the remaining deletions occur very close to the tips of the trees and require selection if $u_3 < u_2$.

The pattern of change for unique sequence insertions, shown by all trees, can again be explained without any selection. However, the data also suggest that $u_3 > u_2$ and indeed that $u_3 = 2u_2$. If the restriction that $u_2 > u_3$ is made, then selection is consistently required by all trees. The minimum amount of selection is approximately $s = 6$–8% of u_1, with $u_2 = u_3 = 4$–18% of u_1 (results not shown).

In all cases, the transposable elements require selection to explain the observed pattern. In contrast to the insertions of unique sequences, a larger rate of loss is not suggested by the data. The reversion classes ([1, 0] and [2, 1]) would be present in the sample if u_3 becomes too large. The pattern suggests that selection must be >50% of the value of u_1, that the rate of gain (the total rate of insertion of transposable elements) is ≈34% of u_1, and that the rate of loss is from 30% to 48% of u_1. The pattern for all of the reconstructed trees is roughly consistent.

DISCUSSION

Despite the biological significance of even very small amounts of selection, the measurement of this evolutionary force in natural populations has proven very difficult (for reviews, see refs. 5–7). Perhaps, the best known test for selection was proposed by Watterson (8). This test detects the effects of overdominant selection by comparing the observed allelic homozygosity in a sample to that expected under selective neutrality (this test and others are reviewed by Kimura, ref. 7; see also Ewens, ref. 9). Watterson's test can also be used to look for the effects of deleterious selection but, in this case, is sensitive only to an order of the selection coefficient squared. Linkage disequilibrium has also been used to look for evidence of epistatic selection, but again, this is a second-order effect.

The new data available from restriction map and DNA sequence analyses on genetic variation provide more information than just allelic frequencies. These data allow estimation of the phylogenetic relationships among the alleles segregating in populations. This significant new component of the data was not utilized by previous tests of selective neutrality. In this paper, we have developed an approach to analyze the distribution of various types of mutational events in the evolutionary history of a sample of DNA sequences. We have focused on the information that can be obtained from a reconstructed phylogenetic history and thus, we follow not only the frequency of different allelic types but also the origins of these sequence types. The results can be used to provide conservative estimates of the levels of selection necessary to explain observed patterns.

As an illustration of the application of this procedure, the data of Aquadro *et al.* (1) from the *Adh* region of *D. melanogaster* were analyzed. Simulation results, to be published elsewhere, indicate that the analysis is appropriate. The sensitivities of the general method to the number of substitutions, sample size, and occasional recombination are areas for further research. Because the estimated mutation rates can be sensitive to changes in the data, it may not be worthwhile to place emphasis on their actual values. Rather, the model examines all possible values and tries to determine any plausible explanation for the observed frequencies.

The distribution of transposable elements among the *Adh* region haplotypes indicates deleterious selection. The actual size of the required selection does not have to be very large (a little less than twice the rate of transposable element transposition) but it must be present. Previously, Watterson's test had been applied to this data set assuming each deletion, insertion, or transposable element to be a separate allelic type (1). It was found that the transposable elements did not fit a neutral distribution due to an excess of unique "alleles" (haplotypes containing transposable elements). However, this "allele frequency distribution" could be explained either by selection or by high rates of loss of transposable elements. The approach presented here allows a test of these alternatives because the spontaneous rate of loss can take any value that is necessary to fit a neutral model to the observed distribution. Nonetheless, a high rate of loss was not sufficient to give agreement with the data and the model implicates selection against haplotypes containing transposable elements. Consistent with this interpretation is the distribution of transposable elements in phylogenetic trees for the *Adh* data (1). Haplotypes containing transposable elements are found at the tips of the trees, as illustrated in Fig. 1.

The data also suggest that the observed unique sequence insertions have a very high rate of spontaneous loss. Selection is required to explain the insertions only when it is assumed that the rate of loss of insertions must be smaller than the rate of gain. The pattern of deletions is again a candidate suggesting possible deleterious selection when the deletion "p" is excluded, with $u_3 < u_2$.

Deleterious selection has been suggested for deletions and insertions by Langley *et al.* (10) and Aquadro *et al.* (1). Their arguments were based on the observation that much deletion/insertion variation exists within species but little or none accumulates between closely related *Drosophila* species. Similarly, it has been suggested by Barrie *et al.* (11) that the deletions and insertions of globin genes are deleterious due to their common occurrence within populations but there is an absence of fixations between species. For unique sequence length variation, the selection pressure may arise, in part, from a need to preserve the spacing within and between functionally important sequences. Transposable elements probably cause similar perturbations and a certain component of the selection against them probably arises from this effect. It is also possible that part of the apparent deleterious selection observed for transposable elements in the *Adh* region is due to restrictions on total element copy number (e.g., refs. 12, 13) affecting the fitness of an individual. Restrictions on copy number could arise due to the

adverse effect of an increase in chromosomal rearrangements and other mutations associated with an increase in copy number of a particular family of elements.

In many ways, this analysis is conservative. One feature of the model that demonstrates this is the fact that frequencies do not distinguish individual deletions, insertions, or transposable elements. The analysis does not exclude the possibility that one of the deletions could be positively selected or that one is strongly deleterious. Rather, we have examined the deletions collectively and asked whether they can be explained as a group. The method is additionally conservative because it does not determine whether the data correspond to a model with specific mutation rates. Instead, it finds mutation rates that best conform to the data and finds the minimum amount of selection that can explain the data. If there is some knowledge of bounds on the mutation rates, then a much stronger test could be performed.

As more detailed molecular population genetics data become available for population samples, it will be possible to apply the proposed procedure in a more general and powerful manner, incorporating the described extensions to more than two types of changes and using multiple ancestral sequences. We can expect these future applications to be more powerful in detecting the effects of selection. This increased sensitivity will also require careful investigation of the robustness to recombination, tree construction, and sample size.

We thank Drs. R. R. Hudson, N. Kaplan, and the reviewers for their helpful comments. G.B.G. was supported by Natural Sciences and Engineering Research Council of Canada Grant U0336.

1. Aquadro, C. F., Deese, S. F., Bland, M. M., Langley, C. H. & Laurie-Ahlberg, C. C. (1985) *Genetics*, in press.
2. Laurie-Ahlberg, C. C., Maroni, G., Bewley, G. C., Lucchesi, J. C. & Weir, B. S. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 1073–1077.
3. Felsenstein, J. (1982) *Q. Rev. Biol.* **57,** 379–404.
4. Li, W. H. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 1085–1089.
5. Lewontin, R. C. (1974) *The Genetic Basis of Evolutionary Change* (Univ. of Columbia Press, New York).
6. Wright, S. (1977) *Evolution and the Genetics of Populations* (Univ. of Chicago Press, Chicago, IL), Vol. 3.
7. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, London).
8. Watterson, G. A. (1978) *Genetics* **88,** 405–417.
9. Ewens, W. J. (1979) *Mathematical Population Genetics* (Springer, New York).
10. Langley, C. H., Montgomery, E. & Quattlebaum, W. F. (1983) *Proc. Natl. Acad. Sci. USA* **79,** 5631–5635.
11. Barrie, P. A., Jeffreys, A. J. & Scott, A. F. (1981) *J. Mol. Biol.* **149,** 319–336.
12. Langley, C. H., Brookfield, J. F. Y. & Kaplan, N. (1983) *Genetics* **104,** 457–471.
13. Charlesworth, B. & Charlesworth, D. (1983) *Genet. Res.* **42,** 1–27.