Simple Methods for Testing the Molecular Evolutionary Clock Hypothesis

Fumio Tajima

Department of Population Genetics, National Institute of Genetics, Mishima, Shizuoka-ken 411, Japan

Manuscript received February 15, 1993
Accepted for publication June 17, 1993

ABSTRACT

Simple statistical methods for testing the molecular evolutionary clock hypothesis are developed which can be applied to both nucleotide and amino acid sequences. These methods are based on the chi-square test and are applicable even when the pattern of substitution rates is unknown and/or the substitution rate varies among different sites. Furthermore, some of the methods can be applied even when the outgroup is unknown. Using computer simulations, these methods were compared with the likelihood ratio test and the relative rate test. The results indicate that the powers of the present methods are similar to those of the likelihood ratio test and the relative rate test, in spite of the fact that the latter two tests assume that the pattern of substitution rates follows a certain model and that the substitution rate is the same among different sites, while such assumptions are not necessary to apply the present methods. Therefore, the present methods might be useful.

WHETHER the molecular evolutionary clock hypothesis (ZUCKERKANDL and PAULING 1965) holds or not is one of the most important issues in molecular evolution [*e.g.*, see KIMURA (1983)]. This hypothesis may not hold if natural selection is operating, or if mutation rate is not constant per year.

Although there are several methods for testing this hypothesis, all methods have some problems. (i) Absolute rate of molecular evolution cannot be estimated unless we know the divergence time. (ii) We must know the pattern of substitution rates to estimate the number of nucleotide (or amino acid) substitutions per site. (iii) We must also know the variation of substitution rates among different sites to estimate the number of substitutions per site. (iv) The phylogenetic relationship among nucleotide (or amino acid) sequences must be known before the test is performed.

The relative rate test (SARICH and WILSON 1967; WU and LI 1985) and the likelihood ratio test (MUSE and WEIR 1992) have overcome the first problem, but still have the other problems. WU and LI's (1985) test assumes that the pattern of substitution rates follows KIMURA's (1980) substitution model, and MUSE and WEIR's (1992) test assumes more sophisticated substitution models. However, we usually do not know whether these substitution models are correct, although they may be good approximations in some cases. The third problem is more difficult to overcome. Although it may be possible to assume that the substitution rate varies from site to site according to some specific distributions such as the gamma distribution (JIN and NEI 1990; LI *et al.* 1990), it is not clear whether such an assumption significantly improves the test. In order to conduct the relative rate test and the likelihood ratio test, the phylogenetic relationships among three sequences must be known. For example, the phylogenetic relationships among eutherian mammals are controversial (EASTEAL 1990; GRAUR, HIDE and LI 1991; NOVACEK 1992). If the phylogenetic relationships assumed were wrong, we would obtain the wrong conclusion.

In this report I shall present methods for testing the molecular evolutionary clock hypothesis, which overcome the above problems. These methods are so simple that they can be easily applied to any nucleotide (or amino acid) sequences as long as alignment is possible.

THEORY

Suppose that we have three nucleotide sequences, say sequences 1, 2 and 3, which are already aligned, and consider only the sites without gaps. Let n_{ijk} be the observed number of sites where sequences 1, 2 and 3 have nucleotides i, j and k, respectively. The subscripts i, j and k take the values 1, 2, 3 and 4 for nucleotides A, G, C and T, respectively.

Case where the outgroup is known: We assume that sequence 3 is the outgroup. Then, the expectation of n_{ijk} must be equal to that of n_{jik}, *i.e.*,

$$E(n_{ijk}) = E(n_{jik}), \tag{1}$$

whatever the substitution model is, and even if the substitution rate varies among different sites. If this equality does not hold, we can conclude that the rate is not constant per year. This is the basis of the present method.

To simplify the analysis, we consider only the sites in which exactly two different types of nucleotides exist in the three sequences, and define the observed

number of sites in which nucleotides in sequence 1 are different from those in sequences 2 and 3 by m_1. In the same way we define m_2 and m_3. Namely, we define m_1, m_2, and m_3 as follows:

$$m_1 = \sum_i \sum_{j \neq i} n_{ijj} \tag{2a}$$

$$m_2 = \sum_i \sum_{j \neq i} n_{jij} \tag{2b}$$

$$m_3 = \sum_i \sum_{j \neq i} n_{jji} \tag{2c}$$

When sequence 3 is the outgroup, it is clear from (1) that the expectation of m_1 must be equal to the expectation of m_2, *i.e.*,

$$E(m_1) = E(m_2). \tag{3}$$

This equality can be tested by using chi-square. Namely,

$$\chi^2 = \frac{(m_1 - m_2)^2}{m_1 + m_2} \tag{4}$$

approximately follows the chi-square distribution with one degree of freedom (1 d.f.), noting that the expectations of m_1 and m_2 are both $(m_1 + m_2)/2$ when $m_1 + m_2$ is given. I call this method one-degree-of-freedom method (1D method). Another derivation of (4) is given in the APPENDIX.

For example, when $m_1 = 50$ and $m_2 = 30$, we have $\chi^2 = (50 - 30)^2/(50 + 30) = 5.000$ and $P(\chi^2 \geq 3.841) = 0.05$ for 1 d.f., so that we reject the molecular evolutionary clock hypothesis at the 5% level. This test is analogous to a sing test, and one could also calculate the exact tail probability under the null hypothesis given by (3).

The equality holds if the rate is constant, but it may hold even if the rate is not constant. Therefore, the present test might be conservative.

In some sequences such as mitochondrial DNA sequences, transitional changes occur more often than transversional changes. In such cases it might be better to classify nucleotide differences into transitional and transversional differences. Namely, m_i is divided into the number of sites (s_i) for transitional differences and the number of sites (v_i) for transversional differences. Noting that A, G, C and T are denoted by 1, 2, 3 and 4, respectively, we define

$$s_1 = n_{211} + n_{122} + n_{433} + n_{344}, \tag{5a}$$

$$v_2 = m_1 - s_1, \tag{5b}$$

$$s_2 = n_{121} + n_{212} + n_{343} + n_{434}, \tag{5c}$$

$$v_2 = m_2 - s_2, \tag{5d}$$

$$s_3 = n_{112} + n_{221} + n_{334} + n_{443}, \tag{5e}$$

$$v_3 = m_3 - s_3. \tag{5f}$$

When sequence 3 is the outgroup, it is clear from (1) that $E(s_1) = E(s_2)$ and $E(v_1) = E(v_2)$, and that

$$\chi^2 = \frac{(s_1 - s_2)^2}{s_1 + s_2} + \frac{(v_1 - v_2)^2}{v_1 + v_2} \tag{6}$$

approximately follows the chi-square distribution with two degrees of freedom (2 d.f.), noting that the expectations of s_1 and s_2 are both $(s_1 + s_2)/2$ and those of v_1 and v_2 are both $(v_1 + v_2)/2$ when $s_1 + s_2$ and $v_1 + v_2$ are given. I call this method two-degrees-of-freedom method (2D method). Another derivation of (6) is given in the APPENDIX.

For example, when $s_1 = 30$, $v_1 = 20$, $s_2 = 25$ and $v_2 = 5$, we have $\chi^2 = (30 - 25)^2/(30 + 25) + (20 - 5)^2/(20 + 5) = 9.455$ and $P(\chi^2 \geq 9.210) = 0.01$ for 2 d.f. so that we reject the molecular evolutionary clock hypothesis at the 1% level.

Case where the outgroup is unknown: Even when the outgroup is unknown, we can still perform the chi-square test. In this case I suggest the following algorithm: (i) Choose two sequences which have two smallest values among m_1, m_2 and m_3. Suppose $m_2 \leq m_3 \leq m_1$. Then, we choose sequences 2 and 3. (ii) Compute a chi-square value by using these two sequences, in this case either by $\chi^2 = (m_2 - m_3)^2/(m_2 + m_3)$ for 1 d.f. or by $\chi^2 = (s_2 - s_3)^2/(s_2 + s_3) + (v_2 - v_3)^2/(v_2 + v_3)$ for 2 d.f. (iii) Perform the chi-square test as usual. I call the algorithm for 1 d.f. one-degree-of-freedom-with-no-outgroup method (1DN method) and that of 2 d.f. two-degrees-of-freedom-with-no-outgroup method (2DN method), respectively.

For example, when $m_1 = 50$ ($s_1 = 30$, $v_1 = 20$), $m_2 = 30$ ($s_2 = 25$, $v_2 = 5$), and $m_3 = 45$ ($s_3 = 25$, $v_3 = 20$), we choose sequences 2 and 3 because of $m_2 < m_3 < m_1$. Then, we have $\chi^2 = (30 - 45)^2/(30 + 45) + 3.0$ for 1DN method (no significant) or $\chi^2 = (25 - 25)^2/(25 + 25) + (5 - 20)^2/(5 + 20) = 9.0$ for 2DN method (significant at the 5% level).

The reason why the 1DN method is appropriate as a statistical test is as follows. Suppose that $m_2 \leq m_3 \leq m_1$, and that there is a significant difference between m_2 and m_3. There are three possibilities in terms of outgroup. (i) If the true outgroup is sequence 1, then this test is the same as the 1D method. Therefore, we reject the molecular evolutionary clock hypothesis. (ii) If the true outgroup is sequence 3, then m_2 is significantly different from m_1 because of $m_3 \leq m_1$. Therefore, we reject the hypothesis. (iii) If the true outgroup is sequence 2, then the substitution rate in the lineage leading to the outgroup sequence is significantly slower than that in the lineages leading to the other two sequences. Therefore, we reject the hypothesis. In all the cases we reject the hypothesis, so that this method is appropriate as a statistical test.

In the case where s_i and v_i are used for the chi-square test, it may not be very clear whether the 2DN

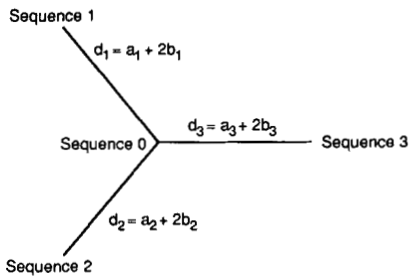

FIGURE 1.—Phylogenetic relationship among three sequences used for computer simulations.

method is applicable. As will be shown in the next section, results of computer simulations suggest that this method is also applicable.

COMPUTER SIMULATION

In order to know whether the present methods are appropriate as statistical tests, as well as to know the powers of the tests, I have conducted computer simulations.

Substitution model: I used KIMURA's (1980) substitution model to generate sequence data. Let $P_{ij}(t)$ be the probability that a site originally having nucleotide i is occupied by nucleotide j after a time t. The subscripts i and j take the values 1, 2, 3 and 4 for nucleotides A, G, C and T, respectively, as before. The substitution rates are defined by $P_{12}(dt) = P_{21}(dt) = P_{34}(dt) = P_{43}(dt) = \alpha dt$ for transitional changes and $P_{13}(dt) = P_{14}(dt) = P_{23}(dt) = P_{24}(dt) = P_{31}(dt) = P_{32}(dt) = P_{41}(dt) = P_{42}(dt) = \beta dt$ for transversional changes. Then, $P_{ij}(t)$'s are given by

$$P_{ii}(t) = \frac{1}{4} + \frac{1}{4}e^{-4\beta t} + \frac{1}{2}e^{-2(\alpha+\beta)t}$$

for no change, (7a)

$$P_{ij}(t) = \frac{1}{4} + \frac{1}{4}e^{-4\beta t} - \frac{1}{2}e^{-2(\alpha+\beta)t}$$

for transition, (7b)

$$P_{ij}(t) = \frac{1}{4} - \frac{1}{4}e^{-4\beta t}$$

for transversion, (7c)

where $i \neq j$. I used the phylogenetic tree in Figure 1, where sequence 3 is assumed to be the outgroup. For simplicity, αt and βt for the lineage leading to sequence i are defined to be a_i and b_i, respectively. The number of nucleotide substitutions per site (d) is given by $\alpha t + 2\beta t$, so that the number of substitutions per site (d_i) in the lineage leading to sequence i is given by $a_i + 2b_i$.

First, I generated sequence 0 by assuming that each site is occupied by one of four nucleotides with equal probability. Then, sequences 1, 2 and 3 were generated, according to the above probabilities. Finally, chi-square values for four methods were computed from m_i, s_i and v_i (i = 1, 2 and 3). The length of nucleotide sequence (L) is 250, 500, 1000 or 2000, and the number of replications is 10,000 for each set of parameter values.

Distribution of test statistic: In order to confirm that the present methods are appropriate, the distributions of test statistics for several sets of parameter values were examined. The results of computer simulation are shown in Tables 1 and 2, where the numbers of replications which were rejected at particular significance levels given together with the averages of chi-square values.

Parameter sets (a) and (b) in Table 1 assume that the rate of transitional substitution is the same as that of transversional one, whereas parameter sets (c) and (d) assume that the rate of transitional substitution is eight times higher than that of transversional one. The lineage leading to the outgroup sequence (sequence 3) is 1.2 times longer than the lineages leading to sequences 1 and 2 in the cases of (a) and (c), whereas it is 1.5 times longer in the cases of (b) and (d). We can see from this table that the distributions of chi-square values obtained by using the 1D and 2D methods are close to their expectations, and that the averages of chi-square values are not significantly different from their expectations, i.e., one and two for the 1D and 2D methods, respectively. Therefore, we can conclude that these methods are appropriate. On the other hand, the 1DN and 2DN methods tend to give smaller chi-square values than do the 1D and 2D methods, especially when the outgroup lineage is short and when the length of sequence (L) is short. This means that the 1DN and 2DN methods are appropriate although they are conservative. Therefore we can conclude that these methods are useful when the outgroup is unknown.

Table 2 gives the distributions of chi-square values when the pattern of substitution rates in the outgroup lineage is different from that in the two other lineages. In this case the substitution rate in the lineage leading to sequence 1 is the same as that of sequence 2. This table shows that chi-square values obtained by using the 1D and 2D methods are close to their expectations, so that we can conclude that these two methods are appropriate. Chi-square values obtained by using the 1DN method tend to be smaller than those of the 1D method when the outgroup lineage is short and when the length of sequence (L) is short, as in Table 1, so that the 1DN method might be conservative but useful when the outgroup is unknown. Interesting is the 2DN method, which gives larger chi-square values than does the 2D method, especially when the outgroup lineage is short [$d_1 = d_2 = 0.25$ and $d_3 = 0.3$ in parameter sets (a) and (c)]. This means that in some cases the 2DN method can identify unequal patterns

TABLE 1

Distribution of test statistic when the hypothesis is true

	1D (1DN) Method					2D (2DN) Method				
	Significance level					Significance level				
L	5.0%	2.5%	1.0%	0.5%	$\overline{\chi^2}$	5.0%	2.5%	1.0%	0.5%	$\overline{\chi^2}$
(a) $a_1 = a_2 = b_1 = b_2 = 0.0333$ and $a_3 = b_3 = 0.04$										
250	523	264	118	52	1.01	484	232	79	34	1.99
	(264	120	33	19	0.72	308	145	51	23	1.69)
500	508	249	97	45	0.99	469	226	88	42	2.00
	(286	121	32	16	0.75	352	159	51	22	1.76)
1000	494	262	97	46	1.00	548	263	97	48	2.03
	(334	152	49	22	0.83	425	196	69	35	1.87)
2000	530	263	96	42	1.01	477	227	87	39	1.99
	(437	208	70	28	0.93	432	194	69	34	1.91)
(b) $a_1 = a_2 = b_1 = b_2 = 0.0333$ and $a_3 = b_3 = 0.05$										
250	492	214	83	38	1.00	455	211	78	31	1.99
	(375	150	53	27	0.89	393	183	65	20	1.88)
500	504	221	95	44	0.99	451	217	64	30	1.98
	(457	194	77	35	0.94	418	196	52	23	1.94)
1000	491	242	90	53	0.99	478	218	83	44	1.99
	(485	239	89	52	0.98	478	217	82	42	1.98)
2000	444	223	83	37	0.96	489	241	88	44	1.97
	(444	223	83	37	0.96	489	241	88	44	1.97)
(c) $a_1 = a_2 = 0.2$, $b_1 = b_2 = 0.025$, $a_3 = 0.24$ and $b_3 = 0.03$										
250	531	262	96	48	1.02	502	230	86	46	2.00
	(255	101	32	11	0.71	334	140	42	17	1.70)
500	459	231	83	36	0.98	475	224	86	33	1.99
	(262	109	28	10	0.75	356	156	58	25	1.77)
1000	476	251	95	50	1.01	518	267	101	54	2.02
	(341	157	53	24	0.86	426	213	70	37	1.88)
2000	550	269	114	51	1.04	535	260	108	53	2.05
	(472	215	82	34	0.97	482	234	97	50	1.99)
(d) $a_1 = a_2 = 0.2$, $b_1 = b_2 = 0.025$, $a_3 = 0.3$ and $b_3 = 0.0375$										
250	501	235	97	49	0.99	466	211	78	36	1.96
	(391	166	55	34	0.88	402	182	63	27	1.87)
500	463	230	87	48	0.98	461	235	88	47	1.96
	(426	208	76	41	0.94	449	235	85	45	1.94)
1000	479	240	96	43	1.00	496	248	86	29	2.01
	(475	238	94	42	1.00	492	244	84	28	2.01)
2000	477	223	98	48	0.98	469	238	91	46	1.99
	(477	223	98	48	0.98	469	238	91	46	1.99)

of substitution rates among lineages. For example, in parameter set (a) we reject the hypothesis in about 22% and 33% of cases at the 1% and 5% levels, respectively, when $L = 250$.

In this study, when the pattern of substitution rates is different among lineages, we assume that the molecular evolutionary clock hypothesis does not hold even if the overall substitution rate is the same among lineages, since neither the transitional nor transversional substitution rate is constant. Furthermore, it might be very unlikely that the overall rate is constant when the transitional and transversional rates are different. Therefore, it might not cause any serious problem.

Power of statistical test: In order to know the power of the statistical test, I have conducted computer simulations. The results are given in Table 3, where the transitional rate was assumed to be equal to the transversional rate. The results obtained by using the 1D method are shown together with those obtained by MUSE and WEIR (1992). The 1D method appears to be slightly more powerful than the likelihood ratio test (LR) of MUSE and WEIR (1992) and the relative rate test (WL) of WU and LI (1985). In these sets of parameter values, the power of the 1DN method was essentially the same as that of the 1D method.

The results obtained by using the 1D and 2D methods are given in Table 5 together with those obtained by MUSE and WEIR (1992), where the sets of parameter values in Table 4 were used. The power of the 2D method appears to be similar to that of the two-parameter likelihood ratio test (LR2P) of MUSE and WEIR (1992), and that the power of the 1D method appears to be similar to those of the single parameter likelihood ratio test (LR1P) of MUSE and WEIR (1992)

TABLE 2

Distribution of test statistic when the pattern of substitution rates in the outgroup lineage is different from that in the lineages leading to the other two sequences

	1D (1DN) Method					2D (2DN) Method				
	Significance level					Significance level				
L	5.0%	2.5%	1.0%	0.5%	$\overline{\chi^2}$	5.0%	2.5%	1.0%	0.5%	$\overline{\chi^2}$
(a) $a_1 = a_2 = b_1 = b_2 = 0.0833$, $a_3 = 0.24$ and $b_3 = 0.03$										
250	488	242	98	56	0.99	461	220	77	33	1.99
	(173	67	23	13	0.67	3330	2801	2193	1793	5.29)
1000	483	243	94	49	0.98	477	244	99	45	1.97
	(305	128	49	20	0.79	2716	2578	2499	2467	10.58)
(b) $a_1 = a_2 = b_1 = b_2 = 0.0833$, $a_3 = 0.3$ and $b_3 = 0.0375$										
250	487	221	83	41	0.99	483	233	84	37	1.99
	(365	142	41	19	0.88	1322	992	713	563	2.96)
1000	493	264	110	56	1.01	499	244	93	60	2.01
	(485	258	108	52	1.01	554	304	155	122	2.25)
(c) $a_1 = a_2 = 0.2$, $b_1 = b_2 = 0.025$ and $a_3 = b_3 = 0.1$										
250	507	254	113	62	1.00	529	259	110	52	2.03
	(353	182	66	34	0.86	1732	1414	1111	903	3.52)
1000	501	246	90	52	1.00	476	229	88	49	1.99
	(483	230	83	48	0.99	597	361	238	191	2.54)
(d) $a_1 = a_2 = 0.2$, $b_1 = b_2 = 0.025$ and $a_3 = b_3 = 0.125$										
250	490	235	88	43	0.99	476	218	81	38	1.99
	(468	224	79	37	0.97	660	393	231	170	2.20)
1000	489	227	96	47	0.98	493	259	99	50	1.97
	(489	227	96	47	0.98	493	259	99	50	1.97)

TABLE 3

Power of statistical test for a 5% significance level, where $d_1 = 0.05$, $d_3 = 0.3$ and $L = 1000$ were assumed

	Method		
d_2	1D ($\overline{\chi^2}$)	LR[a]	WL[a]
0.05	0.0502 (0.99)	0.060	0.059
0.10	0.8264 (9.20)	0.792	0.789
0.15	0.9988 (24.80)	0.997	0.997
0.20	1.0000 (42.58)	1.000	1.000
0.25	1.0000 (60.63)	1.000	1.000

Results obtained by using the 1DN method are the same as those of the 1D method, except in the case of $d_2 = 0.25$. In this case 1.0000 (60.38) was obtained by using the 1DN method.

[a] Results from MUSE and WEIR (1992).

and the relative rate test (WL) of WU and LI (1985). In these sets of parameter values, the powers of the 1DN and 2DN methods were essentially the same as those of the 1D and 2D methods, respectively. It should be noted that, in set 9, the overall rate is the same between sequences 1 and 2 but the transitional and transversional rates are different between them.

From these results we can conclude that the 1D and 2D methods are as powerful as those of the likelihood ratio test and the relative rate test, in spite of the fact that the 1D and 2D methods have no assumption about the pattern of substitution rates.

To investigate the difference between the 1D and 1DN methods and that between the 2D and 2DN method, as well as the effect of the choice of outgroup on the statistical power, I have conducted computer simulations, where the transitional rate was assumed to be two times as high as transversional rate and only the length (d_3) of the lineage leading to the outgroup was changed. The results are shown in Table 6. As expected, the powers of the 1D and 2D methods increase as d_3 decreases, since the outgroup is known under these methods. Therefore, we can conclude that when the outgroup is known, these methods are more powerful when the outgroup lineage is shorter. On the other hand, the 1DN and 2DN methods show different properties. When $d_3 = 0.08$ ($= d_1$), there is no statistical power. As d_3 decreases, the power increases. This is because, using these methods, we are actually testing the difference between sequences 1 and 3, not between sequences 1 and 2. As d_3 increases from 0.08, the power increases, reaches to its maximum value, and decreases. This is because, as d_3 increases, the probability of making a comparison between sequences 1 and 2 increases.

NUMERICAL EXAMPLE

As an example, I have analyzed hominoid mitochondrial DNA sequences obtained by HORAI *et al.* (1992), whose Figure 1 shows the alignment of six sequences from common chimpanzee (C), pygmy chimpanzee (P), human (H), gorilla (G), orangutan (O) and siamang (S). These sequences are about 4.9 kb in length, and contain the complete genes for NADH dehydrogenase subunit 2 (ND2), cytochrome

TABLE 4

Parameter values for computer simulations

Set	Transition			Transversion			Combined[a]		
	$4a_1$	$4a_2$	$4a_3$	$4b_1$	$4b_2$	$4b_3$	d_1	d_2	d_3
1	0.10	0.10	0.25	0.05	0.05	0.15	0.050	0.050	0.1375
2	0.125	0.10	0.25	0.0625	0.05	0.15	0.0625	0.050	0.1375
3	0.10	0.0875	0.25	0.05	0.0375	0.15	0.050	0.040625	0.1375
4	0.15	0.10	0.25	0.075	0.05	0.15	0.075	0.050	0.1375
5	0.05	0.03	0.10	0.02	0.01	0.05	0.0225	0.0125	0.050
6	0.10	0.10	0.25	0.10	0.05	0.15	0.075	0.050	0.1375
7	0.10	0.05	0.25	0.05	0.05	0.15	0.050	0.0375	0.1375
8	0.10	0.05	0.25	0.05	0.10	0.15	0.050	0.0625	0.1375
9	0.10	0.15	0.25	0.05	0.025	0.15	0.050	0.050	0.1375

[a] $d_i = a_i + 2b_i$.

TABLE 5

Power of statistical test for a 5% significance level, where parameter values in Table 4 and $L = 1000$ were used

Set	Method				
	2D	LR2P[a]	1D	LR1P[a]	WL[a]
1	0.0479	0.044	0.0507	0.040	0.041
2	0.1305	0.115	0.1722	0.158	0.155
3	0.1093	0.112	0.1327	0.129	0.127
4	0.3636	0.388	0.4626	0.483	0.472
5	0.2666	0.286	0.3555	0.375	0.361
6	0.5812	0.611	0.4802	0.497	0.458
7	0.2888	0.297	0.1925	0.214	0.215
8	0.7300	0.718	0.1787	0.165	0.156
9	0.4678	0.497	0.0515	0.049	0.049

Results obtained by using the 1DN method are the same as those of the 1D method, and results obtained by using the 2DN method are the same as those of the 2D method, except in set 5. In this set 0.2665 was obtained by using the 2DN method.

[a] Results from MUSE and WEIR (1992).

TABLE 6

Power of statistical test for a 5% significance level, where $d_1 = 0.08$ ($a_1 = 0.04$, $b_1 = 0.02$), $d_2 = 0.12$ ($a_2 = 0.06$, $b_2 = 0.03$) and $L = 1000$ were assumed

Parameter			Method			
a_3	b_3	d_3	2D	2DN	1D	1DN
0.01	0.005	0.02	0.6473	0.9989	0.7395	0.9997
0.02	0.01	0.04	0.6259	0.8064	0.7307	0.8860
0.03	0.015	0.06	0.6040	0.2275	0.7111	0.2940
0.04	0.02	0.08	0.5859	0.0482	0.6904	0.0491
0.05	0.025	0.10	0.5664	0.1643	0.6740	0.2163
0.06	0.03	0.12	0.5593	0.4026	0.6659	0.5109
0.08	0.04	0.16	0.5144	0.5129	0.6236	0.6219
0.10	0.05	0.20	0.4742	0.4742	0.5785	0.5785
0.20	0.10	0.40	0.3183	0.3183	0.4087	0.4087
0.40	0.20	0.80	0.1489	0.1489	0.1898	0.1898
0.80	0.40	1.60	0.0648	0.0648	0.0704	0.0704
1.60	0.80	3.20	0.0470	0.0470	0.0468	0.0468

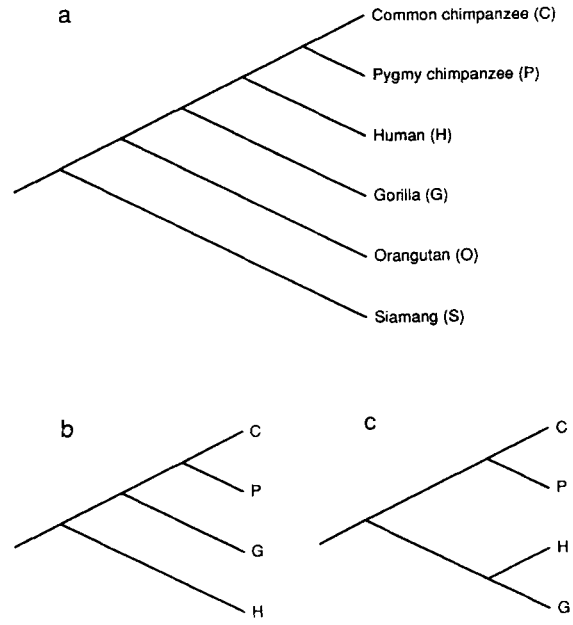

FIGURE 2.—Phylogenetic relationships among hominoid species.

oxidase subunits I and II (COI and COII), ATPase 8, parts of two genes for ND1 and ATPase 6, and 11 interspersed tRNAs. In order to apply the 1D and 2D methods, we must know the phylogenetic relationship among these sequences. I used the phylogenetic relationship shown in Figure 2a. The relationship among (common and pygmy) chimpanzee, human and gorilla is controversial [*e.g.*, see MARKS (1992)], although HORAI *et al.* (1992) strongly suggest that of Figure 2a. Therefore, I also used the phylogenetic relationship shown in Figures 2b and 2c for the relationship among these species.

The results of the tests are given in Table 7, where species 3 was assumed to be the outgroup. It should be noted here that the tests are not mutually independent. In the comparison among C, P and S, significantly large chi-square values were obtained ($\chi^2 = 6.537$ for the 1D and 1DN methods and $\chi^2 = 6.816$ for the 2D and 2DN methods). As shown in the

TABLE 7

Testing hominoid mitochondrial DNA sequences

Species[c]			No. of different sites			Chi-square	
1	2	3	m_1 (s_1, v_1)	m_2 (s_2, v_2)	m_3 (s_3, v_3)	1D	2D
C	P	H	84 (80, 4)	78 (74, 4)	324 (310, 14)	0.222	0.234
C	P	G	83 (80, 3)	72 (69, 3)	407 (358, 49)	0.781	0.812
C	P	O	74 (70, 4)	78 (75, 3)	651 (531, 120)	0.105	0.315
C	P	S	89 (86, 3)	58 (55, 3)	672 (513, 159)	6.537*	6.816*
H	C	G	219 (211, 8)	176 (168, 8)	307 (271, 36)	4.681*	4.879
C	G	H[a]	176 (168, 8)	307 (271, 36)	219 (211, 8)	35.530***	41.984***
G	H	C[a]	307 (271, 36)	219 (211, 8)	176 (168, 8)	14.722***	25.287***
H	P	G	222 (214, 8)	167 (159, 8)	305 (268, 37)	7.776**	8.110*
P	G	H[b]	167 (159, 8)	305 (268, 37)	222 (214, 8)	40.347***	46.513***
G	H	P[b]	305 (268, 37)	222 (214, 8)	167 (159, 8)	13.072***	24.739***
H	C	O	193 (187, 6)	187 (180, 7)	520 (419, 101)	0.095	0.210
H	P	O	189 (182, 7)	188 (181, 7)	526 (424, 102)	0.003	0.003
H	C	S	195 (188, 7)	181 (175, 6)	562 (421, 141)	0.521	0.542
H	P	S	211 (203, 8)	166 (160, 6)	553 (408, 145)	5.371*	5.379
G	C	O	240 (217, 23)	221 (206, 15)	477 (386, 91)	0.783	1.970
G	P	O	234 (211, 23)	220 (204, 16)	486 (393, 93)	0.432	1.374
G	H	O	257 (233, 24)	246 (233, 13)	458 (368, 90)	0.241	3.270
G	C	S	239 (214, 25)	215 (197, 18)	518 (391, 127)	1.269	1.843
G	P	S	250 (223, 27)	195 (179, 16)	509 (382, 127)	6.798**	7.630*
G	H	S	254 (226, 28)	244 (229, 15)	504 (376, 128)	0.201	3.95
O	C	S	356 (295, 61)	306 (269, 37)	398 (312, 86)	3.776	7.076*
O	P	S	376 (314, 62)	296 (261, 35)	383 (294, 89)	9.524**	12.401**
O	H	S	351 (297, 54)	315 (278, 37)	400 (310, 90)	1.946	3.804
O	G	S	349 (297, 52)	324 (278, 46)	397 (309, 88)	0.929	0.995

Data from HORAI *et al.* (1992). Chi-square values obtained by using the 1DN and 2DN methods are the same as those of the 1D and 2D methods, respectively, except in the cases of CGH, GHC, PGH and GHP.

[a] In the cases of CGH and GHC, chi-square values obtained by using the 1DN and 2DN methods are 4.681* and 4.879, respectively.

[b] In the cases of PGH and GHP, chi-square values obtained by using the 1DN and 2DN methods are 7.776** and 8.110*, respectively.

[c] Species 3 was used as the outgroup in the 1D and 2D methods. Species abbreviations are: C, common chimpanzee; P, pygmy chimpanzee; H, human; G, gorilla; O, orangutan; and S, siamang.

* Significant at 5% level; ** significant at 1% level; *** significant at 0.1% level.

previous section, the 1D and 2D methods are more powerful when the outgroup is more closely related. Since chi-square values are not significantly large when H, G and O were used as the outgroup, the molecular evolutionary clock hypothesis cannot be rejected. In the comparison among H, (C or P) and G, significantly large chi-square values were obtained although the values depend on the choice of outgroup. Since the outgroup is unknown, it might be better to apply the 1DN and 2DN methods. Then, we have significantly large chi-square values ($\chi^2 = 4.681$ in the comparison among H, C and G for the 1DN method, and $\chi^2 = 7.776$ and 8.110 in the comparison among H, P and G for the 1DN and 2DN methods), so that the molecular clock hypothesis can be rejected. In the comparison among (H or G), P and S, significantly large chi-square values were obtained. Since chi-square values are not significantly large when the outgroup is O, we cannot reject the molecular clock hypothesis. When the comparison was made among O, (C or P) and S, we have significantly large chi-square values ($\chi^2 = 9.524$ for the 1D and 1DN methods, and $\chi^2 = 7.076$ and 12.401 for the 2D and 2DN methods), so that we can reject the molecular clock hypothesis.

It should be noted that it might be very difficult to apply the likelihood ratio test and the relative rate test to the present data, since the substitution rate varies among different sites. (For example, the substitution rate in tRNAs is different from that in proteins, the rate varies among different proteins, the rate in stem regions of tRNA is different from that in loop regions, the rate varies among different loop regions, and so on.) On the other hand, there is no difficulty in applying the present methods.

DISCUSSION

In this report, simple methods for testing the molecular evolutionary clock hypothesis were developed. In these methods, we do not have to assume a particular pattern of substitution rates. Furthermore, these methods are applicable even when the substitution rate varies among different sites. The 1DN and 2DN methods can be used when the outgroup is unknown. Therefore, the present methods might be useful.

The number of different sites is used in the 1D and

1DN methods, whereas this number is classified into transitional and transversional ones in the 2D and 2DN methods. The number of different sites, however, can be classified into many groups. For example, as a test statistic we can use

$$\chi^2 = \sum_i \sum_{j \geq i} \sum_k \frac{(n_{ijk} - n_{jik})^2}{n_{ijk} + n_{jik}} \tag{8}$$

with 24 d.f., or

$$\chi^2 = \sum_i \sum_{j \neq i} \frac{(n_{ijj} - n_{jij})^2}{n_{ijj} + n_{jij}} \tag{9}$$

with 12 d.f. Since the 2D method is not always more powerful than the 1D method, it is not clear whether (8) and (9) are powerful. Although (9) may be useful in some cases, very long sequences must be necessary to conduct the statistical test. It is also possible, when the substitution rate of transition is substantially higher than that of transversion, that only the transitional differences are classified into four classes, *i.e.*,

$$\chi^2 = \frac{(n_{122} - n_{212})^2}{n_{122} + n_{212}} + \frac{(n_{211} - n_{121})^2}{n_{211} + n_{121}} + \frac{(n_{344} - n_{434})^2}{n_{344} + n_{434}} + \frac{(n_{433} - n_{343})^2}{n_{433} + n_{343}} + \frac{(v_1 + v_2)^2}{v_1 + v_2} \tag{10}$$

with 5 d.f., although it is not clear how powerful (10) is. For example, when we compare mitochondrial sequences from human and common chimpanzee by using gorilla as the outgroup, we have $n_{122} = 10$, $n_{212} = 18$, $n_{211} = 53$, $n_{121} = 43$, $n_{344} = 63$, $n_{434} = 32$, $n_{433} = 85$, $n_{343} = 75$ and $v_1 = v_2 = 8$. Then, we obtain $\chi^2 = 14.068$ and $P(\chi^2 \geq 11.070) = 0.05$ for 5 d.f., so that we reject the molecular evolutionary clock hypothesis at the 5% level. The power of (10) probably depends on the pattern of substitution rates, the length of sequence, the divergence time, and so on.

In this paper we consider only nucleotide sequences. It might be clear, however, that we can also apply the present methods to amino acid sequences.

I thank S. HORAI and A. G. CLARK for their suggestions and comments. This is contribution no. 1966 from the National Institute of Genetics, Mishima, Shizuoka-ken 411, Japan.

LITERATURE CITED

EASTEAL, S., 1990 The pattern of mammalian evolution and the relative rate of molecular evolution. Genetics **124:** 165–173.

GRAUR, D., W. A. HIDE and W. H. LI, 1991 Is the guinea-pig a rodent? Nature **351:** 649–652.

HORAI, S., Y. SATTA, K. HAYASAKA, R. KONDO, T. INOUE, T. ISHIDA, S. HAYASHI and N. TAKAHATA, 1992 Man's place in Hominoidea revealed by mitochondrial DNA genealogy. J. Mol. Evol. **35:** 32–43.

JIN, L., and M. NEI, 1990 Limitations of the evolutionary parsimony method of phylogenetic analysis. Mol. Biol. Evol. **7:** 82–102.

KIMURA, M., 1980 A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.

KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution.* Cambridge University Press, Cambridge.

LI, W. H., M. GOUY, P. M. SHARP, C. O'HUIGIN and Y. W. YANG, 1990 Molecular phylogeny of Rodentia, Lagomorpha, Primates, Artiodactyla, and Carnivora and molecular clocks. Proc. Natl. Acad. Sci. USA **87:** 6703–6707.

MARKS, J., 1992 Genetic relationships among the apes and humans. Curr. Opin. Genet. Dev. **2:** 883–889.

MUSE, S. V., and B. S. WEIR, 1992 Testing for equality of evolutionary rates. Genetics **132:** 269–276.

NOVACEK, M. J., 1992 Mammalian phylogeny: shaking the tree. Nature **356:** 121–125.

SARICH, V. M., and A. C. WILSON, 1967 Immunological time scale for hominid evolution. Science **158:** 1200–1203.

WU, C. I., and W. H. LI, 1985 Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA **82:** 1741–1745.

ZUCKERKANDL, E., and L. PAULING, 1965 Evolutionary divergence and convergence in proteins, pp. 97–166 in *Evolving Genes and Proteins*, edited by V. BRYSON and H. J. VOGEL. Academic Press, New York.

Communicating editor: A. G. CLARK

APPENDIX

Let L be the length of a nucleotide sequence or the number of nucleotides in each sequence. Then, L can be divided into m_1, m_2 and $L - m_1 - m_2$. Since the expectation of m_1 is equal to that of m_2, the expectations of m_1, m_2 and $L - m_1 - m_2$ can be given by $E(m_1) = E(m_2) = pL/2$ and $E(L - m_1 - m_2) = (1 - p)L$, where the unbiased estimate of p is $(m_1 + m_2)/L$. Therefore, $E(m_1)$ and $E(m_2)$ can be estimated by $(m_1 + m_2)/2$, and $E(L - m_1 - m_2)$ by $L - m_1 - m_2$. Following the definition of chi-square, we have

$$\chi^2 = \frac{\{m_1 - E(m_1)\}^2}{E(m_1)} + \frac{\{m_2 - E(m_2)\}^2}{E(m_2)} + \frac{\{(L - m_1 - m_2) - E(L - m_1 - m_2)\}^2}{E(L - m_1 - m_2)} = \frac{(m_1 - m_2)^2}{m_1 + m_2}.$$

There are three categories and we estimate p from the same data, so that the degree of freedom is one. Thus, we obtain (4).

L can be divided into s_1, s_2, v_1, v_2 and $L - s_1 - s_2 - v_1 - v_2$. Since $E(s_1) = E(s_2)$ and $E(v_1) = E(v_2)$, we have $E(s_1) = E(s_2) = qL/2$, $E(v_1) = E(v_2) = rL/2$ and $E(L - s_1 - s_2 - v_1 - v_2) = (1 - q - r)L$, where the unbiased estimates of q and r are $(s_1 + s_2)/L$ and $(v_1 + v_2)/L$, respectively. Therefore, $E(s_1)$ and $E(s_2)$ can be esti-

mated by $(s_1 + s_2)/2$, $E(v_1)$ and $E(v_2)$ by $(v_1 + v_2)/2$, and $E(L - s_1 - s_2 - v_1 - v_2)$ by $L - s_1 - s_2 - v_1 - v_2$. Following the definition of chi-square, we have

$$\chi^2 = \frac{\{s_1 - E(s_1)\}^2}{E(s_1)} + \frac{\{s_2 - E(s_2)\}^2}{E(s_2)}$$

$$+ \frac{\{v_1 - E(v_1)\}^2}{E(v_1)} + \frac{\{v_2 - E(v_2)\}^2}{E(v_2)}$$

$$+ \frac{\{(L - s_1 - s_2 - v_1 - v_2) - E(L - s_1 - s_2 - v_1 - v_2)\}^2}{E(L - s_1 - s_2 - v_1 - v_2)}$$

$$= \frac{(s_1 - s_2)^2}{s_1 + s_2} + \frac{(v_1 - v_2)^2}{v_1 + v_2}.$$

There are five categories and we estimate q and r from the same data, so that the degree of freedom is two. Thus, we obtain (6).

Formulas (8), (9) and (10) can be obtained in the same way as the above.