Rate of Change of Concomitantly Variable Codons

W. M. FITCH

Department of Physiological Chemistry, University of Wisconsin
Madison, Wisconsin 53706

Summary. It was previously shown that about 10 % of the codons in cytochrome *c* are variable in any one mammalian species and any one point in time and that the positions of these *con*comitantly *vari*able cod*ons* (covarions) must change as mutations are fixed. Variability implies the existence of an alternative, non-deleterious amino acid that differs by only one nucleotide replacement from the one presently encoded. This work, in addition to obtaining an independent estimate of the number of covarions, investigates the question: What is the likelihood that a cytochrome *c* covarion will lose its variable status as a result of the fixation of a mutation in another covarion ? The results show: 1, the number of covarions is in the range of 4 to 10 in agreement with the earlier result of 10 but suggesting the variability may be even more circumscribed than originally thought; and 2, the likelihood of a covarion loosing its variable status as a result of fixations elsewhere in the gene may be greater than 0.75, suggesting a high turnover rate among the covarions.

Key-Words: Evolutionary Rates — Molecular Genetics — Cytochrome *c* — Mutations — Codon Variability.

It is abundantly clear in the gene for cytochrome *c* that selective forces narrowly restrict the number of codons that are capable of surviving and fixing a mutation that alters the amino acid encoded. It has been shown that the number of covarions in cytochrome *c* in mammalian species (i.e. concomitantly variable codons that can fix mutations at any one point in time and in a given particular species) is only about ten (Fitch and Markowitz, 1970). Since more than ten codons have fixed mutations, it follows that a different (at least partially) set of codons must be concomitantly variable in different species. This too has been demonstrated by showing that the variable codons in the fungi are largely independent of the variable codons in insects and fish (Fitch, 1971a). But if the concomitantly variable codons (covarions) in one species may not be identical to those in another, it follows that the nature of the new amino acid resulting from a fixation must affect which codons may be capable of fixing the next mutation, a biologically reasonable conclusion. Moreover, if each mutation fixed may alter the codons that belong to the group of covarions, we can immediately understand why there are so few covarions in any one species' gene for cytochrome *c* and yet observe that more than 70 codons have fixed observ-
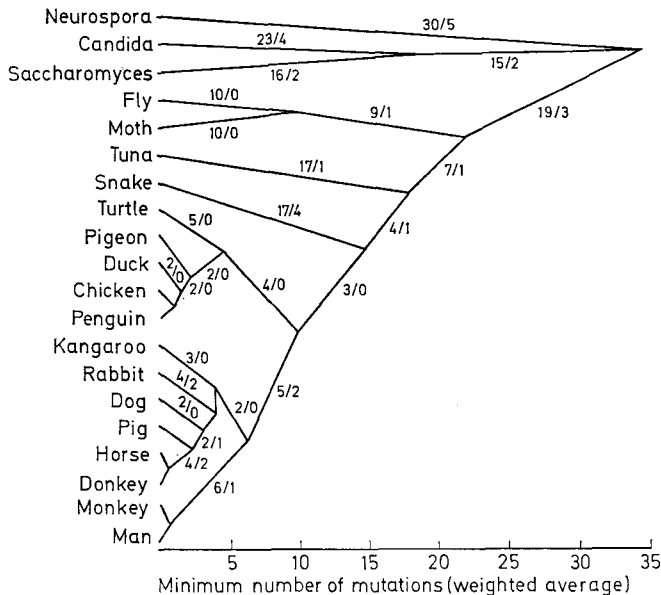
Fig. 1. Phylogeny from cytochrome c. The phylogeny is identical to that previously determined for the twenty species shown utilizing only the amino acid sequence data for its construction (Fitch and Margoliash, 1968). The height of any given node is the weighted average of mutations fixed in the descent from that node. For segments for which two or more mutations were fixed, the lines contain values in the form m/D where m is the total number of mutations fixed and D is the number of double mutations, i.e. the number of codons that contain two of the fixations

able mutations upon examining the cytochromes c in species ranging from fungi to man. This raises, however, an interesting question: What is the rate of turnover of codons in the covarions? Expressed another way, what is the likelihood that a fixation in one covarion will cause other covarions to loose their ability to fix the next mutation?

To answer this question, we first define a double mutation as two observable fixations occurring in a single codon between two successive branch points (nodes) on a phylogenetic tree (Fitch and Margoliash, 1968). For example, residue 19, which is next to the histidine that binds to the heme iron in all known cytochromes c, is threonine (encoded ACX) and occurs in all known cytochromes c except in *Neurospora* which has a glycine (encoded GGX). Given the phylogeny of these species, both the first and second nucleotide *must* have mutated since the most recent ancestor of *Neurospora* that is depicted on that tree. Indeed, it is one of the 5 double mutations on the branch labelled 30/5 in Fig. 1. The "must" depends upon assuming that no mutations will be postulated other than those that are forced upon us in order to account for any observed divergence. Excluding the possibility that the change AC→GG occurred as a single mutational event (which is

reasonable in view of the non-existence of such double changes in *in vitro* studies of point mutations), it then becomes clear that the ability of a codon to fix the second mutation must have depended upon its being among the covarions after fixation of the first mutation. Now the likelihood that a second mutation will be fixed in the same codon as an earlier mutation depends upon two characteristics: (a) the number of covarions, and (b) the persistence of a codon among the covarions. Obviously, the fewer the number of covarions, the greater the probability that the next fixation will occur in any given covarion and, in particular, in a previously mutated covarion. On the other hand, the probability that a covarion remains in the group drops as mutations are fixed in other covarions. The number of covarions, c, and the persistence of variability, v, are the two independent variables that are used to derive the equations given in the appendix. These equations enable one to estimate, depending on the values of c and v assumed, how many double mutations would be expected for any particular number of mutations

Table. *Frequencies of double fixations*

m	n	Observed D	Expected D		
			$c = 4.5$	$c = 7$	$c = 10.0$
2	6	0.167	0.147	0.091	0.063
3	2	0	0.403	0.418	0.439
4	4	1.25	0.533	0.510	0.511
5	2	1.0	0.667	0.611	0.594
6	1	1.0	0.804	0.720	0.690
7	1	1.0	0.946	0.838	0.798
9	1	1.0	1.243	1.100	1.050
10	2	0	1.397	1.244	1.193
15	1	2.0	2.231	2.085	2.076
16	1	2.0	2.409	2.278	2.284
17	2	2.5	2.592	2.478	2.503
19	1	3.0	2.969	2.901	2.972
23	1	4.0	3.769	3.837	4.028
30	1	5.0	5.314	5.749	6.232
		$v = 0.04$		0.12	0.23
		$\chi^2 = 3.16$		3.34	3.62

The values of m are the number of mutations observed to have been fixed in a given internodal interval and n is the number of internodal intervals containing m fixations. The column marked "observed D" contains the total number of codons observed to have fixed two mutations in an interval of size m (see Fig. 1) divided by the number of such intervals to give the observed frequency of double fixations per interval of size m. There were no triple fixations in a single interval. The observed frequency is to be compared to the expected number of double fixations per interval (D) as calculated by the procedure given in the methods section. Expected values of D are calculated for three values of the number of covarions c, and are given for that persistence of variability v that optimizes the fit to the observed data. For 12 degrees of freedom, $\chi^2 = 3.57$ implies that the observed values would, by chance, differ from the expected values to the extent shown $\geq 99\%$ of the time.

m occurring along some observed interval. These expected values may then be compared to values actually observed and the goodness of fit determined by the chi-square statistic.

Fig. 1 shows the 20 species whose cytochromes c were used for this analysis. The topology, the mutations and the double mutations are identical to those given in Fig. 1 of Fitch and Margoliash (1968). Legs containing only one or no mutations can not have double mutations and therefore no numbers are shown for those legs. The remaining legs have two numbers: the first is the total number of mutations observed to have been fixed in that portion of the phylogeny. The second number (following the slash) is the number of double mutations observed in that same portion. In the Table we see the distribution of double mutations found as a function of the number of mutations observed. It is our purpose to find that number of covarions, c, and persistence of variability, v, that would most closely approximate that distribution. A value of v near 1 would mean that a covarion's ability to fix a mutation is largely independent of fixations in the other covarions while a value of $v \ll (1/c)$ would mean that a codon's variability is very unlikely to survive more than one fixation in other covarions. The value of c has been estimated as 10 by an independent method (Fitch and Markowitz, 1970) and the closeness of the agreement between 10 and the number found here will be one indication of the adequacies of the procedures.

Results

The formal procedures are given in the appendix. In this section are presented results intended to provide a feeling for how the variables interact as well as results on cytochrome c in particular.

In asking how many double mutations would be expected after 30 mutations have been fixed, one recognizes that the answer depends upon the number of covarions into which they may be fixed and how long a covarion lasts. Fig. 2 shows the curve of expectation for five covarions as a function of the persistence of variability. If the persistence is 1, then we expect the same five covarions to be always present and by the time 30 mutations were fixed, we expect that all five covarions would have fixed at least two mutations and hence the left-most value is essentially five. As the persistence of variability drops, the curve rises to a maximum, then falls. This is simply a reflection of the fact that if the persistence value is too large, the codons remain variable too long after they have fixed a second mutation, whereas if the value is too small, they do not remain variable long enough to get a second mutation.

In practice we would choose a persistence of variability that would yield the number of double mutations actually observed in the line of descent containing 30 fixations. However, there are other line segments containing
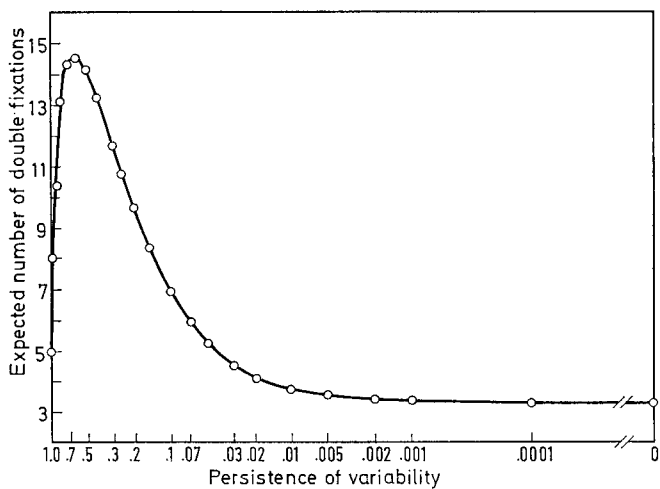
Fig. 2. Number of double mutations as a function of the persistence of variability. The curve gives the expected number of double mutations (D) for various values of the persistence of variability (v) when 30 mutations have been fixed and the number of covarions is five
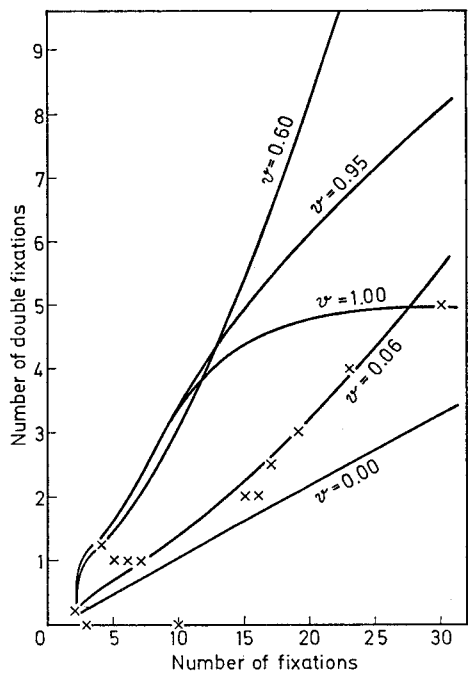

Fig. 3. Double mutations expected for five covarions as a function of the number of fixations. Each line shows the expected number of double mutations (as a function of total fixations) for a different value of v. The points are the actual data (see the Table) that must be fit

other numbers of fixations and we are constrained to choose the same persistence of variability throughout. Fig. 3 shows how the expected number of double mutations varies as a function of the total number of fixations. Each line is for a different persistence with the number of covarions being held at five. The points on the graph show the observational data to be fitted. Some curves fit better than others and chi-square is a measure of the goodness of fit with the lowest value of chi-square being the best fit. The goodness of fit as a function of the persistence of variability is presented, for five covarions, in Fig. 4. The best fit occurs at $v = 0.06$ which is one of the lines in Fig. 3.

But there is nothing to say that five covarions is the correct number of covarions and similar curves may be drawn for other numbers of covarions and Fig. 4 also shows such curves for $c = 2$ and 10. The best fit will be that number of covarions that gives the lowest minimum in such a plot. Portions of the curves in the region of their minimum for $c = 3 \rightarrow 10$ are shown in Fig. 5 and are quite regular in shape. A plot of the minimum as a function of the number of covarions is given in Fig. 6. This minimum occurs for a value of $c = 4.5$. For $c = 4.5$, the minimum comes at a persistence of variability $= 0.04$ which is plotted as the * in Fig. 5. That the optimum value of c is not a whole number is all right in view of our recognition that it is only an average value and must be reasonably considered to vary somewhat from species to species and from time to time.
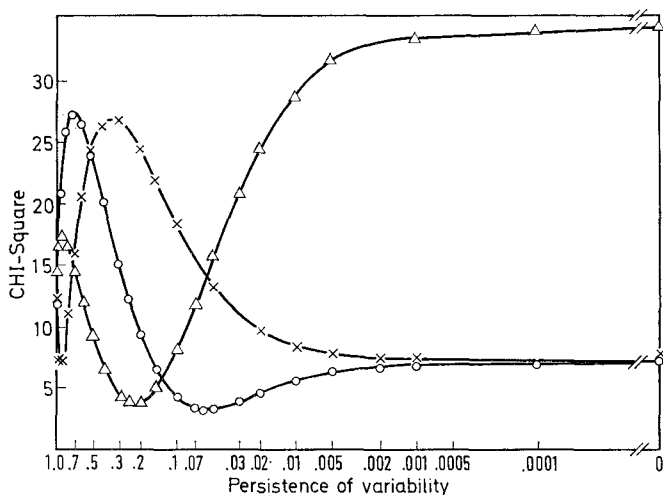

Fig. 4. Goodness of fit as a function of the persistence of variability for several numbers of covarions. The goodness of fit, measured by the chi-square test, is between the actual values of D observed and those expected under the various values of c and v assumed. For all cases, there are 12 degrees of freedom. Values calculated only for the points shown. Curves were fit by eye. Number of covarions are 2 ($\times$), 5 (o), and 10 ($\triangle$)
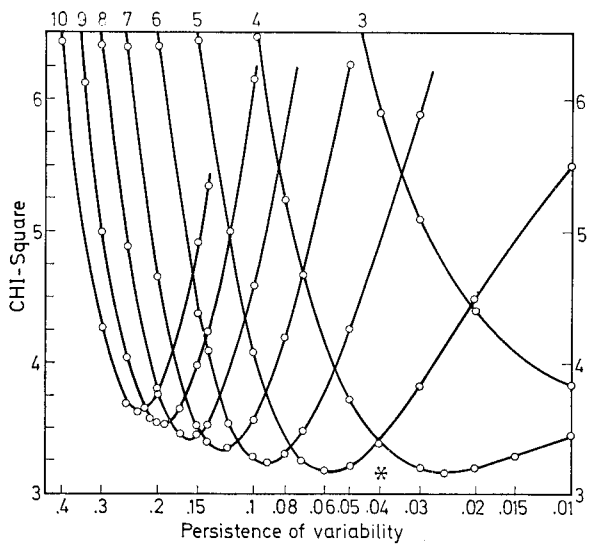
Fig. 5. Goodness of fit as a function of the persistence of variability for several numbers of covarions. Same as Fig. 4 except for restriction of the data to the portion of each curve in the region of its minimum. The number of covarions for each curve is shown at the top of the graph. For $c = 3$, the minimum occurs at $v = 0$ for which $\chi^2 = 3.47$
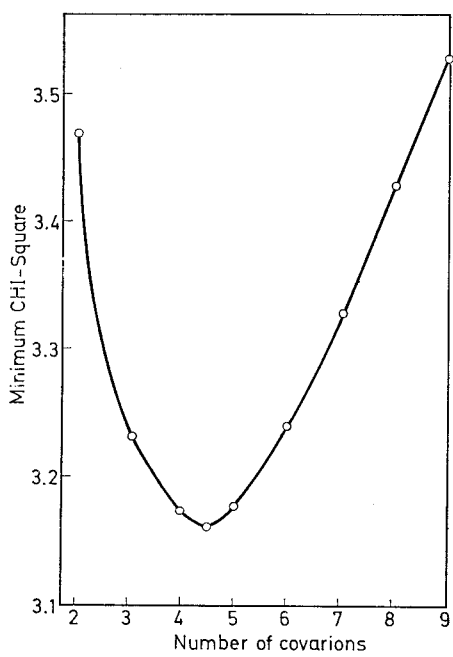

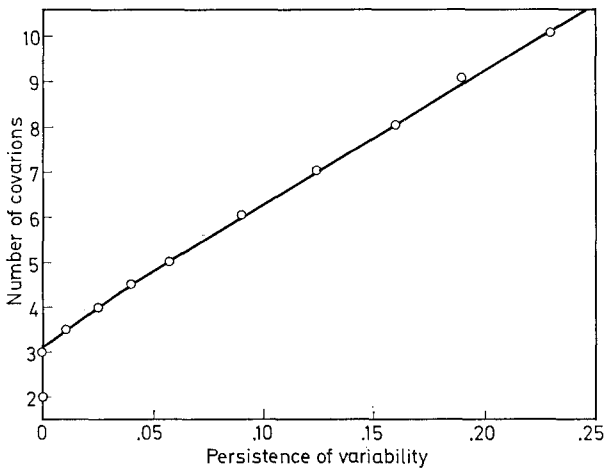Fig. 6. Chi-square for best fit for various numbers of covarions

Fig. 7. Persistence of variability for best fit for various numbers of covarions

Finally, one may ask about the relationship between the number of covarions and the persistence of variability that minimizes chi-square for that value of c. This is shown in Fig. 7. There are two straight lines intersecting at a value of $c = 4.5$.

Discussion

Previous work (Fitch and Markowitz, 1970) suggested that the number of covarions for mammalian cytochromes c was about 10. This study suggests a value of 4.5. Since the methods are different, they verify the conclusion that selective forces greatly restrict the permissible variability in cytochrome c. The actual discrepancy between the two values turns out not to be significant. Note that the minimum chi-square for $c = 10$ is only 3.62. A chi-square value of 3.57 means that if our assumptions are true (including $c = 10$) then we would expect real data to give by chance a worse fit than these do 99 % of the time. We are suffering from the fact that the probability surface is more of a valley than a well. The steep rise of the curves in figure five shows, in effect, the cross sections of the valley, the minima show the gentle rise as one travels up the valley. Given the number of covarions, we can reasonably estimate the persistence of variability and vice versa, but these data do not readily permit the simultaneous estimate of both with great accuracy.

The generally low values for the persistence of variability (optimum values of v are < 0.25 for all values of c checked) are particularly interesting. Most importantly, it means that the variable positions are largely interdependent and that a mutation at any one of them effects most of the others to a considerable degree. This reemphasizes the very limited tolerance that

cytochrome c has for variation in its structure. On the other hand, it should also be recognized that variability may be lost as a result of fixations in genes whose products interact with cytochrome c. We know nothing of these rates, but the longer the interval between fixations in the cytochrome c gene (and it is longer than most), the greater the likelihood that the loss of variability is the result of fixations elsewhere. This does not effect the computations but it must temper their interpretation.

The problem of explaining a double mutation in the terms of a strict selectionist theory in which every mutation fixed confers an advantage is interesting in terms of these results. Consider the replacement of proline (CCX) by valine (GUX) which occurs at position 44 of the rabbit. The intermediate possibilities are alanine (GCX) or leucine (CUX). The strict selectionist asserts that either alanine or leucine must have been superior to proline and replaced it and that subsequently valine was superior to the replacing amino acid and in turn replaced it. The valine superiority assures that the once mutated codon is among the covarions. On the other hand, this analysis suggests that covarions do not persist long so that in effect the two mutations must have generally followed each other successively or nearly so. But if that is true, then one of the intervening amino acids must have provided an intermediate degree of evolutionary fitness. This leaves us with an apparent contradiction. On the one hand, the genetic code seems to be fashioned so that single nucleotide replacements minimize deleterious changes and maximize the possibilities of advantageous changes. On the other hand, the optimum amino acid is very frequently two nucleotide replacements away and one of the two intervening amino acids has an intermediate fitness. In the present data there are 223 mutations of which 64 (32 doubles) are involved in this particular form of change. Thus 29 % of the mutations fixed were to get to a more fit amino acid two nucleotide replacements away from that originally encoded. Does this really square with the idea that cytochrome c is highly evolved and tolerates little change? And what we see can only be those potential improvements for which an intervening amino acid has intermediate fitness. Surely there must be others for which the intervening amino acids are deleterious in which case "you just can't get there from here". But if these latter cases that we can't observe number anywhere near as many as those that we have been able to observe in the former, then the fraction of amino acid substitutions that confer an advantage but require two nucleotide replacements becomes unreasonable.

From this study I would conclude: I, the number of covarions in the gene for cytochrome c averages between 4 and 10; II, the turnover averages 75 % or more among the covarions not fixing the last mutation; and III the frequency with which codons incorporate replacements in two of their nucleotide positions in relatively close succession argues more for the flex-

ibility permitted at that site than for successive selective improvements. This last conclusion is consistent with the observation that the genes for alpha hemoglobin (Fitch, 1971 b) and cytochrome *c* and fibrinopeptide A (Fitch and Markowitz, 1970) are evolving at the same rate per covarion.

Appendix

It was previously shown that only about ten of the codons in the gene for mammalian cytochromes *c* could accept (fix) a mutation at any one point in time (Fitch and Markowitz, 1970). These were called the *concomitantly variable codons* (covarions). But this number is considerably less than the 70 or more codons that are known to have had fixations in them. These two facts are not inconsistent under the hypothesis that, as mutations were fixed, the codons that could fix subsequent mutations changed so that while few codons in any one species could change at any one time, many could change during the evolution of many species. It is the purpose of this appendix to examine the rate at which previously concomitantly variable codons are replaced by new ones and to estimate the number of such covarions by a procedure quite different from that given originally (Fitch and Markowitz, 1970). This is done by examining the rate at which second mutations are fixed in codons that have already fixed a mutation as a function of the total number of mutations fixed and showing that multiple fixations in a single codon have the property of a Poisson function showing contagion.

Method

The basic question is, after m observable mutations have been fixed in a given gene, how many codons in that gene will have fixed two or more mutations? To calculate the number of different codons fixing two or more mutations in the course of a total of m mutations being fixed we need to know: (A), how many covarions c there are in which the mutations can be fixed; (B), the probability v that, following a fixation in another codon, a variable codon will remain variable, that is, that it will retain its ability to fix a mutation; (C) the expected number e_i of covarions that are still variable and have fixed at least one mutation after i fixations; (D), the probability f_i that the mutation following the ith is fixed in any given one of the e_i previously mutated covarions [this, in conjunction with (B), permits us to determine the probability r_{i+1} that the mutation following the ith is fixed in a previously *un*mutated codon]; (E), the probability p_k that, following a codon's first fixation, it has not fixed a second in k subsequent fixations; and finally (F), the probability s_{ih} that at least one of h subsequent fixations occurs in the same codon as fixed the ith mutation.

Our problem is to calculate the expected number of doubly mutated codons D that will have fixed two (or possibly three) observable mutations after a total of m fixations in the gene. If we know the probability r_i that the ith fixation is in a previously unmutated codon and the probability s_{ij} that at least one of the ensuing $m - i = j$ observable fixations will be in that same codon, the product $r_i s_{ij}$ is the expectation that the ith fixation is the first of two (or three) observable fixations in that codon. The expected number of observable double fixations is simply the sum of such products for all values of i from 1 to $m - 1$. This may then be compared with observed frequencies of multiple fixations to determine the adequacy of this model. The problem now is to determine r and s.

Throughout the ensuing discussion, the only mutations being considered are those that are fixed and are observable. We assume that the particular codons classified as concomitantly variable are highly dependent upon the nature of the particular amino acids in various parts of the protein so that we would not necessarily expect the same

codons to be concomitantly variable in the gene for a fungal cytochrome c as in the gene for the dog cytochrome c. Neither do we expect that the number of covarions c is constant from species to species nor from time to time, but we do assume that treating as a constant the average value taken by c in many species over long periods of time will give a useful approximation.

As each mutation is fixed in the population, there must be a probability ($0 \leq v \leq 1$) that any given codon will remain among the concomitantly variable group i.e., v is the persistance of variability. All other variables in this discussion are dependent solely upon c and v. These are our given parameters. We shall assume that if a codon is removed from the group of covarions, it is replaced by another codon that has not recently[1] been among the concomitantly variable codons. We assume that v applies to each codon *except* the codon fixing the most recent mutation. The codon fixing the last mutation we assume remains variable. Such an assumption must be correct when the last fixation has no significant selective advantage. It need not, but could also be true if the last fixation was selected for.

The probability that any one of the covarions will fix the first observable mutation is $1/c$. However, we are only interested in observable mutations and our ability to observe subsequent fixations depends upon prior fixations. If a codon has not fixed a mutation previously, all mutations which change the amino acid encoded are observable. If a mutation has already been fixed in that codon, then the next one will be observable only if a nucleotide is altered in a position different from that changed in the first fixation. This is because successive nucleotide fixations such as $A \rightarrow C \rightarrow G$ between an ancestor and a descendant appear to the outside observer as $A \rightarrow G$ there being no record of the intermediate C. Thus the probability that the next *observable* fixation will occur in a once mutated codon is only 0.604 that of a codon which has not previously fixed a mutation[2]. If f_i is the probability that the next mutation is fixed in a covarion that has already fixed a mutation and f_i' is the probability that it is fixed in a previously unmutated covarion, then $f_i = 0.604\,f_i'$. Since e_i is the expected number of previously mutated covarions, $c - e_i$ is the expected number of covarions without prior mutations being fixed in them. And since the next fixation occurs somewhere with probability one, $f_i e_i + f_i'\,(c - e_i) = 1$. Solving,

$$f_i = 0.604/(c - 0.396\,e_i). \tag{1}$$

Thus f_i is the probability that after i fixations, the next fixation occurs in a particular previously mutated codon. We shall let $g_i = 1 - f_i$. After the first fixation there is precisely 1 covarion that has fixed a mutation and therefore $e_1 = 1$, from which f_1 may be calculated. To find the subsequent values of f_i requires the subsequent values of e_i for which we now derive a recursion relation.

---

1 Recently is here a relative term. The calculations are based upon mutations between an ancestral form (the node of some phylogenetic tree) and its next immediately known descendant (the next node or a presentday species). The model, as formally stated, does not permit a once mutated codon to be removed from the variable group and subsequently returned to that group in the same internodal period. Nevertheless, even if removal and return occur in the same internodal interval, the model still works if v is interpreted as a function that reflects the average amount of time a concomitantly variable codon is likely to be variable.

2 The value 0.604 was arrived at as follows. Excluding mutations involving termination codons, there are precisely 166, 176, and 50 ways, of altering the first, second and third nucleotides of codons, respectively, so as to change the coding from one amino acid to another. This is a total of 392 ways. The probability p_i that the ith nucleotide is involved in an observable mutation is therefore assumed to be 166/392, 176/392 and 50/392 or 0.423, 0.449 and 0.128 for $i = 1$, 2 and 3 respectively. The probability that a second mutation will be fixed in a nucleotide position other than the position of the first fixation is $\sum\limits_{i=1}^{3} p_i (p_j + p_k)$ where $i \neq j \neq k \neq i$.

To find e_{i+1} we note that the fixation following the ith occurs among the previously mutated covarions with probability $e_i f_i$. The remaining $e_i - e_i f_i$ previously mutated covarions are subject to loss of their covarion status by virtue of the effect of this last mutation so that only $e_i(1 - f_i) v$ of them remain variable. To this group must be added the covarion that fixed the last mutation so that

$$e_{i+1} = e_i(1 - f_i) v + 1 = e_i g_i v + 1. \tag{2}$$

Thus given any e_i, f_i can be calculated from equation 1 and given any e_i and f_i, e_{i+1} can be calculated from equation[3]. Since e_1 is 1, all e_i and f_i are obtainable.

Now, since the probability that the $i^{th} + 1$ fixation occurs in a previously mutated covarion is $e_i f_i$ then

$$r_{i+1} = 1 - e_i f_i \tag{3}$$

is the probability that the $i^{th} + 1$ fixation occurs in a previously unmutated covarion. The first fixation must occur in an unmutated covarion, hence $r_1 = 1$. All other r_i can be obtained from Eq. (3).

We now proceed to calculate the probability s_{ih} of a second fixation in a covarion in the h fixations following the ith fixation. Consider a specific codon which fixes its first mutation at the ith fixation. After this initial fixation, there follow k subsequent fixations in the gene. We shall let p_{ik} be the probability (a), that the ith fixation occurred in a previously unmutated codon AND (b), that the ith fixing codon was still variable *before* the kth subsequent fixation AND (c), that the kth subsequent fixation was not fixed in that same i^{th} fixing codon. For brevity we will temporarily drop the subscript i. We shall develop a recursion relation for p_k. Since the kth fixation occurred in a different codon, the probability that the ith fixing codon remains variable after the kth subsequent fixation but has not fixed a second mutation is vp_k, while $(1 - v) p_k$ is the probability that, as a result of the kth subsequent fixation, it will be removed from the variable group and therefore couldn't fix a future mutation. Furthermore, letting $j = i + k$, the probability that the next mutation will be fixed in that codon is then $f_j vp_k$ and that the next one won't be fixed there, even though it is possible, is $g_j vp_k$. To summarize regarding a codon fixing its second mutation with the next fixation:

p_k is the probability it could have but hasn't, $1 - p_k$ that it couldn't or has already;

vp_k is the probability it still can, $(1 - v) p_k$ that it can't;

$f_j vp_k$ is the probability it will, $g_j vp_k$ that it won't by chance. The probability, p_{k+1}, that a second fixation will not have occurred in this location is the sum of the probabilities that it can't (because the codon is no longer variable) and that it won't (because of random processes). Thus, $p_{k+1} = p_k(1 - v + g_j v)$ or, returning the subscript i,

$$p_{i, k+1} = p_{i, k}(1 - f_j v) \tag{4}$$

is the probability that the codon fixing the ith mutation will not have received a second mutation after the $k^{th} + 1$ subsequent fixation. Since we know that the mutation following the ith is not fixed in the ith fixing codon with probability $1 - f_i = g_i$, then we also know that $p_{i1} = g_i$. Thus, knowing p_{i1}, v and all f_i, permits one to calculate all values of p_{ik}. This proves to be $p_{ik} = g_i \prod\limits_{j=i+1}^{i+k-1} (1 - f_j v)$.

Since f_i is the probability that the next fixation will be fixed in any particular one of the previously mutated covarions, it must then also be the probability that the first subsequent mutation fixed will be in the same codon as fixed the ith. We know further that $f_j vp_{ik}$ is the probability that the second fixation will occur in the ith fixing codon on the $j^{th} + 1$ trial given that the ith fixing codon failed to fix a second

---

3 Note that for large values of i, $e_{i+1} \cong e_i$. Thus setting $\hat{e} = \hat{e}gv + 1$ yields the limiting value $\hat{e} = 1/(1 - gv)$.

mutation in the previous k trials. Therefore,

$$s_{ih} = f_i + \sum_{k=1}^{k-1} f_j v p_{ik} \qquad (5)$$

is the probability that, in the course of h subsequent fixations, a second fixation will occur in the same codon as fixed the ith. As before, $j = i + k$. Note that $s_{i1} = f_i$. The recursion relation for s is $s_{ih+1} = s_{ih} + f_j v p_{ih}$. Note also that the probability that a second fixation occurs, does not preclude a third observable fixation in the same codon.

If a total of m mutations has been fixed, how many codons will have had two or more fixations? Consider the ith fixation. The probability that the ith fixation is an initial fixation is r_i and that the ith fixation will have a second fixation in the same codon in the following $j = m - i$ fixations is s_{ij}. Thus, the expectation that the ith fixation is in a previously unmutated codon and will be followed by another fixation in the same codon is $r_i s_{ij}$ and the expected number of codons fixing two or more mutations, given m total mutations fixed,

$$\text{is } D = \sum_{i=1}^{m-1} r_i s_{ij}.$$

A computer program has been prepared that calculates D (double fixations) for all values of m for any given set of values for c and v. These D are compared with the number of double mutations found in the descent of 29 species of cytochrome c to determine, by a chi-squared estimate, how well the assumed values of c and v fit the observed data. Trial and error techniques were utilized to obtain the best fitting values of c and v.

References

Fitch, W. M.: Systematic zoology. In press (1971a).
— Haematologie und Bluttransfusion. Munich: J. F. Lehman 1971b (in press).
— Margoliash, E.: Brookhaven Symp. in Biol. 21, 217–242 (1968).
— Markowitz, E.: Biochem. Genet. 4, 579–593 (1970).

Walter M. Fitch
Department of Physiological Chemistry
University of Wisconsin
Madison, Wisconsin, 53706