J. Mol. Evol. 16, 111–120 (1980)

Journal of
Molecular Evolution

A Simple Method for Estimating Evolutionary Rates of Base Substitutions Through Comparative Studies of Nucleotide Sequences

Motoo Kimura

National Institute of Genetics, Mishima 411, Japan

Summary. Some simple formulae were obtained which enable us to estimate evolutionary distances in terms of the number of nucleotide substitutions (and, also, the evolutionary rates when the divergence times are known). In comparing a pair of nucleotide sequences, we distinguish two types of differences; if homologous sites are occupied by different nucleotide bases but both are purines or both pyrimidines, the difference is called type I (or "transition" type), while, if one of the two is a purine and the other is a pyrimidine, the difference is called type II (or "transversion" type). Letting P and Q be respectively the fractions of nucleotide sites showing type I and type II differences between two sequences compared, then the evolutionary distance per site is $K = -(1/2) \ln \{(1-2P-Q)\sqrt{1-2Q}\}$. The evolutionary rate per year is then given by $k = K/(2T)$, where T is the time since the divergence of the two sequences. If only the third codon positions are compared, the synonymous component of the evolutionary base substitutions per site is estimated by $K'_S = -(1/2) \ln (1-2P-Q)$. Also, formulae for standard errors were obtained. Some examples were worked out using reported globin sequences to show that synonymous substitutions occur at much higher rates than amino acid-altering substitutions in evolution.

Key words: Molecular evolution – Evolutionary distance estimation – Synonymous substitution rate

During the last few years, rapid sequencing of DNA has become feasible, and data on nucleotide sequences of various parts of the genome in diverse organisms have started to accumulate at an accelerated pace. Each new report on such sequences

Contribution No. 1330 from the National Institute of Genetics, Mishima, 411 Japan

0022–2844/80/0016/0111/$ 02.00

invites evolutionary considerations through comparative studies. Therefore, it is desirable if good statistical methods are established for estimating evolutionary distances between homologous sequences in terms of the number of nucleotide base substitutions.

Recently, Miyata and Yasunaga (1980) proposed a new method for this purpose. Theii method consists in tracing, through successive one step changes, all the possible paths (restricted by the assumption of "the minimum substitution number") for each pair of codons compared, giving each step its due weight based on relative acceptance rates of amino acid substitutions. Using this method they succeeded in revealing some interesting properties of synonymous base substitutions. One drawback of their method, however, is that it is too tedious.

The purpose of this note is to derive some simple formulae which enable us to obtain reliable estimates on the evolutionary distances between two nucleotide sequences compared. The present method of estimation is not only handy but also has the merit of incorporating the possibility that sometimes "transition" type substitutions may occur more frequently than "transversion" type substitutions.

Let us compare two homologous sequences and suppose that n nucleotide sites are involved. In what follows, we express sequences in terms of RNA codes, so that the four bases are designated by letters U, C, A and G.

Now, we fix our attention on one of the n sites, and investigate how the homologous sites in two species differentiate from each other in the course of evolution, starting from a common ancestor T years back. Since there are 4 possibilities with respect to a base occupied at each site, there are 16 combinations of base pairs when the homologous sites in two species are compared. These are listed in Table 1. For example, UU in the first line represents the case in which homologous sites in the first and second species are both occupied by base U. Similarly, UC in the second line represents the case in which homologous sites in the first and second species are occupied by U and C.

Table 1. Types of nucleotide base pairs occupied at homologous sites in two species. Type I difference includes four cases in which both are purines or both are pyrimidines (line 2). Type II difference consists of eight cases in which one of the bases is a purine and the other is a pyrimidine (lines 3 and 4).

Same	UU	CC	AA	GG	Total
(Frequency)	(R_1)	(R_2)	(R_3)	(R_4)	(R)
Different,					
Type I	UC	CU	AG	GA	Total
(Frequency)	(P_1)	(P_1)	(P_2)	(P_2)	(P)
	UA	AU	UG	GU	
Different, Type II	(Q_1)	(Q_1)	(Q_2)	(Q_2)	Total
	CA	AC	CG	GC	(Q)
(Frequency)	(Q_3)	(Q_3)	(Q_4)	(Q_4)	

When the homologous sites are occupied by different bases, we shall distinguish two types of differences, namely, type I and type II differences. As shown in Table 1, the type I difference includes 4 cases in which both are either purines or pyrimidines, and the type II difference includes 8 cases in which one is a purine and the other is a pyrimidine.

Let α and β be the rates of base substitutions as shown in Fig. 1. In other words, α is the rate of transition type substitutions, and 2β is that of transversion type substitutions, so that the total rate of substitutions per site per unit time (year) is $k = \alpha + 2\beta$. Note that α and β refer to evolutionary rates of mutant substitutions in the species rather than the ordinary mutation rates at the level of an individual. To simplify the following treatments, we assume that these rates are equal for all bases as shown in Fig. 1.

We now define two probabilities denoted by P and Q, where P is the probability of homologous sites showing a type I difference, while Q is that of these sites showing a type II difference. More detailed specifications of probabilities of individual combinations of bases are given in Table 1. For example, the probability of UC (and also CU) is denoted by P_1. Let T be the time since divergence of the two species (measured in years), and denote by P(T) and Q(T) the probabilities of type I and type II differences at time T. Note that the probability of identity at homologous sites which we denote by R(T) is equal to $1 - P(T) - Q(T)$.

Then, we can derive the equations for P and Q at time $T + \Delta T$ in terms of P, Q, and R at time T as follows, where ΔT stands for the length of a short time interval. Consider a particular base pair of type I, say UC, which occurs with probability $P_1\ (T + \Delta T)$ (see second line in Table 1). We can distinguish three ways by which UC at time $T + \Delta T$ is derived from various base pairs at time T.

(i) UC is derived from UC (which occurs with probability $P_1(T)$) when both U and C remain unchanged. Since the probability of occurrence of substitution per site during a short time interval of length ΔT is $(\alpha + 2\beta)\Delta T$, the probability of no change occurring at both homologous sites is $[1 - (\alpha + 2\beta)\Delta T]^2$ so that the contribution coming from this class is $[1 - (\alpha + 2\beta)\Delta T]^2 P_1(T)$, which reduces to $[1 - 2(\alpha + 2\beta)\Delta T]P_1(T)$ if we neglect small terms of the order $(\Delta T)^2$. We also ignore the rare possibility that CU changes to UC by a double change.

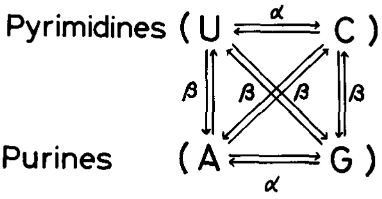

Fig. 1. Scheme of evolutionary base substitutions and their rates per unit time

(ii) UC is derived from UU when U in the second species is replaced by C, and from CC when C in the first species is replaced by U (see the first line in Table 1). Since UU and CC occur with respective frequencies $R_1(T)$ and $R_2(T)$, and since each changes to UC at the rate α, the total contribution coming from this class is $\alpha\Delta T\ [R_1(T) + R_2(T)]$. We neglect cases in which substitutions occur simultaneously at both sites because such probabilities are of the order of $(\Delta T)^2$.

(iii) UC can also be derived from UA, UG, AC and GC (see the third line in Table 1) which occur with respective frequencies $Q_1(T)$, $Q_2(T)$, $Q_3(T)$ and $Q_4(T)$, and each of which changes to UC at the rate β. The contribution of this class to UC at $T + \Delta T$ is $\beta\Delta T\,[Q_1(T) + Q_2(T) + Q_3(T) + Q_4(T)]$ or $\beta\Delta T \cdot Q(T)/2$.

Combining all these contributions coming from various base pair classes at time T, we get

$$P_1(T + \Delta T) = [1 - (2\alpha + 4\beta)\,\Delta T]\,P_1(T) + \alpha\Delta T[R_1(T) + R_2(T)] + \beta\Delta T \cdot Q(T)/2 \ .$$

Similarly, for the base pair AG,

$$P_2(T + \Delta T) = [1 - (2\alpha + 4\beta)\,\Delta T]\,P_2(T) + \alpha\Delta T[R_3(T) + R_4(T)] + \beta\Delta T \cdot Q(T)/2$$

Summing these two equations, and noting $P(T) = 2P_1(T) + 2P_2(T)$ and $R(T) = R_1(T) + R_2(T) + R_3(T) + R_4(T) = 1 - P(T) - Q(T)$, and writing $\Delta P(T) = P(T + \Delta T) - P(T)$ we get

$$\Delta P(T)/\Delta T = 2\alpha - 4(\alpha + \beta)\,P(T) - 2(\alpha - \beta)\,Q(T) \ . \tag{1}$$

Carrying out a similar series of calculations for base pairs of type II, we obtain

$$\Delta Q(T)/\Delta T = 4\beta - 8\beta Q(T) \ . \tag{2}$$

From these two finite difference equations (Eqs. 1 and 2), we obtain the following set of differential equations

$$\begin{aligned} \frac{dP(T)}{dT} &= 2\alpha - 4(\alpha + \beta)\,P(T) - 2(\alpha - \beta)\,Q(T) \\ \frac{dQ(T)}{dT} &= 4\beta - 8\beta Q\,(T) \end{aligned} \tag{3}$$

The solution of this set of equations which satisfies the condition

$$P(0) = Q(0) = 0 \ , \tag{4}$$

i.e., no base differences exist at T = 0, is as follows.

$$P(T) = \frac{1}{4} - \frac{1}{2}e^{-4(\alpha+\beta)T} + \frac{1}{4}e^{-8\beta T} \tag{5}$$

$$Q(T) = \frac{1}{2} - \frac{1}{2}e^{-8\beta T} \ . \tag{6}$$

Writing P_T and Q_T for P(T) and Q(T), we get, from these two equations,

$$4(\alpha + \beta)T = -\log_e (1 - 2P_T - Q_T) \tag{7}$$

and

$$8\beta T = -\log_e (1 - 2Q_T) \ , \tag{8}$$

so that

$$4\alpha T = -\log_e (1 - 2P_T - Q_T) + (1/2)\log_e (1 - 2Q_T) \ . \tag{9}$$

Since the rate of evolutionary base substitutions per unit time is

$$k = \alpha + 2\beta \ ,$$

the total number of substitutions (including revertant and superimposed changes) per site which separate the two species (and therefore involve two branches each with length T) is

$$K = 2Tk = 2\alpha T + 4\beta T \quad ,$$

where αT and βT are given by Eqs. (8) and (9). Then, omitting the subscript T from P_T and Q_T, we obtain

$$K = -\frac{1}{2}\log_e \{(1 - 2P - Q)\sqrt{1 - 2Q}\} \quad . \tag{10}$$

It is remarkable that, as can be seen from Eqs. (3), at equilibrium $2P = Q = 1/2$, even when $\alpha \neq \beta$.

This equation may be used to estimate the evolutionary distance between two sequences in terms of the number of base substitutions per site that have occurred in the course of evolution extending over T years. In this equation, $P = n_1/n$ and $Q = n_2/n$, where n_1 and n_2 are respectively the numbers of sites for which two sequences differ from each other with respect to type I ("transition" type) and type II ("transversion" type) substitutions and n is the total number of sites compared.

In the special case of $\alpha = \beta$, Eqs. (5) and (6) reduce to

$$P_T = Q_T/2 = (1 - e^{-8\alpha T})/4 \quad . \tag{11}$$

Then, substituting $P = Q/2$ in (10), we get

$$K = -\frac{3}{4}\log_e (1 - \frac{4}{3}\lambda) \quad , \tag{12}$$

where $\lambda = P + Q = 3Q/2$ is the fraction of sites for which two sequences differ from each other. This formula is well-known (see Kimura and Ohta 1972), and it was first obtained by Jukes and Cantor (1969). However, in actual situations, particularly when the third positions of codons are compared, P is often larger than Q, and therefore the assumption of $\alpha = \beta$ or $P = Q/2$ is not always realistic. This is one reason why the new formula (10) is better than (12). Eq. (10) also has a desirable property in that as P and Q get small, it converges to $K = P + Q$ independent of α and β.

Since a large fraction of substitutions at the third positions are synonymous, that is, they do not cause amino acid changes, it would be interesting to estimate the synonymous component of the substitution rate at this position.

As is evident from the standard RNA code table, for a given pair of bases in the first and the second codon positions, roughly speaking there are two situations; either the third position is completely synonymous (four-fold degeneracy) or synonymy is restricted within purines or pyrimidines (two-fold degeneracy). These two situations occur roughly in equal numbers. Thus, the synonymous component of substitutions at the third position which we denote by k_S' may be estimated by

$$k_S' = \frac{1}{2}(\alpha + 2\beta) + \frac{1}{2}\alpha = \alpha + \beta \quad . \tag{13}$$

Let $K_S' = 2Tk_S' = 2(\alpha + \beta)T$ be the synonymous component of the distance, then, we get

$$K_S' = -\frac{1}{2}\log_e (1 - 2P - Q) \quad , \tag{14}$$

if we apply Eq. (7), so that the synonymous component of substitution rate per unit time may be estimated from $k_S' = K_S'/(2T)$. Writing $k'_{nuc(S)}$ rather than k_S' to emphasize that this refers to the rate per nucleotide site, we get, using Eq. 14, the following formula.

$$k'_{nuc(S)} = -\frac{1}{4T} \log_e (1 - 2P - Q) \quad . \tag{15}$$

The corresponding formula for amino acid-altering substitutions may be obtained by

$$k_{nuc(A)} = K''/(2T) \quad , \tag{16}$$

where K" is the evolutionary distance computed by applying Eq. (10) to a set of the first and second positions of codons (i.e., by excluding the third positions in the sequence comparison).

We can also derive formulae for the error variances of the estimates K and K_S'. Let δK, δP and δQ be respectively small changes in K, P and Q. Then

$$\delta K = a\delta P + b\delta Q$$

where

$$a = \frac{1}{1 - 2p - Q} \tag{17}$$

and

$$b = \frac{1}{2}\left(\frac{1}{1 - 2P - Q} + \frac{1}{1 - 2Q}\right) \quad , \tag{18}$$

so that

$$\sigma_K{}^2 = E\,\{(\delta K)^2\} = a^2 E\,\{(\delta P)^2\} + 2abE\,\{\delta P\delta Q\} + b^2 E\,\{(\delta Q)^2\} \quad ,$$

where E stands for the expectation operator. Then noting that the sampling variances and the covariance are $E\,\{(\delta P)^2\} = P(1 - P)/n$, $E\,\{(\delta Q)^2\} = Q(1 - Q)/n$ and $E\,\{\delta P\delta Q\} = - PQ/n$. The standard error of K is then

$$\sigma_K = \frac{1}{\sqrt{n}} \left|\sqrt{(a^2P + b^2Q) - (aP + bQ)^2}\right| \quad . \tag{19}$$

In a similar manner, we can derive the standard error of K_S' which turns out to be as follows.

$$\sigma_{K'S} = \frac{\sqrt{4P + Q - (2P + Q)^2}}{2(1 - 2P - Q)\sqrt{n}} \quad . \tag{20}$$

As an example, let us compare the nucleotide sequence of the rabbit β globin (Efstratiadis and Kafatos 1977) with that of chicken β globin (Richards et al. 1979). There are 438 nucleotide sites that can be compared, corresponding to 146 amino acid sites (codons). Among these sites, we find that there are 58 sites for which these two sequences have type I differences, and 63 sites with type II differences. Thus, $P = 0.132$, $Q = 0.144$ and we obtain $K = 0.348$. Mammals and birds probably diverged during the Carboniferous period (see Romer 1968), so we tentatively take $T = 300 \times 10^6$ years. The evolutionary rate per site is then $k_{nuc} = K/(2T) = 0.58 \times 10^9$ per year. This is the overall rate per site, but it is much more interesting to estimate separately the evolutionary rates for the three codon positions. For the first position, there are 146 nucleotide sites compared, and we find P = 15/146 and Q = 21/146, giving $K_1 = 0.300$, where subscript 1 denotes that it refers to the first codon position. Similarly, for the second position, we find P = 7/146 and Q = 18/146 so that $K_2 = 0.195$. Finally, for the third position, P = 36/146 and Q = 24/146, and we get $K_3 = 0.635$, which is much higher than the corresponding estimates for the first and second positions. We can also estimate the synonymous component of the evolutionary distance per third codon position. From Eq. (14), this turns out to be $K_S' = 0.535$.

In Table 2, results of similar calculations are listed for various comparisons involving the human β (Marotta et al. 1977) and the mouse β globin sequences (Konkel et al. 1978), in addition to the chicken and rabbit β-globin sequences. Except for the last two comparisons involving the abnormal, globin like α-3 gene (Nishioka et al. 1980) it is clear that the relationship $K_2 < K_1 < K_3$ holds generally; the evolutionary mutant substitutions are most rapid at the third position, and this is followed by the first position and then, at the second position the substitutions are the slowest.

This can be readily interpreted by the neutral theory of molecular evolution (Kimura 1968; King and Jukes 1969; see also Kimura 1979) as follows. Among the three codon positions, base substitutions at the second positions tend to produce more drastic changes in the physico-chemical properties of amino acids than those at the first positions. Take for example a codon for Pro (CCN). Substitutions for base C at the first position of U, A, and G, lead respectively to Ser, Thr and Ala. In terms of Miyata's distance (based on polarity and volume differences between an amino acid pair; see Miyata et al. 1979), they are 0.56, 0.87 and 0.06 units appart from Pro, with the average distance of about 0.5. On the other hand, the corresponding average distance resulting from substitutions for C at the second position of the codon turns out to be about 2.5.

This means that mutational changes at the first position have a higher chance of not being harmful (i.e., selectively neutral or equivalent) than those at the second position, and therefore, have a higher chance of being fixed in the species by random drift (Kimura and Ohta 1974). This type of reasoning applies more forcibly to the third position (as compared with the first and the second positions), since a majority of mutational changes at this position do not cause amino acid changes.

The ratio per site of synonymous to amino acid-altering substitutions as measured by $2K_S'/(K_1 + K_2)$, is 4.17 for the human β-rabbit β comparison ($T = 8 \times 10^7$ years), 2.16 for the chicken β-rabbit β comparison ($T = 3 \times 10^8$ years) but only 1.41 for the rabbit α-rabbit β comparison ($T = 5 \times 10^8$ years). It looks as if synonymous substitutions occur more frequently during the later (i.e. more recent) stages of globin evolution than its early stages, but such a tendency is probably more apparent than real. In my opinion, this is due to lower detectability of synonymous substitutions for more remote comparisons; as more and more synonymous substitutions accumulate at the third positions of codons, it becomes progressively difficult to detect all of them. In fact, as compared with the first and the second positions, the third position shows marked deviation of the base composition from equality (e.g., G 36%, C 30%, U 27%, A 7% in human β) so that the present method of estimation may become imprecise for a very large value of T.

Table 2. Evolutionary distances in terms of the number of base substitutions estimated for several comparisons of globin sequences. K_1, K_2, and K_3 respectively denote the number of base substitutions at the first, second, and third positions of codons, while K_S' stands for the estimated number of substitutions due to synonymous changes in the third position. Estimated values of these parameters together with their standard errors are listed. The primary sequence of rabbit α-globin is taken from Heindell et al. (1978)

Comparison	Evolutionary distances per nucleotide site			
	K_1	K_2	K_3	K_S'
Human β vs. Mouse β	0.17 ± 0.04	0.13 ± 0.03	0.34 ± 0.06	0.28 ± 0.05
Rabbit β vs. Mouse β	0.16 ± 0.04	0.13 ± 0.03	0.43 ± 0.07	0.36 ± 0.07
Human β vs. Rabbit β	0.06 ± 0.02	0.06 ± 0.02	0.28 ± 0.06	0.25 ± 0.05
Chicken β vs. Rabbit β	0.30 ± 0.05	0.19 ± 0.04	0.64 ± 0.11	0.53 ± 0.10
Rabbit α vs. Rabbit β	0.54 ± 0.09	0.44 ± 0.07	0.90 ± 0.15	0.69 ± 0.13
Rabbit α vs. Mouse α-1	0.12 ± 0.03	0.11 ± 0.03	0.54 ± 0.09	0.47 ± 0.09
Rabbit α vs. Mouse α-3	0.27 ± 0.06	0.28 ± 0.06	0.69 ± 0.13	0.56 ± 0.12
Mouse α-1 vs. Mouse α-3	0.16 ± 0.04	0.20 ± 0.05	0.30 ± 0.06	0.22 ± 0.05

The last two comparisons in Table 2 involve the globin-like α-3 gene recently sequenced in the mouse (Nishioka et al. 1980). This gene completely lacks two intervening sequences normally present in all the α and β globin genes, and it does not encode globin. In other words, it is inactive in the production of stable mRNA. However, from a comparison of this sequence with the normal, adult mouse α globin (α-1) and the rabbit α globin nucleotide sequences, it is evident that this gene evolved from a normal ancestral α globin gene through duplication and subsequent loss of its intervening sequences. This must have occurred after the mouse and the rabbit diverged from their common ancestor some 80 million years ago. This α-globin-like gene acquired in its coding region a number of insertions and deletions of nucleotides. In making sequence comparisons, therefore, I chose only those codons of the α-3 gene which do not contain such changes and which are either identical with or differ only through base substitutions from the corresponding (homologous) codons of the mouse α-1 and the rabbit α genes.

As seen from the last three lines of Table 2, this unexpressed α-3 gene evolved at a much faster rate than its normal counterpart (mouse α-1 gene), particularly with respect to the first and the second codon positions. This is easy to understand from the neutral theory. Under a normal situation, each gene is subject to a selective constraint coming from the requirement that the protein which it produces must function normally. Evolutionary changes are restricted within such a set of base substitutions. However, once a gene is freed from this constraint, as is the case for this globin-like α-3 gene, practically all the base substitutions in it become indifferent to Darwinian fitness, and the rate of base substitutions should approach the upper limit set by the mutation rate (This holds only if the neutral theory is valid, but not if the majority of base substitutions are driven by positive selection; see Kimura 1977). If the rates of synonymous substitutions are not very far from this limit (Kimura 1977) we may expect that the rates of evolution of a "dead gene" are roughly equal to those of synonymous substitutions. Recently Miyata (personal communication) computed the evolutionary rate of a mouse pseudo alpha globin gene (ψα 30.5) which was sequenced by Vanin et al. (1980) and which appears to be essentially equivalent to the α-3 gene. He also obtained a result supporting this prediction.

Finally, it would be interesting to estimate base substitution rates in non-coding regions such as introns. I use data presented by van Ooyen et al. (1979) who investigated similarity between the nucleotide sequences of rabbit and mouse β-globin genes. They list (see their Table 2) separately the numbers of "transition" and "transversion" type differences between the homologous parts of these sequences. For the small introns excluding 5 gaps that amount to 6 nucleotides, $P = 27/113$, $Q = 18/113$ and $n = 113$. Using Eqs. 10 and 19, we get $K = 0.60 \pm 0.12$. This is not significantly different from the substitution rate at the third position $K_3 = 0.43 \pm 0.07$. The large introns of rabbit and mouse β globin genes differ considerably in length, being separated from each other by 14 gaps (determined by optimization of alignment of the sequences) which amount to 109 nucleotides. Excluding these parts, $P = 113/557$ and $Q = 179/557$, from which we get $K = 0.90 \pm 0.07$. This value is significantly larger than K_3. It is likely, as pointed out by van Ooyen et al. (1979) that insertions and deletions occur rather frequently in this part in addition to point mutations, and that they inflate the estimated value of the "nucleotide substitution rate," since a majority of these changes may also be selectively neutral and subject to random fixation by genetic drift.

Aknowledgments. I thank Drs. Y. Tateno, N. Takahata, and J.F. Crow for reading carefully the first draft of this paper and for giving many useful suggestions for improving its presentation.

References

Efstratiadis A, Kafatos FC (1977) Cell 10:571–585

Heindell HC, Liu A, Paddock GV, Studnicka GM, Salser WA (1978) Cell 15:43–54

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed), Mammalian protein metabolism, II. New York, Academic Press, p 21

Kimura M (1968) Nature 217:624–626

Kimura M (1977) Nature 267:275–276

Kimura M (1979) Sci Am 241 (No.5, Nov.):94–104

Kimura M, Ohta T (1972) J Mol Evol 2:87–90

Kimura M, Ohta T (1974) Proc Natl Acad Sci USA 71:2848–2852

King JL, Jukes TH (1969) Science 164:788–798

Konkel DA, Tilghman SM, Leder P (1978) Cell 15:1125–1132

Marotta CA, Wilson JT, Forget BG, Weissman SM (1977) J Biol Chem 252:5040–5053

Miyata T, Miyazawa S, Yasunaga T (1979) J Mol Evol 12:219–236

Miyata T, Yasunaga T (1980) J Mol Evol 16:23–36

Nishioka Y, Leder A, Leder P (1980) Proc Natl Acad Sci USA 77:2806–2809

Richards RI, Shine J, Ullrich A, Wells JRE, Goodman HM (1979) Nucleic Acids Res 7: 1137–1146

Romer AS (1968) The procession of life. Weidenfeld and Nicolson, London.

Vanin EF, Goldberg GI, Tucker PW, Smithies O (1980). Nature 286:222–226

van Ooyen A, van den Berg J, Mantel N, Weissmann C (1979) Science 206:337–344

Received July 24, 1980