Slow Molecular Clocks in Old World Monkeys, Apes, and Humans

Soojin Yi, Darrell L. Ellsworth,† and Wen-Hsiung Li**

*Department of Ecology and Evolution, University of Chicago; and †Windber Research Institute, Windber, Pennsylvania

Two longstanding issues on the molecular clock hypothesis are studied in this article. First, is there a global molecular clock in mammals? Although many authors have observed unequal rates of nucleotide substitution among mammalian lineages, some authors have proposed a global clock for all eutherians, i.e., a single global rate of 2.2 $\times 10^{-9}$ substitutions per nucleotide site per year. We reexamine this issue using noncoding, nonrepetitive DNA from Old World monkeys (OWMs), chimpanzee, and human. First, using the minimal date of 6 MYA for the human-chimpanzee divergence and more than 2.5 million base pairs of genomic sequences from human and chimpanzee, we estimate a maximal rate of 0.99×10^{-9} for noncoding, nonrepetitive genomic regions for these two species. This estimate is less than half of the proposed global rate and much smaller than the commonly used rate (3.5×10^{-9}) for eutherians. Further, using a minimal date of 23 MYA for the human-OWM divergence, we estimate a maximal rate of 1.5×10^{-9} for both introns and fourfold degenerate sites in humans and OWMs. In addition, with the New World monkey (NWM) lineage as an outgroup, we estimate that the rate of substitution in introns is 30% higher in the OWM lineage than in the human lineage. Clearly, there is no global molecular clock in eutherians. Second, although many studies have indicated considerable variation in the mutation rate among regions of the mammalian genome, a recent study proposed a uniform rate. Using new and existing intron sequence data from higher primates, we find significant rate variation among genomic regions and a positive correlation between the rate of substitution and the GC content, refuting the claim of a uniform rate.

Introduction

The molecular clock hypothesis has been a central issue in molecular evolution since it was proposed in 1965 (Zuckerkandl and Pauling 1965). This issue deserves much attention for two reasons. First, if a molecular clock can be applied to all taxa, then estimating divergence times and reconstructing phylogenetic trees can be done by implementing relatively simple methods. Second, the regularity of the molecular clock has significant implications for understanding the mechanisms of molecular evolution. Under the neutral theory of molecular evolution, the rate of evolution is equal to the rate of mutation (Kimura 1968). Therefore, the degree of rate variation among lineages may provide insight into factors affecting the rate of evolution in different genomes (Zuckerkandl and Pauling 1965; Li 1997, pp. 215–235; Nei and Kumar 2000, pp. 187–191).

For similar reasons, it is also useful to determine the degree of rate variation among different regions of a genome. Although there has been general agreement that considerable variation in substitution rates exists among different regions of the mammalian genome (Filipski 1988; Wolfe, Sharp, and Li 1989; Bernardi, Mouchiroud, and Gautier 1993; Casane et al. 1997; Matassi, Sharp, and Gautier 1999; Lercher, Williams, and Hurst 2001; Ebersberger et al. 2002), some authors have proposed a uniform mutation rate along the mammalian genome (Kumar and Subramanian 2002).

There have been generally two drawbacks in previous studies on the molecular clock hypothesis. First, estimates of nucleotide substitution rates often involved uncertain assumptions about divergence dates between taxa. For example, Kumar and Subramanian (2002) relied mainly on divergence dates estimated under the assumption of a global protein clock in vertebrates and the assumption of 310 MYA for the mammal-reptile split to examine mutation rate differences within and among genomes of diverse mammalian lineages (e.g., Kumar and Hedges 1998). However, the assumption of a global protein clock in vertebrates is unlikely to be true because, for example, the rate of amino acid substitution has been shown to be much higher in the rodent lineage than in the primate lineage (Gu and Li 1992). Further, some of the fossil dates they used may be questionable; e.g., they assumed the same divergence date of 90 MYA for several pairs of mammalian orders. Second, the sequence data used in many studies were usually limited and often biased. For example, Kumar and Subramanian (2002) excluded all sequences that failed their disparity test; e.g., they excluded 41% of the available sequence data in the comparison between primates and rodents. In all their between-taxa comparisons, the excluded sequences have evolved, on average, considerably faster than the sequences retained in their analysis (see figure 2c of Kumar and Subramanian [2002]). This could be a major reason for their considerably lower estimated rates than previous estimates.

In this article, we address the above issues in three ways. First, we show that the molecular clock in Old World monkeys (OWMs), humans, and apes (i.e., catarrhines) runs at a rate much lower than rates estimated for other mammals. For this purpose we take advantage of the large amount of genomic sequence data from human and chimpanzee and a recent discovery of an early hominid fossil. As questions regarding the rate of neutral substitutions can be adequately addressed only using selectively neutral sequences, and as repetitive sequences tend to evolve at different rates than the surrounding genomic regions, we use only noncoding, nonrepetitive sequences. Then we use the dating of a recently discovered hominid fossil (Brunet et al. 2002; Vignaud et al. 2002) as a calibration point. This fossil allows us to

Key words: molecular clocks, mutation rate, hominoid rate-slowdown, Old World monkeys, humans, apes.

Address for correspondence and reprints: Wen-Hsiung Li, Department of Ecology and Evolution, University of Chicago, 1101 East 57th street, Chicago, Illinois 60637. E-mail: whli@uchicago.edu.

Mol. Biol. Evol. 19(12):2191–2198. 2002

estimate a minimal divergence time for these two lineages, which in turn can be used to infer a ''maximal'' mutation rate for these lineages. In a similar manner, we obtain a maximal rate for the human and OWM lineages and show that it is much lower than the rates estimated for other mammals.

Second, we revisit the hominoid rate-slowdown hypothesis. This was first proposed on the basis of immunological data (Goodman 1961, 1962) and further debated using DNA sequence data (Bailey et al. 1991; Easteal 1991; Herbert and Easteal 1996; Li et al. 1996). In this study we have obtained new intron sequence data and have retrieved existing intron sequence data and fourfold degenerate site data from higher primates. Using the relative rate test (Sarich and Wilson 1967; Wu and Li 1985) with a New World monkey (NWM) species as the outgroup, we show that the rate of nucleotide substitution is significantly lower in the hominoid lineage than in the OWM lineage. This and the above results together reject the proposal of a global molecular clock in mammals.

Third, using new and existing intron sequence data we show that the rate of nucleotide substitution is not uniform among genomic regions. This result corroborates previous observations (Filipski 1988; Wolfe, Sharp, and Li 1989; Casane et al. 1997; Matassi, Sharp, and Gautier 1999; Lercher, Williams, and Hurst 2001; Ebersberger et al. 2002). Further, using the same data set, we show that the rate heterogeneity is in part due to variation in GC content.

Materials and Methods

Human and Chimpanzee Genomic Sequence Data

A total of 24 complete chimpanzee contigs was obtained from GenBank (March 2002 freeze); these included all the chimpanzee contigs used in Chen et al. (2001). The accession numbers of each chimpanzee contig and the corresponding human contig are shown in table 1. A total of 2.73 million base pairs of genomic sequences were available for comparison between the two species. We used each chimpanzee contig to BLAST against the human genome sequences, with repeats intact or masked, to find the corresponding human contig and the parallel coordinates. Because each chimpanzee contig was 64–213 kb in length, the corresponding human contig (or joined contigs) was easily identified even with repeats intact. From the matched human and chimpanzee contigs, we extracted only the noncoding intergenic regions, based on information given in the GenBank summary files. We then separately aligned the homologous segments of the chimpanzee and human contigs, using a large-scale alignment program by A. C. Shih and W.-H. Li (unpublished data), which first identifies well-conserved regions to serve as alignment anchors and then uses Clustal W to align segments between anchors. A total of 875,353 bp of intergenic regions was aligned between the two species and only the nonrepetitive sequences (447,330 bp) identified by the Repeat Masker (courtesy of the Bioinformatics group, the Institute for Systems Biology, Seattle, WA and the University of Washington) were used in our analysis.

Primate Intron Sequences

Genomic DNA from OW (baboon, *Papio cynocephalus*) and NW (squirrel monkey, *Saimiri sciureus*) primates was isolated from homogenized liver tissue. Available nucleotide sequences from divergent taxa were aligned to identify conserved regions in exons flanking introns of interest to construct oligonucleotide primers. Subsequent PCR reactions amplified the following regions: intron 3 of APOC3, intron 5 of CA7, intron 5 of IFNAR1, intron 7 of LPL, and intron 6 of CPA3. Primer sequences, PCR conditions, and detailed sequencing protocols are available upon request.

In addition, 75 regions including exons and introns were retrieved from GenBank for at least one OW monkey species and at least one NW monkey species. From these data we excluded sequences from the X and Y chromosomes because the rate of molecular evolution on the sex chromosomes is also affected by other evolutionary factors (e.g., Bachtrog and Charlesworth 2002; Makova and Li 2002). After exclusion, 27 intron sequences from various genomic regions were available for alignment. RepeatMasker was used to exclude repeats before alignment. Manual adjustments were made, if necessary, to finalize each alignment.

Retrieval of Primate Fourfold Degenerate Sites

We extracted protein-coding exon sequences from 41 protein-coding regions that fit the above criteria in the sequences retrieved from GenBank. Amino acid sequences were then produced from the exon-only DNA sequences and aligned with those from other species using the Clustal option of the MegAlign (part of the DNA STAR package). Sequence lengths were trimmed to obtain a complete alignment set for the three species. Then we back translated amino acid sequences to produce DNA sequence alignments. We only used fourfold degenerate sites for further analyses.

Statistical Analysis

The number of nucleotide substitutions per site (K) between two sequences was estimated using Kimura's two-parameter method (Kimura 1980) and the Tamura-Nei method (Tamura and Nei 1993) available in the DAMBE program (Xia and Xie 2001). Because these methods provided essentially the same results for all sequence comparisons, only the results obtained by Kimura's two-parameter method are shown. The relative rate tests were performed using Wu and Li's method (Wu and Li 1985) only on introns longer than 100 bp.

The disparity index test (Kumar and Gadagkar 2001) was done for each set of orthologous sequences using MEGA2 (Kumar et al. 2001). We used a stringent criterion such that if the disparity index test failed in any of the three pairwise comparisons (Homo-OW monkey, OW-NW monkeys, Homo-NW monkeys), then we considered that the set of three sequences has failed the

Table 1
Number of Nucleotide Substitutions Per 100 Sites (*K*) Between Human and Chimpanzee Noncoding, Nonrepetitive Sequences

Homo Contig/ Chromosome#	*Pan* Contig	Aligned Intergenic Regions (bp)	Length (bp) Used, Excluding Repeats	*K*
NT_030004/7	AC087834	12,407	4,794	1.05
NT_030004/7	AC087834	14,147	3,460	0.99
NT_030004/7	AC087834	9,104	7,063	0.96
NT_030004/7	AC087835	44,494	21,895	1.24
NT_005479/2	AC097335	24,196	12,249	1.19
NT_005479/2	AC097335	29,529	17,824	1.69
NT_005479/2	AC097335	2,707	985	2.91
NT_007867/7	AC087778	12,517	5,774	1.35
NT_007867/7	AC087568	14,147	3,460	0.99
NT_007867/7	AC087568	12,039	4,242	0.90
NT_007867/7	AC091296	35,625	11,134	1.02
NT_007867/7	AC093135	9,343	4,261	1.66
NT_007867/7	AC093135	6,495	3,007	0.74
NT_007867/7	AC093135	10,013	3,606	0.73
NT_007867/7	AC091504	9,632	3,874	1.15
NT_007867/7	AC093709	10,030	3,664	1.07
NT_007867/7	AC093709	30,407	6,710	0.67
NT_007927/7	AC087512	30,807	15,405	2.33
NT_007927/7	AC087265	18,674	14,712	1.26
NT_007927/7	AC087265	19,247	10,149	1.26
NT_007927/7	AC087253	10,357	7,328	1.03
NT_007927/7	AC087253	54,167	38,141	0.98
NT_007927/7	AC087264	52,629	37,337	1.03
NT_007927/7	AC087729	21,328	10,646	1.30
NT_007927/7	AC087729	17,946	8,845	1.06
NT_009700/12	AC006582	23,506	7,995	1.11
NT_022222/2	AC097335	11,551	8,041	2.06
NT_024133/10	AC113436	2,692	1,870	1.57
NT_024133/10	AC113436	21,315	3,175	2.27
NT_028391/20	AC096630	39,890	21,310	1.41
NT_028391/20	AC096630	4,314	2,844	1.10
NT_028391/20	AC096630	1,658	128	2.38
NT_028391/20	AC096630	31,260	20,957	1.12
NT_029419/12	AC006582	27,206	14,956	1.11
NT_029419/12	AC006582	8,554	2,006	1.00
NT_029419/12	AC006582	26,522	17,935	1.24
NT_030004/7	AC087777	13,951	8,874	0.76
NT_030004/7	AC087777	5,571	4,006	1.06
NT_030004/7	AC087777	12,294	8,927	0.79
NT_030004/7	AC087730	59,712	23,352	0.89
NT_030005/7	AC092764	14,652	8,994	1.17
NT_030005/7	AC092764	36,183	17,252	1.16
NT_030005/7	AC092764	1,038	966	1.68
NT_030005/7	AC092764	22,487	9,603	1.24
NT_030005/7	AC092764	11,417	4,114	0.96
Total		875,353	447,330	1.19 ± 0.016

disparity index test. We also performed a *G*-test for goodness of fit (Sokal and Rohlf 1995, p. 738) to a uniform mutation rate model for the observed number of substitutions between human and NW monkeys. For each gene, we chose the longest intron sequence, which had to be longer than 250 bp after excluding repeats. A total of 23 introns were thus selected. The same data set was used to investigate the correlation between the GC content and the evolutionary rate between human and a NW monkey species.

Results and Discussion

Exceedingly Slow Molecular Clocks in Humans and Chimpanzees

We are interested in providing an estimate of the substitution rate between the human and chimpanzee lineages that can be confidently regarded as a maximal estimate. We approach this by estimating the sequence divergence using only noncoding, nonrepetitive sequences, and by using well-characterized fossil records.

First, we use the huge amount of sequence data from the human and chimpanzee genomes to obtain a reliable estimate of the mean sequence divergence between these two lineages. Beginning with $\sim 2.73 \times 10^6$ bp of homologous *Pan* and *Homo* genomic contigs available in GenBank, we extracted and aligned approximately 875 kb of noncoding intergenic regions between human and chimpanzee (table 1). This was after exclud-

ing all the predicted and known coding exons, introns, and regulatory regions. We also excluded repetitive sequences. The sequence divergences in aligned regions between human and chimpanzee are listed in table 1.

The degree of sequence divergence differs considerably among regions (0.67%–2.91%, table 1). This has been observed earlier in two different studies (Chen and Li 2001; Ebersberger et al. 2002). It may reflect statistical fluctuations and variation among genomic regions in the degree of preexisting polymorphism in the common ancestral population of humans and chimpanzees. The average divergence from these regions, excluding repetitive sequences, is 1.19% (table 1). This estimate is almost the same as previously reported (1.24%) based on 53 short segments of noncoding intergenic regions and many long contigs (Chen and Li 2001; Chen et al. 2001) and almost 2 million bases of randomly scattered sequences in the entire genome (Ebersberger et al. 2002).

Next we take advantage of the recent discovery of a hominid fossil, *Sahelanthropus tchadensis* (Brunet et al. 2002), which was dated to be 6–7 MYA before present (Vignaud et al. 2002). This is even older than the earliest known fossil hominid so far, the *Ardipithecus ramidus kadabba*, dated to be 5.2–5.8 MYA before present (Haile-Selassie 2001). As the *Sahelanthropus* fossil is hominid, it gives a minimal date of 6 MYA for the divergence between human and chimpanzee. (Note that as long as *Sahelanthropus* was a hominid, the fossil would have postdated the human-chimpanzee divergence, even if it was not a direct ancestor of *Homo sapiens*.) We therefore obtain a *maximal* rate of 0.99×10^{-9} substitutions per site per year for the human and chimpanzee lineages. The actual human-chimp divergence should be older than the *Sahelanthropus* hominid fossil, which can be as old as 7 MYA (Vignaud et al. 2002). If we assume that the date for the human-chimp split is 7.5 MYA, then the average substitution rate in noncoding, nonrepetitive regions becomes 0.79×10^{-9} substitutions per site per year. This result may still overestimate the substitution rate because it did not exclude the effect of preexisting polymorphism in the common ancestor of human and chimpanzee. We understand that fossil dates may not be reliable, but even if we assume 5 MYA for the human-chimpanzee divergence, which is a minimum date commonly used in the literature, we still obtain a rate of only 1.19×10^{-9} substitutions per site per year.

The above values are less than half the ''global rate'' proposed by Kumar and Subramanian (2002) and much lower than estimates obtained from other mammalian lineages. For example, the maximum-likelihood estimate of the numbers of synonymous substitutions per site along the Artiodactyl lineage is 0.327 (Bielawski, Dunn, and Yang 2000), which leads to a rate of $\sim 3.5 \times 10^{-9}$ substitutions per site per year, using the divergence time estimates of 90–95 MYA by Kumar and Hedges (1998). The average sequence divergence between mouse and rat estimated from fourfold degenerate sites is ~18% (Smith and Hurst 1999). The divergence time between mouse and rat is controversial, ranging from 12 MYA based on the fossil record (Flynn, Jacobs, and Lindsay 1985) to over 40 MYA based on a global vertebrate molecular clock (Kumar and Hedges 1998). Even if we use the range of 20 to 40 MYA, the substitution rate range is $2.5–5.0 \times 10^{-9}$. Therefore, the substitution rate in humans and chimpanzees is exceedingly low, contradictory to the proposal of a global clock in eutherians (Kumar and Subramanian 2002).

Slow Molecular Clock in Old World Primates

We also estimate the evolutionary rates in the human and OWM lineages for introns and fourfold degenerate sites. Using new and existing sequence data, we estimate a divergence of 6.85% for introns (table 2). We also divide the introns used into those that failed the disparity test and those that passed the test. Note that, as expected, the average divergence (7.05%) for the first group of introns is substantially higher than that (6.77%) for the second group of introns (table 2). In the following calculations we shall include all introns because we are interested in obtaining a maximal estimate of the substitution rate. We also obtain a divergence of 7.17% for fourfold degenerate sites (data not shown) between the *Homo* and OWM lineages. The simple average of the estimates from introns and fourfold degenerate sites is 7.01%. As the earliest fossil record of true hominoids is represented by the genus *Proconsul*, which lived in Africa 23–27 MYA during the Early Miocene (Delson 2000, pp. 595–597), the human-OWM divergence would have predated this fossil. If we assume 23 MYA, a commonly used fossil calibration point for the human-cercopithecoid divergence (Goodman et al. 1998), the average rate of nucleotide substitution for the two lineages is $\sim 1.5 \times 10^{-9}$. If we assume a divergence date of 30 MYA, the average rate becomes only 1.17×10^{-9}. Clearly, these estimates are much lower than the values cited above for other mammals. Note that our estimated rates for humans, apes, and OW monkeys are inflated to some extent because we did not exclude those sequences that failed the disparity test as was done by Kumar and Subramanian (2002). Therefore, our results strongly contradict the proposal of a global clock in mammals.

Rate Differences between the Human and OWM Lineages

To reexamine the issue of rate differences between the *Homo* and OWM lineages, we used new and existing intron sequence data to perform relative rate tests, with a NWM species as the outgroup. Because the pattern of evolution in the repetitive sequences differs from that of nonrepetitive sequences, we excluded all repetitive sequences from our analyses. Overall, we compared 15,304 bp of intron sequences (table 2).

In the majority of cases, the rate of nucleotide substitution is faster in the OWM lineage (table 2). When all the sites are combined the rate of nucleotide substitution in the OWM lineage is significantly faster ($P < 0.001$). The average ratio of the substitution rate in the OWM lineage to that in the *Homo* lineage is 1.33, which

Table 2
Differences in the Number of Nucleotide Substitutions Per 100 Sites and the Relative Substitution Rates Between The *Homo* (Species 1) and the Old World Monkey (Species 2) Lineages with a New World Monkey (Species 3) as a Reference

Intron/ Chromosome#	Nucleotides Compared	K_{12}	K_{13}	K_{23}	$K_{13} - K_{23}$ (σ)	Ratio
Introns that passed the disparity index test						
APOC3/11	815	8.65	16.88	18.48	−1.60 (1.00)	1.45
CA7/16	501	7.23	9.67	11.14	−1.46 (1.36)	1.51
CPA3/3	1,285	5.53	12.64	13.15	−0.51 (0.77)	1.20
DAFint/1	504	6.30	9.24	9.66	−0.42 (1.24)	1.14
HBE1int1/11	121	3.41	3.38	6.94	−3.57 (1.80)	—
HBE1int2/11	812	4.98	12.39	12.40	−0.01 (0.91)	1.00
HBG1int1/11	122	5.16	10.78	12.75	−1.97 (2.50)	2.24
HBG1int2/11	804	7.39	12.89	15.30	−2.41 (1.10)	1.97
HBG2int1/11	122	6.05	10.72	9.66	1.07 (2.52)	0.70
HBG2int2/11	816	7.52	14.30	16.73	−2.43 (1.14)	1.95
IL3int3/5	122	1.62	6.05	7.94	−1.89 (1.26)	—
IL3int4/5	101	4.13	7.41	11.96	−4.55 (2.33)	—
INSint1/11	161	5.85	11.63	12.44	−0.81 (2.22)	1.32
INSint2/11	699	8.10	16.09	18.28	−2.19 (1.32)	1.74
LPL/8	1,168	7.86	13.77	13.55	0.22 (0.95)	0.95
OPN1SWint3/7	282	6.44	12.38	14.69	−2.31 (1.80)	2.12
OPN1SWint4/7	851	5.85	11.08	12.74	−1.67 (0.95)	1.80
PRM2/16	159	8.62	6.58	5.83	0.75 (2.43)	0.84
TUBB4Qint3/4	349	13.09	18.19	19.56	−1.37 (2.42)	1.23
Subtotal	9,794	6.77	12.67	13.86	−1.10 (0.30)***	1.36
Introns that failed the disparity index test						
FYint/1	394	6.18	11.59	14.48	−2.89 (1.47)	6.00
GHRint/5	116	7.27	9.19	11.09	−1.90 (2.76)	1.71
IFNAR1/21	885	7.59	14.00	14.03	−0.03 (1.67)	1.01
IGF2/11	1,589	6.41	14.19	15.83	−1.64 (0.76)	1.69
IL3int2/5	384	4.54	12.63	14.43	−1.80 (1.25)	2.31
IRBP/10	612	7.09	16.35	15.48	0.87 (1.32)	0.78
OPN1SWint1/7	279	5.62	9.19	9.19	0.00 (1.56)	1.00
OPN1SWint2/7	318	5.64	5.96	6.64	−0.68 (1.42)	1.27
TUBB4Qint1/4	219	13.11	22.57	23.82	−1.25 (3.16)	1.21
VWF/12	714	9.30	13.60	11.77	1.83 (1.30)	0.67
Subtotal	5,510	7.05	13.46	14.01	−0.50 (0.40)	1.17
All introns	15,304	6.85	12.98	13.95	−0.98 (0.25)***	1.33

NOTE.—K_{ij} = the number of substitutions per 100 nucleotide sites between species i and j.
*** Significant at the 0.001 level.

implies a 33% faster rate in the OWM lineage. When we restrict our analyses to the introns that passed the disparity index test, we reach the same conclusion (table 2). In fact, the rate difference between the *Homo* and OWM lineages is more pronounced in the introns that passed the disparity index test (table 2).

The above result supports the hominoid rate-slowdown hypothesis, which postulates that the rate of molecular evolution has become slower in hominoids (humans and apes) after their separation from the OW monkeys (Goodman 1961, 1962). The same trend has been observed earlier (Bailey et al. 1991; Seino, Bell, and Li 1992; Li et al. 1996) but has been challenged by Easteal (1991) and Herbert and Easteal (1996). Although Herbert and Easteal (1996) did find a ~30% rate increase in the OWM lineage in the noncoding regions they studied, they doubted the universality of this trend because most of the data (83%) were from the β-globin gene region (including the Ψη-globin gene). In our analysis, many different genomic regions were analyzed and the Ψη-globin region was not included (table 2). Moreover, we obtained the same result ($K_{13} - K_{23} = 0.9\% \pm 0.27\%$) when we excluded all the sequences from the β-globin gene region. This suggests the generality of our observation. Kumar and Subramanian (2002) reported a 10.9% increase in the evolutionary rate in the OWM lineage compared with the human lineage. However, their result was obtained from a much smaller data set, less than 4,000 fourfold degenerate sites from 23 genes (Kumar and Subramanian 2002). At any rate, the average rate difference from their and our estimates can be somewhere between 20% and 30%.

The basis of the hominoid rate-slowdown hypothesis lies in the notion that the molecular clock should run faster in organisms with a short generation time, for they go through many more generations per unit time than do organisms with a long generation (Ohta 1993; Li et al. 1996). This entails that the replication-dependent mutational errors such as errors in DNA replication in germ cells are the major source of mutations that are responsible for lineage effects. This view is strengthened by other observations. For example, the male-to-female

ratios of germ cell division in mice and primates roughly correspond to the estimated male-to-female ratios of the mutation rate (Li et al. 1996). Also, when compared with autosomes, the X-linked sequences usually show the lowest sequence divergence between species (Smith and Hurst 1999; Lercher, Williams, and Hurst 2001; Castresana 2002; Ebersberger et al. 2002). The generation-time effect hypothesis is also supported by the observation that the rate of substitution in the chloroplast genome is more than five times higher in grasses than in palms (Gaut et al. 1992). Nevertheless, the large variation in substitution rate among autosomal regions suggests that replication-independent factors may also be important sources of mutation (Lercher, Williams, and Hurst 2001; Castresana 2002; Ebersberger et al. 2002).

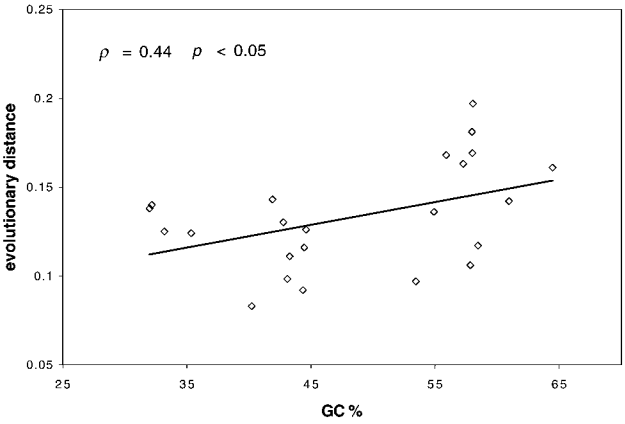

FIG. 1.—Relationship between the nucleotide substitution rate and the GC content of intron sequences.

Rate Variation Among Genomic Regions

Next let us examine the proposal of a uniform mutation rate along the mammalian genome. For this purpose, we use the human and NW monkey lineages because they show a moderate degree of sequence divergence (~14%, table 2) and the sequences can be reliably aligned. We choose introns longer than 250 bp after exclusion of repeats. This is to minimize selective constraints due to splicing signals and other functional motifs. Furthermore, only one intron (the longest one) from a given gene is examined; the purpose is to avoid the correlation of data from the same region, which is an important consideration when we compute the correlation between substitution rate and GC content. In total we include 23 introns from 17 chromosomes, with varying GC contents. We test a simple model of independence between genomic regions by applying the *G*-test statistic of heterogeneity to the counts of substituted and conserved sites in different introns (Sokal and Rohlf 1995, p. 738). The test rejects the model ($G^2 = 63.6$, df $= 22$, $P < 0.001$), implying that the mutation rate differs significantly among genomic regions.

We now determine whether the substitution rate in an intron is correlated with its GC content. Because we have no knowledge of the ancestral GC contents, we take the average of the GC contents of the introns from both species. We find a significant correlation ($\rho = 0.44$; fig. 1). When we use only human or only NWM species GC content, the obtained correlation coefficient varies only slightly ($\rho = 0.42$ and 0.45, respectively), as expected from the relatively small divergence between the two taxa.

Variation of substitution rate among genomic regions has been observed previously (Filipski 1988; Wolfe, Sharp, and Li 1989; Casane et al. 1997; Matassi, Sharp, and Gautier 1999; Lercher, Williams, and Hurst 2001; Castresana 2002; Ebersberger et al. 2002), but the factors responsible for rate variation have not been well understood. One proposed factor is variation in GC content. However, previous studies of the relationship between the GC content and the synonymous substitution rate in the mammalian genome have produced contradictory conclusions (Filipski 1988; Wolfe, Sharp, and Li 1989; Bernardi, Mouchiroud, and Gautier 1993; Matassi, Sharp, and Gautier 1999; Bielawski, Dunn, and Yang 2000; Castresana 2002). Recently, Bielawski, Dunn, and Yang (2000) reported a positive correlation between the synonymous substitution rate and the GC content at the third codon positions of a sequence when the sequence data were analyzed using a maximum-likelihood model that takes account of the transitional bias and unequal codon usage. Our finding of a significant correlation between GC content and substitution rate in introns corroborates this conclusion. However, the correlation found is relatively weak (~40%), which may explain why some studies found no correlation. At any rate, our study rejects the proposal of a uniform mutation rate along the mammalian genome.

Concluding Remarks

Our analyses of the rates of nucleotide substitution in noncoding, nonrepetitive sequences in Old World primates show that the rates of nucleotide substitution in these lineages are exceedingly slow when compared with those from other taxa. Therefore, application of a molecular clock to estimate divergence dates should be exercised with great caution even in relatively closely related taxa. That is, a molecular clock calibrated for some lineages may not be applicable to other lineages because the assumption of rate constancy among lineages may not hold, as shown in the case of higher primates. Furthermore, the rate estimated from one genomic region may not be applicable to another region because the mutation rate varies among genomic regions. This difference is apparently due in part to variation in sequence contexts.

Acknowledgments

We thank Robert Martin and Richard Kay for suggestions on the fossil record. This work was supported by NIH grants GM55759, HD38287, and GM30998. We thank two anonymous reviewers for helpful suggestions.

LITERATURE CITED

BACHTROG, D., and B. CHARLESWORTH. 2002. Reduced adaptation of a non-recombining neo-Y chromosome. Nature **416**:323–326.

BAILEY, W. J., D. H. A. FITCH, D. A. TAGLE, J. CZELUSNIAK, J. L. SLIGHTOM, and M. GOODMAN. 1991. Molecular evolution of the ψη-globin gene locus: Gibbon phylogeny and the hominoid slowdown. Mol. Biol. Evol. **8**:155–184.

BERNARDI, G., D. MOUCHIROUD, and C. GAUTIER. 1993. Silent substitutions in mammalian genomes and their evolutionary implications. J. Mol. Evol. **37**:583–589.

BIELAWSKI, J. P., K. A. DUNN, and Z. YANG. 2000. Rates of nucleotide substitution and mammalian nuclear gene evolution: approximate and maximum-likelihood methods lead to different conclusions. Genetics **156**:1299–1308.

BRUNET, M., F. GUY, D. PILBEAM ET AL. (38 co-authors). 2002. A new hominid from the upper Miocene of Chad, Central Africa. Nature **418**:145–151.

CASANE, D., S. BOISSINOT, B. H.-J. CHANG, L. C. SHIMMIN, and W.-H. LI. 1997. Mutation pattern variation among regions of the primate genome. J. Mol. Evol. **45**:216–226.

CASTRESANA, J. 2002. Genes on human chromosome 19 show extreme divergence from the mouse orthologs and a high GC content. Nucleic Acids Res. **8**:1751–1756.

CHEN, F.-C., and W.-H. LI. 2001. Genomic divergence between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am. J. Hum. Genet. **68**:444–456.

CHEN, F.-C., E. J. VALLENDER, H. WANG, C. S. TZENG, and W.-H. LI. 2001. Genomic divergence between human and chimpanzee estimated from large-scale alignments of genomic sequences. J. Hered. **92**:481–489.

DELSON, E. 2000. Encyclopedia of human evolution and prehistory. Garland Publishing Co., New York.

EASTEAL, S. 1991. The relative rate of cDNA evolution in primates. Mol. Biol. Evol. **8**:115–127.

EBERSBERGER, I., D. METZLER, C. SCHAWARTS, and S. PÄÄBO. 2002. Genomewide comparison of DNA sequences between humans and chimpanzees. Am. J. Hum. Genet. **70**:1490–1497.

FILIPSKI, J. 1988. Why the rate of silent codon substitutions is variable within a vertebrate genome. J. Theor. Biol. **34**:159–164.

FLYNN, L. J., L. L. JACOBS, and E. H. LINDSAY. 1985. Problems in muroid phylogeny: relationships to other rodents and origin of major groups. Pp. 589–616 *in* W. P. LUCKETT and J.-L. HARTENBERGER, eds. Evolutionary relationships among rodents. Plenum, New York.

GAUT, B. S., S. V. MUSE, W. D. CLARK, and M. T. CLEGG. 1992. Relative rates of nucleotide substitution at the *rbc*L locus of monocotyledonous plants. J. Mol. Evol. **35**:292–303.

GOODMAN, M. 1961. The role of immunochemical differences in the phyletic development of human behavior. Hum. Biol. **33**:131–162.

———. 1962. Evolution of the immunologic species specificity of human serum proteins. Hum. Biol. **34**:104–150.

GOODMAN, M. C., C. A. PORTER, J. CZELUSNIAK, S. L. PAGE, H. SCHNEIDER, J. SHOSHANI, G. GUNNELL, and C. P. GROVES. 1998. Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. Mol. Phyl. Evol. **9**:585–598.

GU, X., and W.-H. LI. 1992. Higher rates of amino acid substitution in rodents than in humans. Mol. Phylogenet. Evol. **1**:211–214.

HAILE-SELASSIE, Y. 2001. Late miocene hominids from the middle Awash, Ethiopia. Nature **412**:178–181.

HERBERT, G., and S. EASTEAL. 1996. Relative rates of nuclear DNA evolution in human and old world monkey lineages. Mol. Biol. Evol. **13**:1054–1057.

KIMURA, M. 1968. Evolutionary rate at the molecular level. Nature **217**:624–626.

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

KUMAR, S., and B. HEDGES. 1998. A molecular timescale for vertebrate evolution. Nature **392**:917–920.

KUMAR, S., and S. R. GADAGKAR. 2001. Disparity index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. Genetics **158**:1321–1327.

KUMAR, S., and S. SUBRAMANIAN. 2002. Mutation rates in mammalian genomes. Proc. Natl. Acad. Sci. USA **99**:803–808.

KUMAR, S., K. TAMURA, I. B. JAKOBSEN, and M. NEI. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17**:1244–1245.

LERCHER, M. J., E. J. WILLIAMS, and L. D. HURST. 2001. Local similarity in evolutionary rates extends over whole chromosomes in human-rodent and mouse-rat comparisons: implications for understanding the mechanistic basis of the male mutation bias. Mol. Biol. Evol. **18**:2032–2039.

LI, W.-H. 1997. Molecular evolution. Sinauer Associates, Sunderland, Mass.

LI, W.-H., D. L. ELLSWORTH, J. KRUSHKAL, B. H.-J. CHANG, and D. HEWETT-EMMETT. 1996. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. Mol. Phyl. Evol. **5**:182–187.

MAKOVA, K. D., and W.-H. LI. 2002. Strong male-driven evolution of DNA sequences in humans and apes. Nature **416**:624–626.

MATASSI, G., P. M. SHARP, and C. GAUTIER. 1999. Chromosomal location effects on gene sequences evolution in mammals. Curr. Biol. **9**:786–791.

NEI, M., and S. KUMAR. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York.

OHTA, T. 1993. An examination of the generation-time effect on molecular evolution. Proc. Natl. Acad. Sci. USA **90**:10676–10680.

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

SEINO, S., G. I. BELL, and W.-H. LI. 1992. Sequences of primate insulin genes support the hypothesis of a slower rate of molecular evolution in humans and apes than in monkeys. Mol. Biol. Evol. **9**:193–203.

SMITH, N. G. C., and L. D. HURST. 1999. The causes of synonymous rate variation in the rodent genome: can substitution rates be used to estimate the sex bias in mutation rate? Genetics **152**:661–673.

SOKAL, R. R., and F. J. ROHLF. 1995. Biometry: the principles and practice of statistics in biological research. 3rd edition. W. H. Freeman and Company, New York.

TAMURA, K., and M. NEI. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. **10**:512–526.

VIGNAUD, P., P. DURINGER, H. T. MACKAYE ET AL. (21 co-authors). 2002. Geology and palaeontology of the upper miocene Toros-Menalla hominid locality, Chad. Nature **418**:152–155.

WOLFE, K. H., P. M. SHARP, and W.-H. LI. 1989. Mutation rates differ among regions of the mammalian genome. Nature **337**:283–285.

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.

XIA, X., and Z. XIE. 2001. DAMBE: data analysis in molecular biology and evolution. J. Hered. **92**:371–373.

ZUCKERKANDL, E., and L. PAULING. 1965. Evolutionary divergence and convergence in proteins. Pp. 357–417 *in* V. BRYSON and H. J. VOGEL. eds. Evolving genes and proteins. Academic Press, New York.

DAVID IRWIN, reviewing editor

Accepted August 9, 2002