Detecting Non-neutral Heterogeneity Across a Region of DNA Sequence in the Ratio of Polymorphism to Divergence

John H. McDonald
Department of Biology, University of Delaware

Natural selection, in the form of balancing selection or selective sweeps, can result in a decoupling of the amounts of molecular polymorphism and divergence. Thus natural selection can cause some areas of DNA sequence to have greater silent polymorphism, relative to divergence between species, than other areas. It would be useful to have a statistical test for heterogeneity in the polymorphism to divergence ratio across a region of DNA sequence, one that could identify heterogeneity greater than that expected from the neutral processes of mutation, drift, and recombination. The only currently available test requires that a region be arbitrarily divided into sections that are compared with each other, and the subjectivity of this division could be problematic. Here a test is proposed in which runs of polymorphic and fixed sites are counted, where a “run” is a set of one or more sites of one type preceded and followed by the other type. The number of runs is smaller than otherwise expected if polymorphisms are clumped together. By simulating neutral evolution and comparing the observed number of runs to the simulations, a statistical test is possible which does not require any a priori decisions about subdivision.

Introduction

Imagine a set of DNA sequence data from several individuals of species A and from one individual of a related species B. An area of sequence that is small enough that we can ignore recombination has a phylogeny that can be divided into two parts, within-species branches and between-species branches (fig. 1). Any substitution occurring on the within-species branches, those lineages connecting the sequences from species A with their most recent common ancestor, will appear as a polymorphism in the data set. Any substitution on the between-species branches, those lineages connecting species B with the common ancestor of the sequences from species A, will appear as a fixed difference between the species.

For such a tree, five lengths of time can be defined that will be useful in this discussion. The species-tree coalescence time, $T_{c(tot)}$, is the length of time from the present back to the time when the ancestral species split into two descendant species. The within-species coalescence time, $T_{c(w)}$, is the length of time from the present back to the most recent common ancestor of the $n$ sequences from species A. The polymorphism time, $T_{p}$, is the sum of all the lengths of the branches in species A going back to the most recent common ancestor. The between-species coalescence time, $T_{c(b)}$, is the length of time from the present back to the most recent common ancestor of the $n$ sequences from species A and the one sequence from species B. The fixation time, $T_{f}$, is the length of time from the most recent common ancestor of species A, back to the common ancestor of the sequences from the two species, plus all the time from species B to the between-species common ancestor. In other words, $T_{f} = 2T_{c(b)} - T_{c(w)}$.

The neutral theory of molecular evolution predicts that the amount of polymorphism in an area of DNA sequence is determined by the product of the mutation rate, the proportion of mutations that are neutral, and the polymorphism time. The number of fixed differences between two species under the neutral model is determined by the product of the mutation rate, the proportion of mutations that are neutral, and the fixation time. Under the assumption that the mutation rate and the proportion of mutations that are neutral have not changed during the history of the two species, the ratio of expected polymorphisms to expected fixed differences for a particular area of DNA sequence is therefore the ratio of the polymorphism time to the fixation time for that area.

Due to recombination, different areas within a region of DNA sequence may have different phylogenies. They therefore may have different polymorphism and fixation times, and thus different ratios of expected polymorphisms to expected fixed differences. This random phylogenetic variation is in addition to the sampling variation that is caused by observing a small number of substitutions. The phylogenetic and sampling variation in polymorphism to divergence ratio among areas under neutrality can be estimated easily using simulations of a Wright–Fisher model with a constant-rate neutral mutation process (Hudson 1983, 1990).

Natural selection can cause greater heterogeneity in the polymorphism to divergence ratio than expected under the neutral model. Near a balanced polymorphism that is much older than the average coalescence time,
Fig. 1.—Idealized phylogeny of a sample of three sequences from species A and one sequence from species B. The coalescence time for the n sequences from species A, $T_{nca}$ is the length of lineage a. The polymorphism time, $T_p$, is $a + b + c + d$. The coalescence time for species A with species B, $T_{coab}$, is $a + e$. The coalescence time for the 12 sequences from species A with the one sequence from species B, $T_{coab1}$, is $a + e + f$. The fixation time, $T_s$, is $e + f + g + h$.

The within-species branches of the phylogeny will be longer. Thus the area will have a greater number of neutral polymorphisms, and fewer fixed differences, than areas further away that are unaffected by selection (Hudson, Kreitman, and Aguadé 1987). The fixation of an adaptive substitution by natural selection, or selective sweep, has the opposite effect. The coalescence time for linked sites will be approximately the time that the adaptive mutation occurred, so if that time is much more recent than the average coalescence time, there will be fewer polymorphisms and more fixed differences than expected under the neutral model.

A statistical test for comparing the levels of polymorphism between different regions is well established (Hudson, Kreitman, and Aguadé 1987), and this HKA test has found evidence both for balancing selection (Kreitman and Hudson 1991) and for selective sweeps (Berry, Ajioka, and Kreitman 1991; Begun and Aquadro 1992). However, using the HKA test to search for heterogeneity within a region in the ratio of polymorphisms to fixed differences is problematic, since there is no single a priori method for dividing up a region into sections to be compared. Examples of divisions used include coding region versus flanking region (Hudson et al. 1987), three equal-sized sections (Aguadé, Miyashita, and Langley 1992), and five equal-sized sections (Kreitman and Hudson 1991); many other combinations of number, size, and position of sections could be imagined. One possible technique would be to decide, before looking at the data, on a single method of dividing a sequence into sections. The risk is that this would miss something real, by putting a boundary between sections right in the middle of a peak or valley of polymorphism, or by dividing the region into so many sections that there is no statistical power, or by dividing the regions into so few sections that smaller-scale heterogeneity is overlooked. However, if several methods of dividing the region into sections are tried and the results are corrected for the multiple comparisons, the test could become so weak that it would be useless. Choosing a single method of dividing the region after examining the distribution of polymorphisms and diverged sites would, of course, be inappropriate.

Here I describe a test for heterogeneity in the ratio of polymorphisms to fixed differences across a single region that does not require any a priori decisions about the size, position, or number of sections.

Materials and Methods

Several DNA sequences from one species, and one sequence from a closely related species, are aligned. Any areas aligned with gaps, where part of the sequence is not present in all individuals, are omitted. Differences from the consensus sequence are classified into two kinds of substitutions, polymorphisms (nucleotides present only in some sequences from the first species) and fixed differences (nucleotides present only in the second species). The number of runs of polymorphisms and of fixed differences is counted. A run is a set of one or more substitutions of one type (polymorphism or fixed difference) preceded and followed by substitutions of the other type. The first and last sets of like substitutions are also counted as runs.

Where a single nucleotide site has three or four bases, they are treated as if they were at adjacent sites. For example, a site that has C, T, and G in species A would be counted as two polymorphic substitutions. Where a single nucleotide site has both a polymorphism and a fixed difference, they are treated as adjacent sites, and to be conservative they are put in the order with the greater number of runs. For example, if species A has C and T and species B has G at one site, that is counted as one polymorphic site and one fixed site, and if it is preceded by a polymorphism and followed by a fixed site, they would be put in the order PFPF rather than PPFE.

To test whether the number of runs is smaller than that expected from a neutral model, the observed number of runs is compared to the number of runs found in repeated Monte Carlo simulations of a model which includes both sampling and random phylogenetic variation. First, some times are estimated from the observed data, based on the assumption that the population size...
of species A has remained constant at \( N \) diploid individuals. The mean time for \( j \) lineages to coalesce into \( j - 1 \) lineages, measured in \( 2N \) generations, is

\[
E[T_j] = \frac{1}{j} \left( \frac{j}{2} \right) = 2N(j - 1)
\]

(Hudson 1990). Therefore the mean coalescence time for \( n \) sequences is

\[
E[T_{c(n)}] = \sum_{j=2}^{n} \frac{2}{j(j - 1)}
\]

and the mean polymorphism time is

\[
E[T_p] = \sum_{j=2}^{n} \frac{2}{j(j - 1)}
\]

If there are \( S_p \) polymorphic substitutions and \( S_f \) fixed substitutions, the fixation time can be estimated by \( E[T_j] = E[T_{c(n)}]S_f/S_p \). Assuming that the population size of the ancestral species is also \( N \), the expected time from the species-tree coalescence back to the coalescence of the \( n \) sequences from species A with the one sequence from species B is \( 2N \) generations. The time back to the species-tree coalescence can therefore be estimated by \( E[T_{c(species)}] = (E[T_j] + E[T_{c(n)}] - 2) \).

To simulate neutral evolution, the \( m \)-block model of Hudson (1983) is used, with each of the \( m \) blocks being a single nucleotide site. A value of \( R \) is chosen, where \( R = 4Nr \) and \( r \) is the recombination rate per generation between the most distant sites. A constant population size of \( N \) diploid individuals is assumed for species A, species B, and the ancestral species. Coalescence and recombination are simulated for the \( n \) sequences of species A from the present back to the estimated species split time. Coalescence and recombination are likewise simulated for the one sequence of species B from the present back to the species-tree coalescence. Coalescence and recombination are simulated within the ancestral species from the species-tree coalescence back until all the sequences have coalesced into one sequence at each nucleotide site. (Note that if all the sequences from species A have not coalesced by the species-tree coalescence time, it is possible for the sequence from species B to coalesce with some sequences from species A more recently than the sequences from species A coalesce with each other.)

After coalescence and recombination have been simulated, the polymorphism time and fixation time for each site are recorded. Total polymorphism time is defined as the sum across all sites of the polymorphism times; total fixation time is defined the same way. The observed number of polymorphisms is distributed randomly among the sites, with the probability of a polymorphism being placed at a site being equal to the fraction of the total polymorphism time found at that site. Fixed differences are distributed the same way. Each simulation thus ends up with the same number of polymorphisms and fixed differences as the observed data, a reasonable approach when the observed data are used to estimate parameters (Hudson 1993). If a site contains more than one substitution, the substitutions are ordered randomly. Finally, the number of runs of polymorphisms and fixed differences is counted for the simulated data set.

A large number of simulations is conducted, and the proportion of the simulations in which the number of runs is equal to or less than the observed number is counted. This proportion, \( P_R \), is the estimated probability of getting the observed or fewer number of runs under a neutral model with a recombination parameter of \( R \). To be conservative, different values of \( R \) are tried, starting with 0 and increasing until \( P_R \) starts to decrease. The largest \( P_R \) is used as the statistical significance of the deviation of the observed data from a neutral model. Note that this is a one-tailed test, which is appropriate since there is no known biological process that will produce a more even distribution of polymorphisms than expected under neutrality.

A Macintosh computer program that will perform these simulations, DNA Runs, is available by anonymous FTP from ftp.bio.indiana.edu in the directory /molbio/mac. It is also available from the author.

In order to test the computer program, simulations of a sequence 201 bases long were carried out for several values of \( R \), and parameters for which there are theoretical expectations were recorded: the mean and variance in polymorphism time at a single site, the covariance of polymorphism time between two sites (sites between these two sites.

To demonstrate the use of the test, published data on polymorphism and divergence between *Drosophila melanogaster* and *Drosophila simulans* were analyzed (table 1). Only silent substitutions (synonymous and noncoding) were used. When there were multiple sequences from both species, it was necessary to choose one sequence from the outgroup species. If the original authors chose an outgroup sequence, it was used. Otherwise, for convenience, the outgroup sequence that was adjacent to the other species' sequences in the published figure was used. Simulations were run with \( R = 1, 2, 4, 8, 16, \ldots \) until the simulated probability of obtaining the observed number of runs or fewer started to decline. For each value of \( R \), 1,000 simulations were run. When any initial run indicated that \( P \) was less than 0.10, an independent set of 10,000 simulations was run, to obtain a somewhat more accurate estimate of the probability.
Table 1
Application of the Runs Test to Sequence Polymorphism and Divergence in *Drosophila melanogaster* and *D. simulans*

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>S_p</th>
<th>S_f</th>
<th>K</th>
<th>L</th>
<th>P</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>D. mel.</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>boss</td>
<td>7</td>
<td>15</td>
<td>30</td>
<td>21</td>
<td>1,565</td>
<td>0.668</td>
<td>Ayala and Hartl 1993</td>
</tr>
<tr>
<td>Est</td>
<td>13</td>
<td>36</td>
<td>47</td>
<td>41</td>
<td>1,755</td>
<td>0.594</td>
<td>Cooke and Oakeshott 1989</td>
</tr>
<tr>
<td>G6pd</td>
<td>32</td>
<td>26</td>
<td>37</td>
<td>34</td>
<td>1,705</td>
<td>0.831</td>
<td>Eanes, Kirchner, and Yoons 1993</td>
</tr>
<tr>
<td>Mlc1</td>
<td>16</td>
<td>21</td>
<td>22</td>
<td>15</td>
<td>991</td>
<td>0.028</td>
<td>Leicht et al. 1995</td>
</tr>
<tr>
<td>Ms120A</td>
<td>10</td>
<td>22</td>
<td>43</td>
<td>23</td>
<td>1,257</td>
<td>0.067</td>
<td>Agudáé, Miyashita, and Langley 1992</td>
</tr>
<tr>
<td>per</td>
<td>6</td>
<td>28</td>
<td>54</td>
<td>34</td>
<td>1,869</td>
<td>0.342</td>
<td>Kliman and Hey 1993</td>
</tr>
<tr>
<td>Sod</td>
<td>41</td>
<td>54</td>
<td>59</td>
<td>49</td>
<td>1,410</td>
<td>0.116</td>
<td>Hudson et al. 1994</td>
</tr>
<tr>
<td>AdhlAdhr</td>
<td>11</td>
<td>82</td>
<td>157</td>
<td>83</td>
<td>4,720</td>
<td>0.004</td>
<td>Kreitman and Hudson 1991</td>
</tr>
<tr>
<td><em>D. sim.</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>boss</td>
<td>5</td>
<td>39</td>
<td>23</td>
<td>34</td>
<td>1,565</td>
<td>0.959</td>
<td>Ayala and Hartl 1993</td>
</tr>
<tr>
<td>Est</td>
<td>4</td>
<td>113</td>
<td>99</td>
<td>102</td>
<td>2,998</td>
<td>0.752</td>
<td>Karotam, Boyce, and Oakeshott 1995</td>
</tr>
<tr>
<td>G6pd</td>
<td>12</td>
<td>16</td>
<td>37</td>
<td>24</td>
<td>1,705</td>
<td>0.709</td>
<td>Eanes, Kirchner, and Yoons 1993</td>
</tr>
<tr>
<td>per</td>
<td>6</td>
<td>46</td>
<td>43</td>
<td>40</td>
<td>1,869</td>
<td>0.356</td>
<td>Kliman and Hey 1993</td>
</tr>
<tr>
<td>prune</td>
<td>4</td>
<td>12</td>
<td>33</td>
<td>15</td>
<td>1,395</td>
<td>0.224</td>
<td>Simmons et al. 1994</td>
</tr>
</tbody>
</table>

Note.—*n*, number of sequences; *S_p*, number of polymorphisms; *S_f*, number of fixed differences; *K*, number of runs; *L*, length of the sequence in nucleotides; *P*, proportion of the simulations, when done with the recombination parameter yielding the greatest *P*, that had *K* or fewer runs.

To compare the power of the HKA test (Hudson, Kreitman, and Aguadé 1987) to the runs test proposed here, data on the AdhlAdhr region (Kreitman and Hudson 1991) were analyzed. The sequence was divided up into two, three, four, five, or six equal-sized sections, the silent polymorphisms and diverged sites within each section were counted, and the data were subjected to the HKA test. "Diverged sites" were those that differed between the *Was* allele of *D. melanogaster* and the *D. simulans* sequence. Two kinds of "equal sized sections" were used: each section containing the same number of nucleotide sites, or each section containing approximately the same number of silent substitutions.

Results

The distribution of polymorphic and fixed substitutions in proportion to the polymorphism time and fixed time at each site is simple, but simulating these times is sufficiently complex that some verification is warranted. An approximate equation suggested by Hudson (1983) for covariance of polymorphism time at two sites,

$$\text{Cov}_{n}[T_{n(1)}, T_{n(2)}] = \left( \sum_{i=1}^{n-1} \frac{1}{i^2} \right) \frac{R + 18}{R^2 + 13R + 18},$$

is quite accurate for values of *R* less than 10 (Kaplan and Hudson 1985). Because this covariance is a function of both the sample size and the recombination rate, it is a sensitive test of the computer algorithm. Results from simulations using the algorithm are in excellent agreement with the theoretical expectation (fig. 2). The mean and variance in polymorphism time at a single site, and the correlation of polymorphism time between two sites, also exhibited excellent agreement between simulations and theoretical expectations (results not shown).

As the recombination parameter used in the simulations is increased, the probability of getting the observed or fewer number of runs increases at first (fig. 3). A greater recombination rate means that there will be less correlation between the phylogenies of different sections of the sequence, and thus there is an increasing chance that one part of the sequence will have an unusually large polymorphism time and contain a cluster of polymorphisms. Eventually the probability of getting an observed or fewer number of runs decreases, as the recombination rate increases to the point where a section of sequence with an unusually long polymorphism time would be so small that it would contribute little to the clumpiness of the polymorphisms. For all the data sets
tested, there is a fairly broad range of values of $R$ that yield approximately the same $P_R$. Therefore the exact value of $R$ that is simulated does not seem critical to the accuracy of the results.

Of published data sets on DNA sequence polymorphism and divergence in *D. melanogaster* and *D. simulans*, two have significantly fewer runs than expected under neutrality. One is the Adh1Adhr region in *D. melanogaster*, where there is a cluster of silent polymorphisms around the balanced F/S polymorphism in *Adh* and relatively few polymorphisms in *Adhr* (Kreitman and Hudson 1991). The second significant result is the Mlc1 locus in *D. melanogaster*. There, polymorphisms appear to be clustered in the middle of the sequenced region (fig. 4a). The Mst26A region in *D. melanogaster* does not quite meet the $P < 0.05$ criterion for significance, but the clustering of polymorphisms at the 3' end of the region is intriguing (fig. 4b); further exploration of this possible peak of polymorphism may be warranted.

**Discussion**

The neutral model simulated by this statistical test assumes there is no selection. If the results indicate that there are significantly fewer runs in the observed data than expected, it is tempting to interpret this as evidence for natural selection. Before accepting this interpretation, it is important to determine whether deviations from the other assumptions of the neutral model could also cause a reduced number of runs. Here I examine the assumptions of the neutral model.

**No multiple hits.** The simulations do not allow multiple polymorphisms at a single nucleotide site.

**Constant mutation rate:** If the mutation rate dramatically decreased at just the right moment, between...
the longest and shortest coalescence times for sections, those sections with longer coalescence times might have even greater polymorphism than expected. However, there is little evidence for the kind of dramatic changes in mutation rate through short periods of time that this would require.

**Uniform recombination rate:** The model assumes that there is an equal recombination rate between each adjacent pair of nucleotides. A model of a few discrete blocks of sequence with no recombination within blocks and free recombination between blocks would maximize the random phylogenetic heterogeneity among blocks and therefore would be more conservative (Hudson, Kreitman, and Aguadé 1987). Simulations with this extreme model yielded only a slight increase, generally 1% or 2%, in the probability of getting the observed or fewer number of runs with the data from table 1 (results not shown). Moderate heterogeneity in the recombination rate should therefore have little effect on the accuracy of the test.

**Random union of gametes:** Geographic subdivision of the species might produce increased heterogeneity, but only under the following conditions: species A had been two isolated populations; these populations were well isolated enough that divergence could build up between them; they were isolated for much longer than the coalescence time for each population; the populations have merged recently, but long enough ago that intragenic recombination has occurred; and the sample just happens to include sequences in which the phylogeny of some sections of sequence includes the second population, and the phylogeny of other sections does not. If all this is true, those sections of sequence where the phylogeny does include the second population will have more polymorphisms and fewer fixed differences than other sections. This source of heterogeneity, while not unimaginable, requires a lot of coincidences, and its effects would likely be seen at many regions throughout the genome.

**Constant population size:** A recent increase in population size would actually decrease the heterogeneity across a region, since all phylogenies would be likely to have coalescence times around the time of the expansion. A recent decrease in population size could increase heterogeneity, but only if it was at such a time that the phylogenies of some sections of sequence had coalescence times since the population crash, but other sections had coalescence times back in the time of the larger population. This would also seem to require a delicate balance between population size and time, and it would also affect many regions.

**Point mutations only:** Occasionally, one sees what appears to be the substitution of two or more adjacent bases by a different set of bases. If this represents a single mutational event, counting it as several point substitutions could artificially inflate the number of polymorphisms or fixed sites and thus increase the expected number of runs. The results of the test should be interpreted with caution if the data set contains a number of these possible multiple-base substitutions.

**Silent substitutions are neutral:** Although this test does not assume that all silent mutations are neutral, it does assume that those silent substitutions that are seen in a data set are neutral. There is increasing evidence that this assumption is incorrect, in that the fate of many synonymous mutations is at least partly determined by selection (Akashi 1994, 1995). Substitutions that change a favored codon to an unfavored codon may reach intermediate frequencies if they are weakly selected against, but they are unlikely to become fixed. Areas of coding region where this is occurring thus might have a higher ratio of polymorphism to divergence than introns and flanking regions. Conversely, a species that has recently expanded in population size might have eliminated mildly deleterious polymorphic synonymous substitutions, while these could have accumulated as fixed differences when the population size was smaller. This could lead to a smaller ratio of polymorphism to divergence in coding regions, compared with introns and flanking regions. Further research will be needed to determine how plausible it is that weak selection on synonymous mutations could result in a significant runs test. For now, this kind of weak selection on synonymous sites cannot be ruled out as a possible cause of statistically significant clustering of polymorphic sites.

Since a statistical test already exists for examining heterogeneity in the polymorphism to divergence ratio (Hudson, Kreitman, and Aguadé 1987), it is important to determine whether the runs test proposed here offers any improvement in the power to detect selection. When the Adh/Adhr region in *D. melanogaster* is divided up into sections in a number of simple, obvious ways and each set of sections is subjected to the HKA test, none of the test results is significant (table 2). The runs test proposed here does yield a significant result (table 1).

---

**Table 2**

<table>
<thead>
<tr>
<th>Equal Sites per Section</th>
<th>Equal Substitutions per Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sections</td>
<td>$P$</td>
</tr>
<tr>
<td>2</td>
<td>0.86</td>
</tr>
<tr>
<td>3</td>
<td>0.75</td>
</tr>
<tr>
<td>4</td>
<td>0.65</td>
</tr>
<tr>
<td>5</td>
<td>0.28</td>
</tr>
<tr>
<td>6</td>
<td>0.28</td>
</tr>
</tbody>
</table>
There are many other ways of dividing the Adh/Adhr region into sections, some of which do yield significant HKA tests (Kreitman and Hudson 1991). However, the failure of several obvious subdivisions to reveal the heterogeneity in the polymorphism to divergence ratio suggests that the runs test may be more powerful in at least some situations.

Of the data sets analyzed here, only the Adh/Adhr and the Mlc1 region in D. melanogaster exhibit significant heterogeneity in the polymorphism to divergence ratio. In the Adh/Adhr region, silent polymorphisms are clustered around the replacement substitution (Kreitman and Hudson 1991). Patterns of geographic variation also indicate that the F/S polymorphism is affected by selection (Oakeshott et al. 1982; Berry and Kreitman 1993). The object of selection in the Mlc1 locus is not obvious, since there is no amino acid polymorphism in the region sequenced. Levels of interspecific sequence conservation within introns of this gene suggest that the introns may play an important regulatory role (Leicht et al. 1995), so it is possible to speculate that there may be balancing selection on some of the sequence variation within introns. Another possible explanation for the pattern in Mlc1 would be selective sweeps near the ends of the sequenced region.

There was prior evidence suggesting the effects of balancing selection on G6pd (Eanes, Kirchner, and Yoon 1993) and Sod (Hudson et al. 1994) in D. melanogaster, and Est-6 in both D. melanogaster and D. simulans (Karotam, Boyce, and Oakeshott 1995), but none of these exhibit significant heterogeneity in the polymorphism to divergence ratio. If they are under balancing selection, these amino acid polymorphisms may have arisen too recently to have accumulated excess silent polymorphism at linked sites, as suggested by the lack of silent DNA variation within a single locus for selection. It is also possible that the recombination rate around these genes is so low that the scale of heterogeneity is larger than the sequenced regions, or that the recombination rate is so high that any evidence of selection has been erased.

Acknowledgments

I thank A. J. Berry, V. Dilegge, L. Geyer, R. R. Hudson, A. Hunt, C. Maier, and M. Kreitman for their detailed and insightful comments on the manuscript.

Literature Cited


the myosin alkali light chain in *Drosophila*. Genetics **139**: 299–308.


STANLEY SAWYER, reviewing editor

Accepted August 30, 1995