This paper concludes that the statistical properties of protein evolution are compatible with a particular model of evolution by natural selection. The argument begins with a statistical description of the molecular clock based on a Poisson process with a randomly varying tick rate. If the time scale of the change of the tick rate of the molecular clock is assumed to be much less than the average time between substitutions, then it is shown that the substitution process must be episodic, with bursts of substitutions being separated by long periods of time with no substitutions. This analysis generalizes the recent work of Gillespie (1984a). The second part of the argument shows that a simple model of evolution by natural selection—one that incorporates a changing environment, the molecular landscape, and a simple form of epistasis—exhibits dynamics that are identical to those inferred from the statistical analysis. This leads to the conclusion that natural selection is a viable explanation for protein evolution. In addition, a correction formula for multiple substitutions is given that does not require that the substitution process be a Poisson process, and some comments on the inability of the neutral allele theory to account for the dynamics of the substitution process are presented.

Introduction

The forces responsible for the evolution of proteins remain a mystery despite many years of intensive study and debate. The two most conspicuous theories to account for molecular evolution are the neutral allele theory (recently reviewed by Kimura [1983]) and a less precisely defined theory of evolution by natural selection. From a mathematical point of view, the neutral allele theory is the more thoroughly studied and, partly for this reason, is the more generally cited as the probable cause of many of the differences that are observed in the amino acid sequences both between and within species. One of the most convincing arguments for neutrality and against natural selection has been the near constancy of the evolutionary rates of proteins. Kimura, in particular, has repeatedly argued that natural selection would lead to different rates of evolution in different lineages whereas neutrality would give similar rates. Thus, the issue of the mechanism of evolution is intimately connected with the variability of the rates of molecular evolution.

Despite the central importance of the variability of rates of evolution on our understanding of the mechanism of evolution, there appears to be very little published work that attempts to estimate the variability of rates using sequence data. When this has been attempted, as in Ohta and Kimura (1971), a naive model of rate variation is usually invoked. In this model the variation in the rate of evolution is due to dif-

1. Key words: molecular evolution, protein evolution, doubly stochastic Poisson process, neutral allele theory, molecular clock, natural selection, rates of evolution, mutational landscape.
ferences in rates between the lineages under study rather than to variation in rates within a lineage. This model allows us to infer the variability of the rates of evolution between lineages in a relatively straightforward fashion and has led to the generality that the rates of evolution do not exhibit much variation.

Recently, Gillespie (1984a) examined a somewhat different model of rate variation, one in which the rates of substitutions are allowed to vary through time down each of the lineages under study. Such a model allows variation both between and within lineages. This seemingly innocent generalization led to a complex statistical analysis that suggests that the variability of rates of evolution may be much larger than previously suspected. Paradoxically, this model has the property that large variabilities in rates of evolution are not necessarily manifest in large variabilities in the numbers of substitutions in different lineages. With a few additional assumptions, this analysis went even further to suggest that molecular evolution may be an episodic process, with bursts of substitutions being separated by long periods without any substitutions. It is this property of molecular evolution that suggested that the molecular clock be called an "episodic clock."

These results are purely statistical, i.e., they do not depend in any way on a particular mechanism of molecular evolution. They are, however, strongly suggestive of a model of molecular evolution by natural selection that will be described in this paper. The model is particularly attractive because it exhibits the same episodic dynamics that were described in the statistical analysis. It is a model that appears to fit the statistical aspects of the sequence data as well as does the neutral allele model and that shares with that model a very appealing biological description.

Unlike the neutral allele model, the most important properties of this model of molecular evolution by natural selection cannot be described in a couple of simple equations. To make the results accessible to those unfamiliar with the requisite mathematics, the next section will provide an overview of the remainder of the paper. One could even view the next section as the paper proper and all of the subsequent sections as extended appendices.

Overview

The present paper consists of two logically independent parts: a statistical part that provides a description of the properties of the molecular clock and a mechanism part that describes a model of molecular evolution by natural selection. The main observation of the paper is that both of these parts arrive at the same description of the stochastic dynamics of molecular evolution. From this we conclude that natural selection is compatible with the statistical aspects of the sequence data.

The statistical part of the paper is further subdivided into two parts: the first part derives a very general formula for correcting sequence data for multiple substitutions at a site, and the second part generalizes the statistical procedure of Gillespie (1984a) for characterizing the episodic nature of the clock. The data for this statistical investigation come from "star" or "bush" phylogenies. A star phylogeny is one in which the radiation into a series of \( n \) species has occurred in a small period of time relative to the length of the lineages. Kimura (1983) was the first to recognize the value of star phylogenies for estimating the statistical properties of the substitution process. The best-known example of a star phylogeny is the mammalian radiation.

The object of study is \( N_i(t) \), the number of substitutions that occurred on the \( i \)th lineage of a radiation that occurred \( t \) years ago. Our goal is to estimate the mean and variance of the \( N_i(t) \): \( \mu = E[N_i(t)] \), \( \sigma^2 = \text{Var}[N_i(t)] \), where \( E \) stands for "expected value.
of." The data at hand for the estimation are the observed number of amino acids that differ between species \( i \) and \( j \), \( D_{ij} \).

To estimate the moments of \( N_i(t) \) using the \( D_{ij} \), corrections must be made for the fact that the \( D_{ij} \) are correlated with one another and for the occurrence of multiple substitutions at a site. The details of this correction are given in the next section. The results are as follows. Let \( M \) be Kimura's statistic that is one-half the average value of the \( D_{ij} \):

\[
M = \frac{1}{n(n-1)} \sum_{i<j} D_{ij}. \tag{1}
\]

Let \( S^2 \) be Kimura's statistic that is proportional to the variance in the \( D_{ij} \):

\[
S^2 = \frac{1}{(n-1)(n-2)} \sum_{i<j} (D_{ij} - 2M)^2. \tag{2}
\]

The results of the next section suggest that an estimator for the mean number of substitutions per lineage, \( \mu \), that corrects for multiple substitutions at a site, is given by

\[
\hat{\mu} = \log\left(1 - \frac{2M}{m}\right)/2 \log\left(1 - \frac{1}{m}\right). \tag{3}
\]

The estimator for the variance, \( \sigma^2 \), is given by

\[
\hat{\sigma}^2 = (S^2 - VC)(1 - \frac{1}{m})^{-\hat{\mu}}. \tag{4}
\]

In these expressions, \( m \) is the number of amino acids in the protein and \( VC \) is a small correction for compulsive types. When \( VC \) is set equal to zero, an error of no more than 5% will result in typical protein data. This error is much smaller than the sampling variance. Finally, much of the subsequent discussion will focus on the ratio of these two estimators:

\[
\hat{r} = \frac{\hat{\sigma}^2}{\hat{\mu}}, \tag{5}
\]

which is an estimator of the ratio of the variance to the mean of the number of substitutions per lineage,

\[
R(t) = \frac{\text{Var}\{N_i(t)\}}{\text{E}\{N_i(t)\}}. \tag{6}
\]

These estimators are used on actual data in table 2. Of particular interest in this table are the estimates of \( R(t) \). It has generally been assumed (incorrectly) that, under the neutral allele model or for constant rates of evolution, \( R(t) \) will equal one. This is because in both of these situations the \( N_i(t) \) are supposed (again incorrectly) to be Poisson distributed. When various authors (e.g., Ohta and Kimura 1971; Langley and Fitch 1974; Kimura 1983) have claimed to show that the rates of evolution are not constant, they have really shown that \( R(t) \) is significantly larger than one or that the \( N_i(t) \) are not Poisson distributed. Thus most of the discussion over the variability of rates is really a discussion about the values of \( R(t) \). As in other work, our results show that \( R(t) \) is always larger than one. The striking aspect of the estimates, however, is the narrow range of \( R(t) \) values: 1.0 < \( R(t) < 3.4 \). This is an awkward result: if one favors neutrality, then one must account for the fact that \( R(t) \) is larger than one (perhaps...
by variable rates); if one favors natural selection, then one must account for the fact that \( R(t) \) is not larger than 3.4. It is the latter problem that will be solved in this paper.

A statistical description of the substitution process with a fluctuating rate of evolution is presented in the second of the two statistical sections. The aim of this section is to generalize the results of Gillespie (1984a). The starting point is a process called a doubly stochastic Poisson process, or Cox process, which is used as the statistical model for the substitution process, \( N_s(t) \). This process is just a Poisson process with a randomly changing rate of evolution. One major assumption is introduced in this section: the rate of change of the rate of molecular evolution is assumed to occur on a time scale that is much shorter than the length of the lineages under study. A typical lineage as used in table 2 is tens to hundreds of millions of years long. If the changes in the rates of evolution occur on a time scale that is on the order of thousands to tens of thousands of years—the time scale of the recent ice ages—then this assumption will be met. Given this assumption, the data support the view that the substitution process is an episodic process with bursts of substitutions followed by long periods with no substitutions. Table 3 gives estimates for the mean number of episodes and for the mean number of substitutions per episode for six proteins using two particular examples of the Cox process.

It is also shown in this section that the substitution process may be represented as a Poisson sum of positive random variables:

\[
N(t) = Y_1 + Y_2 + \ldots + Y_{M(t)}.
\]  

In this representation the \( Y_i \) represent the number of substitutions that occurred in the \( i \)th episode of evolution and \( M(t) \) represents the number of episodes that occurred in the lineage during a time interval of \( t \) generations. In this formula, \( M(t) \) is a Poisson process. This representation formula is the main result of the statistical portion of the present paper. It gives a representation of the substitution process that will be used as a guide for the development of a model of molecular evolution by natural selection in the subsequent section.

The following section of the paper presents a model of molecular evolution by natural selection whose dynamics may also be represented by a formula of the form given above in equation (7). The model has three components: (1) a changing environment, (2) an epistatic scheme in which each environmental change presents a challenge to the species that may be met by substitutions at one of several nearly equivalent loci, and (3) a mutational landscape that makes it unlikely that any particular locus will experience the substitution of an allele that is two mutational steps away from the allele that is currently fixed in the population. To achieve the representation of equation (7), we must assume that the rate of environmental change is large and that the number of loci that are capable of responding to each environmental change is also large. A remarkable property of this model is that the values of \( R(t) \) that it predicts are in a very restricted range, say from 1.0 to \( \sim 3.5 \), just as observed in the data. This agreement, and the fact that the dynamics are compatible with the episodic structure of the data, show that a very plausible model of molecular evolution by natural selection fits the sequence data just as well as does the neutral allele model. Various caveats about the interpretation of these results are given in the general discussion at the end of this paper.
Correcting for Multiple Substitutions

In this section a correction formula for multiple substitutions in a single amino acid will be derived that does not require the number of substitutions per lineage to be Poisson distributed. The basic setting is that of a star phylogeny in which $n$ lineages radiate from a single common ancestor at some point in the remote past. We will represent each of the $m$ amino acids in a protein in each of the $n$ extant species under study by an urn and each of the $N_i$ amino acid substitutions that occur in the $i$th lineage by a ball (in this section the argument $t$ in $N_i[t]$ will be suppressed). If $N_i$ substitutions have occurred in the $i$th lineage, then we throw $N_i$ balls randomly into the $m$ urns for that lineage with the probability of a ball going into the $k$th urn being $p_k$. The number of balls in a lineage will be viewed as a random quantity. Models of molecular evolution frequently assume that $N_i$ is Poisson distributed. Here we will assume only that the $N_i$, $i = 1, \ldots, m$, are independent, identically distributed, non-negative, integer-valued, random variables. The number of substitutions that separate two species is then the total number of balls found in their two sets of urns. Note that the assumption of independence of the $N_i$ is a strong assumption that may be violated if the autocorrelation of the rate of molecular evolution is high relative to the length of the lineages.

The statistical problem arises because it is possible for two or more balls to land in the same urn. As experimentalists we can only say that a particular amino acid site differs between two species. We cannot know if that difference arises from one, two, or even more substitutions at a site. Thus we need to infer the number of substitutions that occurred when all that we know is the number of sites in which at least one substitution occurred. In terms of the urn model, we would like to infer the number of balls by observing the number of occupied urns. In fact, this goal is too ambitious. All that we can do is to estimate the mean and variance of $N_i$ using the mean and variance of the number of occupied urns.

Let $D_{ij}$ represent the number of urns that are occupied in either species $i$ or species $j$: that is, to calculate the contribution of the $k$th urn to $D_{ij}$ we examine urn $k$ in species $i$ and species $j$ and increment $D_{ij}$ if the urn is occupied in either or both of the two species. $D_{ij}$ is therefore a reasonable representation of the number of amino acids that are observed to be different between two species. It is not a perfect representation because we are not allowing the possibility that if two or more balls land in the same urn then a back substitution may occur. While this possibility could be included in the model, its occurrence in the data to be examined is so rare that ignoring it will not introduce any significant bias into the analysis.

Kimura (1983) suggested the two summary statistics based on $D_{ij}$, $M$ and $S^2$, given in equations (1) and (2) of the previous section. His development differed somewhat from ours in that he first transformed the observed number of differences between two species, $D_{ij}$, by a function that is supposed to correct for multiple substitutions at a site. From that point on he assumed that $D_{ij} = N_i + N_j$. In our development the correction will be carried out after the summary statistics are calculated. Our primary goal is to use these statistics to estimate the mean and variance of $N_i$. This will be accomplished by first evaluating the expectations of $M$ and $S^2$ and then using these results to suggest suitable estimators.

In urn models it is often more convenient to describe the number of unoccupied urns rather than the number of occupied urns. Thus we will let $U_{ij}$ represent the number of urns that are unoccupied in both species $i$ and $j$:
The moments of the statistics $M$ and $S^2$, expressed in terms of the moments of $U_{ij}$, are

\[ E(M) = [m - E(U_{ij})]/2, \]
\[ E(S^2) = (n + 1)[2(n - 1)]^{-1}\text{Var}(U_{ij}) - 2(n - 1)^{-1}\text{Cov}(U_{ij}, U_{ik}). \]

Thus the task before us is to find approximations to the moments of the $U_{ij}$, plug these into the expressions for the expectations of $M$ and $S^2$, and from these guess some good estimators of the moments of $N_i$. (The details of the approximations to the moments of $U_{ij}$ are complex, tedious, and inappropriate for this journal; they will not be given here. The author will supply the details of the calculations to anyone who requests them.)

The approximations are

\[ E(U_{ij}) \approx m(1 - 1/m)^2u, \]
\[ \text{Var}(U_{ij}) \approx 2\sigma^2(1 - 1/m)^4u, \]
\[ \text{Cov}(U_{ij}, U_{ik}) \approx \sigma^2(1 - 1/m)^4u. \]

In arriving at these approximations, we took as our main assumption that the variation in the $p_k$, i.e., the probability of a substitution in the $k$th urn—is not excessive. To be precise, if we write $p_k = (1 + a_k)/m$, then we require that the sum

\[ m^{-1} \sum_i a_i^2 \]

be around one or less. In using this formula it may be necessary to exclude some sites that are known not to evolve.

Using these approximations, we get

\[ E(M) \approx m(1 - 1/m)^3u/2, \]
\[ E(S^2) \approx \sigma^2(1 - 1/m)^4u, \]

which immediately suggest the estimators

\[ \hat{\mu} = \log(1 - 2M/m)/2 \log(1 - 1/m), \]
\[ \hat{\sigma}^2 = S^2(1 - 1/m)^{-4u}, \]
\[ \hat{r} = \hat{\sigma}^2/\hat{\mu}. \]

At the level of approximation used in arriving at these estimators we have

\[ E(\hat{\mu}) \approx E[N_i(t)], \]
\[ E(\hat{\sigma}^2) \approx \text{Var}[N_i(t)]. \]

In deriving these approximations we assumed that the variance introduced by
the random placement of balls in urns is small relative to the variance introduced by the variation in $N_i$. In some cases, when $\mu$ and $\sigma^2$ are small, it may be necessary to take account of this extra element of variation. This is accomplished by using the estimator

$$\hat{\sigma}^2 = (S^2 - VC)(1 - 1/m)^{-4\mu},$$  \hspace{1cm} (15)$$

where $VC = (n + 1)[2(n - 1)]^{-1}C_1 - 2(n - 1)^{-1}C_2$, and $C_1 = m(1 - 1/m)^{3\mu} + m^2(1 - 1/m)^{4\mu} + m(m - 1)(1 - 2/m)^{2\mu}$, $C_2 = m(1 - 1/m)^{3\mu} - m^2(1 - 1/m)^{4\mu} + m(m - 1)(1 - 1/m)^{2\mu}(1 - 2/m)^{\mu}$. Fortunately, this correction is unnecessary for most data.

Because of the various approximations and biases that are inherent in these estimators, they must be checked for accuracy by simulations. Table 1 gives the results for a protein that has 150 amino acids with the mean number of substitutions per lineage being 14. This is very typical of the actual proteins that will be analyzed. As is obvious, the estimators are very good up to an $R(t)$ of approximately 5–6. Even here the bias in estimating $R$ is only ~5%. The real data fall in the range 1.5 < $R(t)$ < 3.5, so the bias in the estimate of $R(t)$ can safely be ignored. The origin of the bias is from the covariance in $M$ and $S^2$ that is generated by the variability in $N_i$. While a correction for the bias could be derived, the computations are very complex for the small gain in precision. This point of view is reinforced by examining the SD of the estimator of $\hat{r}$ in table 1. The SD exceeds by a large amount any correction that might be undertaken for $\hat{r}$.

In table 2 the estimators are used to examine actual data. A comparison of the estimators obtained by our method with those obtained by Kimura (1983) shows that there is very little difference between them. This is surprising, since the justification

---

### Table 1

Simulation Results Illustrating Accuracy of Estimators for Moments of $N_i$

<table>
<thead>
<tr>
<th>$\mu$</th>
<th>$R(t)$</th>
<th>Species</th>
<th>$M$</th>
<th>$S^2$</th>
<th>$\hat{\mu}$</th>
<th>$\hat{\sigma}^2$</th>
<th>$\hat{r}$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>0</td>
<td>4</td>
<td>12.81</td>
<td>1.27</td>
<td>14.00</td>
<td>0.00</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>6</td>
<td>12.81</td>
<td>1.14</td>
<td>14.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>4</td>
<td>12.76</td>
<td>10.93</td>
<td>13.97</td>
<td>14.12</td>
<td>1.00</td>
<td>0.90</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>6</td>
<td>12.78</td>
<td>10.76</td>
<td>13.98</td>
<td>14.04</td>
<td>1.00</td>
<td>0.67</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>4</td>
<td>12.73</td>
<td>20.49</td>
<td>13.95</td>
<td>28.35</td>
<td>2.00</td>
<td>1.71</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>6</td>
<td>12.72</td>
<td>20.32</td>
<td>13.93</td>
<td>28.14</td>
<td>2.00</td>
<td>1.82</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>4</td>
<td>12.67</td>
<td>29.85</td>
<td>13.93</td>
<td>42.52</td>
<td>2.98</td>
<td>2.50</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>6</td>
<td>12.67</td>
<td>29.6</td>
<td>13.88</td>
<td>41.93</td>
<td>2.98</td>
<td>1.93</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>4</td>
<td>12.64</td>
<td>38.89</td>
<td>13.89</td>
<td>56.36</td>
<td>3.93</td>
<td>3.31</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>6</td>
<td>12.67</td>
<td>38.86</td>
<td>13.90</td>
<td>55.95</td>
<td>3.94</td>
<td>2.53</td>
</tr>
<tr>
<td>14</td>
<td>6</td>
<td>4</td>
<td>12.57</td>
<td>57.13</td>
<td>13.85</td>
<td>85.06</td>
<td>5.83</td>
<td>4.88</td>
</tr>
<tr>
<td>14</td>
<td>6</td>
<td>6</td>
<td>12.58</td>
<td>56.94</td>
<td>13.81</td>
<td>84.37</td>
<td>5.85</td>
<td>3.78</td>
</tr>
</tbody>
</table>

**Note.**—Each number represents the average of 50,000 replicates. For $R = 1$ a Poisson distribution for $N_i$ was used; for $R > 1$ a Polya-Aeppli distribution was used. $\mu$ = true mean of the number of substitutions per lineage; $R(t)$ = true value of the variance to mean ratio of the number of substitutions per lineage; $M$ and $S^2$ estimators given by eq. (1) and (2), respectively; $\hat{\mu}$ = estimated mean number of substitutions per lineage; $\hat{\sigma}^2$ = estimated variance in the number of substitutions per lineage; $\hat{r}$ = estimated value of $R(t)$; and $\sigma$ = average SE of estimator of $R(t)$. 

Downloaded from https://academic.oup.com/mbe/article-abstract/3/2/138/981135 by guest on 03 April 2019
Table 2

Estimates of Moments of $N(t)$ for Six Proteins

<table>
<thead>
<tr>
<th>Protein*</th>
<th>$N$</th>
<th>$m$</th>
<th>$M$</th>
<th>$S^2$</th>
<th>$\hat{\mu}$</th>
<th>$\hat{\sigma}^2$</th>
<th>$\hat{r}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$-Hemoglobin</td>
<td>6</td>
<td>146</td>
<td>13.43</td>
<td>31.19</td>
<td>14.79</td>
<td>44.93</td>
<td>3.04</td>
</tr>
<tr>
<td>$\alpha$-Hemoglobin</td>
<td>6</td>
<td>141</td>
<td>11.57</td>
<td>11.29</td>
<td>12.59</td>
<td>14.74</td>
<td>1.17</td>
</tr>
<tr>
<td>Myoglobin</td>
<td>6</td>
<td>153</td>
<td>11.37</td>
<td>15.15</td>
<td>12.27</td>
<td>19.67</td>
<td>1.60</td>
</tr>
<tr>
<td>Cytochrome c</td>
<td>4</td>
<td>104</td>
<td>7.58</td>
<td>19.81</td>
<td>8.16</td>
<td>26.26</td>
<td>3.22</td>
</tr>
<tr>
<td>Ribonuclease</td>
<td>4</td>
<td>123</td>
<td>17.42</td>
<td>25.14</td>
<td>20.39</td>
<td>43.81</td>
<td>2.15</td>
</tr>
<tr>
<td>$\alpha$-Crystallin</td>
<td>6</td>
<td>175</td>
<td>4.10</td>
<td>10.42</td>
<td>4.19</td>
<td>11.36</td>
<td>2.71</td>
</tr>
</tbody>
</table>

**NOTE.** $N =$ no. of species; $m =$ no. of amino acids; $M$ and $S^2 =$ estimators given by eqs. (1) and (2) respectively; $\hat{\mu} =$ estimated mean number of substitutions per lineage; $\hat{\sigma}^2 =$ estimated variance in the number of substitutions per lineage; and $\hat{r} =$ estimated value of $R(t)$.

* Data for all of the proteins except $\alpha$-crystallin were taken from tables 4.3-4.4 in Kimura (1983). The $\alpha$-crystalline data is from De Jong (1982) for human, rat, rabbit, dog, horse, and cow.

for Kimura's method rests on the assumption that the underlying process is a Poisson process. The resolution of this paradox must be that, for the data analyzed, the occurrence of multiple substitutions at a site is sufficiently rare that the different techniques for compensating for multiple substitutions are adjusting an already small bias.

In the analysis presented in table 2 it was assumed that $m$ equals the total number of amino acids. However, for most proteins the number of amino acids that are free to evolve is somewhat less than the total number. Within the context of our data, there is no way to judge whether any of the sites should be considered invariable or not. However, we can assume that $m$ is less than the total number of amino acids and examine the consequences. For example, if we assume that $m$ for cytochrome c is 52 rather than 104, then the estimate of $R(t)$ is 4.18 rather than 3.22. Similarly, for the $\alpha$-hemoglobin if we use $m = 72$ rather than $m = 141$, the estimate of $R(t)$ increases from 1.17 to 1.50. From these examples we see that the estimate of $R(t)$ will increase with decreasing $m$. Thus the values of $\hat{r}$ in table 2 should be viewed as underestimates.

The General Clock

The estimates of $R(t)$ and the mean rate of substitution should provide some insight into the process of molecular evolution. Such insight will only come after the adoption of a specific stochastic model of molecular evolution. In this section, the model to be employed is a statistical model, being totally independent of any genetically based models. The intent is to use such a model to gain insight into the dynamics of the process without reference to the underlying mechanism. Discussion of the mechanism will be deferred until the next section.

As was done in Gillespie (1984a), assume that the substitution process along any lineage is a doubly stochastic Poisson process, $N_\alpha(t)$, with the rate of evolution being

$$\lambda \theta(t),$$

where $\theta(t) > 0$, $E[\theta(t)] = 1$, $\text{Cov}[\theta(t), \theta(t + x)] = r(x)$; that is, the probability of a substitution occurring in a small period of time, $dt$, is

$$P_r[dN_\alpha(t) = 1] = E[dN_\alpha(t)]dt = \lambda \theta(t)dt.$$

(17)
A good treatment of Cox processes is given by Cox and Isham (1980). The process has been indexed by \( \tau \), which will play a role in the limiting operation that will be described shortly. A well-known result that follows immediately from these assumptions is that the total number of substitutions in a lineage, \( N_\tau(t) \), given a single trajectory of \( \theta_\tau(t) \), is Poisson distributed with mean

\[
\lambda \xi_\tau(t) = \lambda \int_0^t \theta_\tau(x)dx.
\]  

(18)

In Gillespie (1984a) it was argued that the time scale of change of the “molecular clock,” \( \theta_\tau(t) \), is probably much shorter than the length of a typical lineage under study. This assumption is motivated by the fact that major environmental changes, such as the recent ice ages, occur on a time scale of thousands to tens of thousands of years while the time between substitutions is typically on the order of millions of years. If the rate of evolution is changing on a time scale of thousands of years, the autocorrelation of the rate will approach zero very quickly on the time scale of the substitution process. Thus it is natural to consider the behavior of the integral \( \xi_\tau(t) \) as the autocorrelation as the process approaches zero. However, as the autocorrelation approaches zero, so does the variance in \( \xi_\tau(t) \) as given by the equation

\[
\text{Var}[\xi_\tau(t)] = 2 \int_0^t (t - x)r_\tau(x)dx.
\]  

(19)

In order to keep the variance from going to zero (required by the fact that \( R(t) \) is in the range 1.5–3.5), we must let the variance of \( \theta_\tau(t) \) approach infinity as the autocovariance for nonzero lags approaches zero in such a way that the variance of \( \xi_\tau(t) \) approaches a nonzero limit. As these limits are taken, we want to discover the limiting behavior of the process \( \xi_\tau(t) \).

Suppose we have a sequence of processes, \( \theta_\tau(t) \), such that, as \( \tau \to 0 \),

\[
r_\tau(x) \to 0 \quad \text{for} \quad x \neq 0,
\]  

\[
r_\tau(0) \to \infty,
\]  

(20)

such that \( \text{Var}[\xi_\tau(t)] \to vt \) and such that the limiting process exists and is a process with stationary, independent increments. We will call the limiting process \( \xi(t) \). \( \xi(t) \) will obviously be a process with orthogonal increments. This is true because the integrals of \( \theta_\tau(t) \) over disjoint intervals will be asymptotically uncorrelated since the autocovariance function approaches zero (as \( \tau \to 0 \)) for nonzero lags. However, orthogonality does not imply independence so, to be cautious, we will restrict the domain of our discussion to those sequences of processes with limiting processes having independent increments.

A nondecreasing process with stationary, independent increments, such as \( \xi(t) \), is called a subordinator. Its distribution is determined by the Levy equation

\[
E[e^{-\xi(t)}] = e^{-t\psi(z)},
\]  

(21)

where \( \psi(z) = \int_0^\infty [1 - e^{-zx}]\mu(dx) \) (see Kingman [1975] for a discussion of subordinators.
and Breiman [1968, section 14.4] for the background on processes with independent increments). The measure, \( \mu(x) \), called a Levy measure, characterizes the process. This equation is particularly useful for our purposes. The number of substitutions, \( N(t) \), is a Poisson process randomized by \( \zeta(t) \); thus its probability-generating function is

\[
E[e^{-\zeta(t)\lambda(1-\eta)}] = e^{-\eta t\lambda (1-\eta)}.
\] (22)

(See Feller [1968, chap. 11] for a discussion of probability-generating functions [pgf] and his formula 2.9 for the pgf of the Poisson distribution.)

In order to achieve the desired generalization of the results in Gillespie (1984a), we must see whether or not a random variable with this pgf may be represented as a Poisson sum of positive random variables, say

\[
N(t) = Y_1 + Y_2 + \ldots + Y_{M(t)},
\]

where \( M(t) \) is Poisson distributed and the \( Y_i \) are independent, identically distributed positive random variables. The answer is clearly in the affirmative. If the pgf of \( Y_i \) is

\[
g(s) = \psi(\lambda)^{-1} \int_0^\infty e^{-\lambda x}[e^{\lambda xs} - 1]d\mu(dx),
\]

and if

\[
E[M(t)] = \alpha \psi(\lambda),
\]

then the pgf of the sum will be given by \( e^{-\eta t\lambda (1-\eta)} \), as required.

Somewhat more about the distribution of the \( Y_i \) can be learned from the pgf, \( g(s) \). By expanding \( g(s) \) in powers of \( s \), we have

\[
Pr(Y_i = j) = \lambda[j(\psi(\lambda))]^{-1} \int_0^\infty e^{-\lambda x}x^j d\mu(dx).
\]

The moments of \( Y_i \) may be obtained by differentiating \( g(s) \):

\[
E(Y_i) = \lambda \psi(\lambda)^{-1},
\]

\[
\text{Var}(Y_i) = \lambda \psi(\lambda)^{-1}[\lambda \mu_2 + 1 - \lambda \psi(k)^{-1}],
\]

\[
\mu_2 = \int_0^\infty x^2 d\mu(dx).
\]

From these moments can be derived the moments of \( N(t) \),

\[
E[N(t)] = \lambda t,
\]

\[
\text{Var}[N(t)] = \lambda t(\lambda \mu_2 + 1).
\] (28)

We will now give two examples of processes that fulfil the assumptions of this development. The first is taken from Gillespie (1984a). For \( \theta_d(t) \) we use a two-state Markov process. The process remains in state zero for an exponentially distributed
time with mean $1/\mu_0$ and then jumps to the value $1/\left(\mu_0 \tau\right)$, where it remains for an exponentially distributed time with mean $\tau$ before returning to zero, and so forth. This process will be called the two-state clock. As shown in Gillespie (1984a), as $\tau \to 0$, the limiting process may be represented as a Poisson sum of exponentially distributed random variables:

$$\xi(t) = Z_1 + Z_2 + \ldots + Z_{\xi(t)}.$$  (29)

This process has the Levy measure

$$\mu(dx) = \mu_0^2 e^{-\mu_0 x} dx,$$  (30)

with $\mu_2 = 2/\mu_0$. Using these results we see that $Y_i$ is geometrically distributed with mean $(\mu_0 + \lambda)/\mu_0$ and that $M(t)$ is Poisson distributed with mean $\mu_0 \lambda t/\left(\mu_0 + \lambda\right)$. The distribution of a Poisson sum of geometric random variables is called a Polya-Aeppli distribution (Johnson and Kotz 1969). These results agree with those obtained in Gillespie (1984a) by a very different approach.

Another example that is very different from the previous one is a process that starts at zero, where it remains for an exponentially distributed length of time with mean $\tau$ and then jumps to a gamma-distributed height with mean $l/2$ and variance $t/\tau^3$, where it remains for time $\tau^2$ before returning to zero, and so forth. This process will be called the gamma clock. Again we let $\tau \to 0$ and discover that the limiting process, $\xi(t)$, is gamma distributed with Levy measure

$$\mu(dx) = (\nu x)^{-1} e^{-x/\nu} dx,$$  (31)

with $\mu_2 = \nu$.

This process is examined in detail in Kingman (1975). Unlike the previous case, $\xi(t)$ cannot be represented as a Poisson sum of random variables. Rather, $\xi(t)$ may be thought of as increasing in a countably infinite number of steps, “most” of which are “zero” in magnitude. There is some similarity with the previous case in that the number of jumps that are larger than any number larger than zero is Poisson distributed. For this case $Y_i$ has a logarithmic series distribution with mean $\lambda \nu/\log(1 + \lambda \nu)$ and $M(t)$ is Poisson distributed with mean $t \log(1 + \lambda \nu)/\nu$. Since $N(t)$ is a Poisson sum of a logarithmic series of distributed random variables, it has a negative binomial distribution.

The contrast between these two clocks stems from the fact that one has a Levy measure with finite total mass while the other has infinite total mass. In the former case the clock “ticks” a Poisson number of times in a finite interval, so it is natural to think of each of these ticks as an episode with a rapid rate of evolution even if no substitutions actually occur. For the latter clock, however, one cannot represent the clock as a Poisson sum of random variables (even though $N[t]$ can be represented in this way), so one cannot as naturally conceive of some quantity that represents the number of episodes of high evolutionary rate. Clearly, the common ground between these two clocks is in the representation of the number of substitutions rather than in the number of episodes. We will henceforth refer to an episode of evolution as occurring whenever at least one substitution occurs. This differs from the usage in Gillespie (1984a), where an episode referred only to the two-state clock and then to a case in
which the clock was not zero even if a substitution did not occur. The new usage is dictated by the more general treatment of the clock.

Using the expressions from the previous sections for the expectation of the estimators $\hat{\mu}$ and $\hat{\sigma}^2$, we see that

$$E(\hat{\mu}) \approx E[N(t)] = \lambda t,$$

$$E(\hat{\sigma}^2) \approx \text{Var}[N(t)] - \lambda t(\mu_2 + 1).$$

(32)

A minor calculation using these results shows that, for the two-state clock, $(\hat{t} - 1)/2$ is a consistent estimator of $\lambda/\mu_0$, the mean number of substitutions per episode, and $2\hat{\mu}/(\hat{\sigma}^2 - \hat{\mu})$ is a consistent estimator of the mean number of episodes per lineage. For the gamma clock, $(\hat{\sigma}^2/\hat{\mu} - 1)/\log(\hat{\sigma}^2/\hat{\mu})$ is a consistent estimator for the mean number of substitutions per episode and $\hat{\mu}^2\log(\hat{\sigma}^2/\hat{\mu})/(\hat{\sigma}^2 - \hat{\mu})$ is a consistent estimator of the mean number of episodes per lineage. These estimators are applied to actual data in table 3. One remarkable aspect of this analysis is the similarity of the inferred episodic events for the two clocks. Given the very different characters of the two clocks, this suggests that the inferred episodic structure is robust to the assumptions made about the clock, at least for the restricted range of values of $R(t)$ for the currently available data.

Models of Molecular Evolution

The analysis of the previous section suggests that the substitution process for a single locus may be represented by the random sum

$$N(t) = V_1 + V_2 + \ldots + V_{M(t)},$$

(33)

where the $V_i$ are independent, identically distributed, positive, integer-valued random variables and $M(t)$ is a Poisson process. In this section an attempt will be made to attach some biological significance to this representation by exploring models of molecular evolution based on the action of natural selection.

For various technical reasons theoretical work on models of molecular evolution based on natural selection has lagged far behind the work on neutral models. However,

<table>
<thead>
<tr>
<th>Protein</th>
<th>Mean No. of Substitutions per Episode</th>
<th>Mean No. of Episodes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Two-State</td>
<td>Gamma</td>
</tr>
<tr>
<td>$\beta$-Hemoglobin</td>
<td>2.02</td>
<td>1.83</td>
</tr>
<tr>
<td>$\alpha$-Hemoglobin</td>
<td>1.09</td>
<td>1.08</td>
</tr>
<tr>
<td>Myoglobin</td>
<td>1.30</td>
<td>1.28</td>
</tr>
<tr>
<td>Cytochrome c</td>
<td>2.11</td>
<td>1.90</td>
</tr>
<tr>
<td>Ribonuclease</td>
<td>1.57</td>
<td>1.50</td>
</tr>
<tr>
<td>$\alpha$-Crystalline</td>
<td>1.86</td>
<td>1.71</td>
</tr>
</tbody>
</table>
in a recent series of papers, Gillespie (1983a, 1983b) has developed a boundary layer-
approximation technique for multiple allele-diffusion models that allows one to de-
scribe molecular evolution in a particularly simple way. The approach is similar to
that used by Maynard Smith (1976) in another context. In summary, it was shown
that if the mutation rate to an advantageous allele is very small, say $10^{-9} - 10^{-8}$, and
if the selection coefficient of the allele is large relative to the reciprocal of the population
size, say 100--1,000 times larger, then the advantageous allele will spend an exponen-
tially distributed length of time in the boundary layer, i.e., at a very low frequency,
before sweeping through the population. The time required to sweep through the
population will be short relative to the time spent in the boundary layer. The only
alleles that are likely to exhibit this behavior are alleles that are one mutational step
away from alleles that are currently fixed or segregating at high frequency in the pop-
ulation. If there is more than one advantageous allele, then a particular allele's prob-
ability of being the first one to become fixed is its selective advantage divided by the
sum of the selective advantages of all of the advantageous alleles.

Using this boundary-layer description of the substitution process we can easily
describe a population’s response to a changing environment. Let $P(t)$ be a stationary-
point process that represents the times of change of those aspects of the environment
that cause a change in the relative fitnesses of one or more of the alleles at the locus
under study. With each environmental change the locus could respond with zero, one,
or more substitutions with a boundary-layer dynamic. This suggests representing the
substitution process as the random sum

$$S(t) = X_1 + X_2 + \ldots + X_{P(t)}.$$  \hfill (34)

In this representation, the $X_i$ represent the number of substitutions at the locus that
occur following the $i$th environmental change. In general, we would expect $X_i$ to be
greater than zero only if there are alleles that are both more fit than and one mutation
step away from those alleles that are currently in the population. Given that $X_i \neq 0$,
it's properties are determined to a great extent by the “mutational landscape” as
described in Gillespie (1984b). There it is argued, using some results from extreme
value theory, that the mean value of the $X_i$, given that $X_i > 0$, are likely to be in the
range 1.7--2.7.

The similarity of this representation of molecular evolution to the representation
based on the statistical analysis suggests equating $M(t)$ with $P(t)$ and $Y_i$ with $X_i$. There
are, however, two problems with this equivalence. The first is that $X_i$ can equal zero
while $Y_i$ cannot. The second is that $P(t)$ will not, in general, be a Poisson process. The
representations can be made more similar by modifying $P(t)$ to mark only those events
for which $X_i > 0$ and to substitute for $X_i$ the random variable, call it $Z_i$, whose dis-
btribution is equal to that of $X_i$ conditioned on $X_i > 0$. The substitution process could
now be written as

$$S(t) = Z_1 + Z_2 + \ldots + Z_{Q(t)},$$  \hfill (35)

where $Q(t)$ is a point process for which each event is an event of $P(t)$ that led to a
substitution. This simple expedient obviously handles the fact that the $X_i$ might equal
zero and, less obviously, makes $P(t)$ more Poisson-like. The transformation of the
process $P(t)$ into the process $Q(t)$ is called thinning in the point-process literature. It
is a well-known result that thinning of a point process leads, asymptotically, to a Poisson process (see Cox and Isham 1980, pp. 98–100). However, even though the process \( P(t) \) is made more Poisson-like by considering only those environmental changes that lead to substitutions, there is an extension to this one-locus model that will make the convergence of \( P(t) \) to a Poisson process even more compelling. It is an argument based on a simple form of epistasis.

Suppose that each environmental change presents the population with a challenge that can be met in equivalent ways by substitutions at many different loci. Suppose, for example, that there are \( L \) loci that are capable of responding to an environmental challenge by the fixation of one or more alleles. We suppose, for simplicity, that these \( L \) loci are equivalent in a long-term evolutionary sense. By this we mean that with each environmental challenge, each of the \( L \) loci is equally likely to experience an allelic substitution that will mollify the challenge. However, the reason that a particular locus is the one to experience the substitution depends on a number of factors. First among these is that, as discussed in the one-locus model, the locus must have at least one allele that is more fit than the alleles currently in the population and is a neighbor on the “mutational landscape,” that is, an allele that is more fit and is one mutational step away. As was argued in Gillespie (1984b), alleles that are two or more mutational steps away from those currently in the population are unlikely to enter the population because of the extraordinarily low rate of production of these double mutants. Thus, with each environmental change, some of the \( L \) loci will be sitting on peaks in the mutational landscape, with further evolution in the current environment unlikely, while other loci will be on the sides of peaks, with the potential for allelic substitutions.

When the environment changes, only a subset of the \( L \) loci will be on the sides of peaks on the mutational landscape. Among these, there will be a tremendous range in the strength of selection acting on the advantageous alleles. Those loci whose advantageous alleles have the largest selective advantage are the ones that are most likely to experience the substitutions that satisfy the environmental challenge. This is the second factor that will determine the probability that a particular locus among the \( L \)-equivalent loci will be the one that is chosen.

There are undoubtedly other factors that will contribute to the probability of a particular locus being chosen, but these two appear to be the most important. Thus, with each environmental change there will be a small subset of the \( L \) loci that are likely to respond: those off peaks with the strongest selection. In some cases, there may be no loci that can respond. In those cases when more than one locus are capable of responding with similar selection coefficients, the substitutions should occur with equal probability among them. With each environmental change, we picture another randomly chosen set of loci being the ones that are likely to respond with substitutions. These ideas will now be investigated mathematically through an approximation that assumes that the number of loci and the rate of environmental change are large.

If \( L \) is large, the probability of at least one of the substitutions occurring at any particular locus being small, say equal to \( \alpha/L \), where \( \alpha \) will depend on the particulars of the model. Given that at least one substitution occurs, let the pgf of the total number that occurs at a given locus be \( g_L(s) \). The pgf for the number of substitutions at a locus following an environmental change is then

\[
1 + (\alpha/L)(g_L(s) - 1).
\]
If the total number of environmental changes is represented by the point process \( P(t) \), then the number of substitutions at a particular locus has the pgf

\[
E\{ [1 + (\alpha/L)[g_L(s) - 1] ]^{P(t)} \}.
\]

To arrive at an approximation of this formula equation, let the number of environmental changes and the number of loci increase such that \( P(t)/L \to \lambda t \) and let \( g_L(s) \to g(s) \). The pgf of the total number of substitutions will then become, asymptotically,

\[
E\{ e^{P(t)\log(1 + \alpha/L)[g_L(s) - 1]} \} \\
\sim E\{ e^{P(s) - 1}P(t)/L \} \\
\sim E\{ e^{\lambda tP(s)/L - 1} \}.
\]

This will be immediately recognized as the pgf for a Poisson sum of random variables of precisely the same form as that obtained by the statistical model

\[
S(t) = Z_1 + Z_2 + \ldots + Z_Q(t),
\]

where \( Z_i \) has the pgf, \( g(s) \), and where \( Q(t) \) is a Poisson process with mean \( \lambda t \).

As in the single-locus case, the limiting argument used to achieve this representation involves thinning. The effect of thinning on the variance-to-mean ratio is given by

\[
R(t) = \frac{\text{Var}[S(t)]/E[S(t)]}{\text{Var}(Z_i)/E(Z_i) + (\alpha/L)E(Z_i)\mu(t)},
\]

where \( \mu(t) = \text{Var}[P(t)]/E[P(t)] \) is the index of dispersion of the process of environmental change. This expression shows how the variance-to-mean ratio is lowered as the number of loci is increased.

The expression

\[
\kappa = E(Z_i) + \text{Var}(Z_i)/E(Z_i)
\]

requires some additional discussion. In was shown in Gillespie (1984b) that excursions through the mutational landscape lead to values of \( \kappa \) in the range 2.0–3.5. In the context of the epistatic model, this range would be expected to extend down to one. The reason is that with each environmental change there could be more than one locus that is a candidate for responding with a substitution. If the total number of candidate loci is larger than the total number of substitutions that occurs, then each locus is unlikely to experience more than one substitution. Since \( Z_i \) is conditioned to be greater than zero, \( Z_i \) would equal one with a probability of nearly one, yielding a \( \kappa \) of nearly one. In the original description of the model this would correspond to \( \alpha = 1 \) and \( g(s) = s \). Thus the prediction of our model is that the \( R(t) \) should be in the range \( 1.0 < R(t) < 3.5 \), exactly as observed in the data.

This argument is only valid if \( L \), the number of loci that can respond to a particular environmental change, is large. It may well be the case that for certain environmental
changes \( L \) is very small, perhaps even equal to one. For these cases the value of \( R(t) \) could be rather large if the process of environmental change, \( P(t) \), has a large index of dispersion.

**General Discussion**

This paper presents a statistical analysis of protein sequence data and a model of natural selection that share a common episodic structure. Moreover, the model specifically predicts that the values of \( R(t) \) observed in the data should be in the range \( 1.0 < R(t) < \approx 3.5 \) if enough loci are capable of responding to particular environmental changes. Thus the statistical aspects of the data appear to fit a model of evolution by natural selection even better than they fit a constant-rate neutral model. This conclusion is made more forceful by the recent demonstration (Gillespie 1985) that, under neutrality, \( R(t) \) should be

\[
R(t) = 1 + V(4Nr)4Nu/(1 + t/2N),
\]

where \( r \) is the intragenic recombination rate, \( N \) is the population size, \( u \) is the neutral mutation rate, and \( V(\cdot) \) is a function described in Hudson (1983b) that varies from one to zero as its argument varies from zero to infinity. It had already been pointed out that for neutrality to be compatible with the observed values of \( R(t) \) in a model without intragenic recombination, \( 4Nu \) would have to be too large, say \( 1 < 4Nu < 10 \), to be compatible with the polymorphism data (Gillespie and Langley 1979; Hudson 1983a). When recombination is introduced, the effect is to lower the value of \( R(t) \) and thus make it even more difficult to account for the data under a neutral model. Thus, from the perspective of purely statistical considerations, our model of evolution by natural selection actually seems to fit the data better than does the neutral model.

The neutral allele model could be modified to incorporate variable mutation rates and thus be made more compatible with the data. This has, in fact, been attempted both explicitly (e.g., Wu and Li 1985) and implicitly (e.g., Kimura 1983). In each case the model seems to be the same: the mutation rates for each of the lineages of the star phylogeny are chosen independently and remain fixed from the point of origin of the lineage until the present. For this model, the variance-to-mean ratio of the number of substitutions per lineage, \( R(t) \), is a good measure of the variability of the rates of substitution. However, this model seems quite artificial. Why should the rates of mutation be altered only at the time of origin of the orders of mammals and then remain unaltered for the next 60 Myr even though other lineages are branching off to form families, genera, and species? A more realistic model would have the neutral mutation rates varying through time in a single lineage to the same extent that they vary between lineages. Such a model has not yet been analyzed. We can speculate that it may well result in a prediction of episodic mutation rates whose range would be outside the ranges that are commonly observed in laboratories. One conclusion is clear: the neutral allele theory cannot, at present, account for the statistical patterns in the sequence data without the use of a very artificial model of rate variation. While the theory could be modified to be in better agreement with the data, such modifications will make the theory less compelling as a uniquely parsimonious explanation for molecular evolution.

Another generalization of the neutral model that could elevate the value of \( R(t) \) is Ohta's (1976) theory of mildly deleterious alleles. Under this model, the rate of evolution depends on the population size as well as on the mutation rate. Fluctuations
in the population size could potentially increase the value of \( R(t) \), although there are no published results that suggest that this theory can elevate the values of \( R(t) \) to the level observed in the data.

A basic conclusion of the present paper is that molecular evolution is compatible with an episodic model. To arrive at this conclusion we began with a model with varying rates of evolution. However, other starting points could lead to models without an episodic structure that are nonetheless compatible with the observed values of \( R(t) \). One such model would involve interactions of substitutions at different sites. It is well known (e.g., Blaisdell 1985) that there is some correlation in the identities of neighboring bases in DNA sequences. This suggests that the occurrence of a substitution at one site may increase the likelihood of a substitution at a nearby site. If the increase were adequate, this could lead to an elevation in the value of \( R(t) \) without the necessity of an episodic model. A common name given to this phenomenon is "compensatory substitutions."

A second class of models without an episodic structure would be models in which the rate of change of the molecular clock occurs on a time scale similar to that of the length of the lineages. While this assumption does not alter the conclusion that the variance in the rates of evolution is very much larger than previously thought (Gillespie 1984a), it does not require an episodic process. A complication of this model involves a fundamental issue as to whether the rates of evolution that remain nearly constant for long periods of time are better represented as random quantities chosen from a common probability distribution or as fixed quantities that differ from one part of a lineage to another. This is similar to the distinction between fixed-effects and random-effects models in the analysis of variance.

Clearly, the next step in the development of stochastic models of molecular evolution must be to incorporate all three elements in the model: episodes of evolution, compensatory substitutions, and slow changes in rates. Unfortunately, the current data will probably be inadequate to determine the relative importance of these three elements.

As a final caveat it should be noted that the analysis presented here depends on the fact that the data come from a phylogeny that is closely approximated by a star phylogeny. If this assumption is inappropriate, then the high values of \( R(t) \) could be due to the variation in the lengths of the lineages. The analysis of Langley and Fitch (1974) is not subject to this problem and yields an overall estimate of \( R(t) \) of \( \sim 0.5 \). This is similar to the average value of \( R(t) \) estimated from the proteins used in this study, suggesting that the assumption of a star phylogeny is fairly accurate. As more sequences become available, attention should turn to the distribution of \( M \) and \( S^2 \) for phylogenies other than star phylogenies.

Acknowledgments

This paper has benefitted from many useful comments from Brian Charlesworth, Walter Fitch, and Michael Turelli.

LITERATURE CITED


WALTER M. FITCH, reviewing editor

Received February 21, 1985; revision received December 13, 1985.