Estimate of the Mutation Rate per Nucleotide in Humans

Michael W. Nachman and Susan L. Crowell

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721

Manuscript received July 24, 1999
Accepted for publication May 19, 2000

ABSTRACT

Many previous estimates of the mutation rate in humans have relied on screens of visible mutants. We investigated the rate and pattern of mutations at the nucleotide level by comparing pseudogenes in humans and chimpanzees to (i) provide an estimate of the average mutation rate per nucleotide, (ii) assess heterogeneity of mutation rate at different sites and for different types of mutations, (iii) test the hypothesis that the X chromosome has a lower mutation rate than autosomes, and (iv) estimate the deleterious mutation rate. Eighteen processed pseudogenes were sequenced, including 12 on autosomes and 6 on the X chromosome. The average mutation rate was estimated to be $\sim 2.5 \times 10^{-8}$ mutations per nucleotide site or 175 mutations per diploid genome per generation. Rates of mutation for both transitions and transversions at CpG dinucleotides are one order of magnitude higher than mutation rates at other sites. Single nucleotide substitutions are 10 times more frequent than length mutations. Comparison of rates of evolution for X-linked and autosomal pseudogenes suggests that the male mutation rate is 4 times the female mutation rate, but provides no evidence for a reduction in mutation rate that is specific to the X chromosome. Using conservative calculations of the proportion of the genome subject to purifying selection, we estimate that the genomic deleterious mutation rate (U) is at least 3. This high rate is difficult to reconcile with multiplicative fitness effects of individual mutations and suggests that synergistic epistasis among harmful mutations may be common.

MUTATION is the ultimate source of genetic variation; it is both the substrate for evolution and the cause of genetic disease. Most previous estimates of the human mutation rate have utilized one of three approaches. Two of these approaches rely on phenotypic differences associated with diseases and the third approach relies on direct comparison of DNA sequences without function. These phenotypic and molecular methods are fundamentally different and rely on different assumptions. The first approach, pioneered by Haldane (1932, 1935), assumes diseases are in mutation-selection balance. For recessive mutations, the equilibrium frequency of mutant alleles is $\sqrt{\mu/s}$, where μ is the mutation rate and s is the selective effect of the deleterious mutation (Haldane 1927). The mutation rate can be estimated if the frequency of mutant alleles and the strength of selection are known. Using this method, Haldane (1932) calculated that the per locus rate of mutation for hemophilia in humans is $\sim 10^{-5}$ per generation. The second approach involves counts of affected individuals born to unaffected parents for dominant disorders (Cooper and Krawczak 1993). This method has been used for many dominant diseases in humans, and rates per locus vary from 10^{-6}–10^{-4} per generation (summarized in Vogel and Motulsky 1997). Methods that rely on screens of disease phenotypes may provide an underestimate of the actual mutation rate because not all nucleotide changes in a disease gene will result in disease. On the other hand, some disease genes are likely to have been identified and studied in part because they have high mutation rates. The third approach to measuring the human mutation rate takes advantage of the well-known result that for neutral mutations, the mutation rate is equal to the rate of evolution (Kimura 1968). A direct comparison of stretches of DNA without function can provide an estimate of the mutation rate per generation between species whose divergence time and generation length are known. Comparisons of pseudogenes and of synonymous sites between humans and chimpanzees have suggested mutation rates on the order of 10^{-8} per site per generation (*e.g.*, Kondrashov and Crow 1993; Drake *et al.* 1998). Since many genes may contain $\sim 10^3$ nonsynonymous sites, this estimate is in reasonable agreement with per locus rates of 10^{-5} (Vogel and Motulsky 1997).

Here we are interested in extending this work to obtain a more precise estimate of the rate and pattern of mutation at the nucleotide level in humans. We have sequenced 18 processed pseudogenes in humans and chimpanzees, including 12 on autosomes and 6 on the X chromosome. First, we provide an estimate of the average underlying mutation rate per nucleotide site. Second, we compare mutation rates for different sites and for different classes of mutation to evaluate hetero-

Corresponding author: Michael W. Nachman, Department of Ecology and Evolutionary Biology, Biosciences West Bldg., University of Arizona, Tucson, AZ 85721. E-mail: nachman@u.arizona.edu

geneity of mutation rate. Third, we compare rates of divergence on the X chromosome and on autosomes to evaluate the hypothesis that the X chromosome has a lower mutation rate than the autosomes (McVean and Hurst 1997). Finally, we provide an approximation of the genomic deleterious mutation rate by considering the total mutation rate and the fraction of the genome that is subject to constraint (Kondrashov and Crow 1993).

MATERIALS AND METHODS

Samples: For each locus, two humans and one common chimpanzee were surveyed. Human genomic DNAs were provided by Dr. M. F. Hammer from the Y chromosome consortium DNA repository and represent one African male and one Caucasian male. Male chimpanzee (*Pan troglodytes*) DNA was provided by Dr. O. A. Ryder.

PCR amplification and DNA sequencing: Eighteen processed pseudogenes were PCR amplified (Saiki *et al.* 1988) directly from genomic DNA. The names and chromosomal locations for each pseudogene are given in Tables 1 and 2. PCR conditions were optimized for each locus. Typically, PCR was performed in 25-μl volumes with 40 cycles of 94° 1 min, 55° 1 min, and 72° 2 min. PCR and sequencing primers were designed from published human sequences with the following accession numbers: gamma-cytoplasmic actin (Actgp3), D50-657; α-enolase, X15277; connexin 43, M65189; cytochrome b, AC002087; C4-sterol methyl oxidase (Desp4), U93261; elongation factor 1-α (Ef1 alpha), AC002086; Ferritin, U46066; α-1, 3-galactosyltransferase (Hgt-2), M60263; interferon-induced 56-kD protein (II56), Z74739; lanosterol 14-αdemethylase (Cyp51), U40053; malate dehydrogenase, Z93019; NADH dehydrogenase, Z81369; proliferation-associated gene (Pag), X72297; GPI-anchor synthesis gene (PIGF), D49727; regulatory subunit RI α of cAMP-dependent protein kinase (RI alpha), X73110; adaptor protein (Shc), Y09846; Translin, AC002075; HTLV-1 enhancer-binding protein (Txreb), U03712. For each locus, at least one amplification primer was designed to lie outside of the pseudogene. By utilizing processed pseudogenes, we were able to generate a sequence across the site of integration for each locus in both chimpanzees and humans. This allowed us to confirm that we were comparing orthologous pseudogenes that had integrated prior to the human-chimpanzee divergence. Patterns of substitution further confirmed that all loci were pseudogenes: nonsynonymous substitutions outnumbered synonymous substitutions and many genes had frameshift mutations. Products were cycle-sequenced on both strands and run on an ABI 377 automated sequencer. No differences were detected between strands, suggesting that the error rate due to DNA sequencing is likely to be low. The amount of sequence generated for each locus is given in Tables 1 and 2; the average sequence length for each autosomal locus was 902 bp and the average sequence length for each X-linked locus was 877 bp. A total of 16,089 bp was sequenced in each individual. Sequences have been submitted to GenBank under accession nos. AF-196978–AF197019.

Data analysis: Chromatograms were scored by hand and sequences were aligned manually. Heterozygous sites were confirmed on both strands. Divergence between human and chimpanzee was calculated as the average pairwise difference between the two chimpanzee alleles and the four human alleles using Kimura's two-parameter model (Kimura 1980). Standard errors were also calculated from this model. Rates of

TABLE 1

Rates of evolution for autosomal processed pseudogenes

Locus	Chromosome	Length (bp)	GC content	Non-CpG K_{ts}	Non-CpG K_{tv}	CpG K_{ts}	CpG K_{tv}	K_h	K_i	K_t
Cytoplasmic actin	20	955	0.528	0.0111	0.0066	0.1701	0.0000	0.0245	0.0010	0.0256
α-Enolase	1	1034	0.505	0.0051	0.0010	0.1116	0.0504	0.0119	0.0000	0.0119
Connexin	5	768	0.474	0.0067	0.0040	0.1019	0.0620	0.0148	0.0013	0.0161
Cytochrome b	5	718	0.421	0.0029	0.0000	0.0698	0.0000	0.0042	0.0000	0.0042
Hgt-2	12	1048	0.456	0.0025	0.0030	0.0517	0.0000	0.0072	0.0019	0.0091
Interferon protein	13	713	0.453	0.0075	0.0028	0.1203	0.0000	0.0117	0.0000	0.0117
Lanosterol	13	873	0.462	0.0039	0.0024	0.0000	0.0000	0.0060	0.0000	0.0060
NADH dehydrogenase	22	1151	0.464	0.0056	0.0018	0.1134	0.0000	0.0107	0.0000	0.0107
Pag	9	883	0.470	0.0093	0.0070	0.1203	0.0000	0.0183	0.0034	0.0218
Pigf	5	865	0.384	0.0112	0.0041	0.1585	0.0751	0.0187	0.0023	0.0211
R1 α	1	872	0.426	0.0082	0.0015	0.0883	0.0883	0.0118	0.0034	0.0154
Translin	7	944	0.443	0.0022	0.0008	0.0000	0.0613	0.0050	0.0011	0.0061
Mean autosomal values ± standard errors		902	0.457	0.0063	0.0029	0.0875	0.0232	0.0121	0.0012	0.0133
		39	0.011	0.0008	0.0005	0.0173	0.0086	0.0011	0.0003	0.0011

Divergence per site between human and chimpanzee (Kimura 1980) is given for transitions (K_{ts}) and transversions (K_{tv}) at CpG sites and at non-CpG sites, all single nucleotide changes (K_h), insertion-deletions (K_i), and total changes including both point mutations and insertion-deletions (K_t). Mean values are weighted by the number of sites in each locus.

TABLE 2

Rates of evolution for X-linked processed pseudogenes

Locus	Chromosome	Length (bp)	GC content	Non-CpG K_{ts}	Non-CpG K_{tv}	CpG K_{ts}	CpG K_{tv}	K_h	K_i	K_t
Elongation factor	X	1028	0.467	0.0035	0.0030	0.1156	0.0366	0.0103	0.0019	0.0123
Shc	X	961	0.555	0.0096	0.0032	0.1468	0.0000	0.0147	0.0000	0.0147
Malate dehydrogenase	X	942	0.408	0.0053	0.0032	0.8240	0.0000	0.0107	0.0000	0.0107
Txreb	X	929	0.533	0.0056	0.0033	0.0380	0.0000	0.0098	0.0011	0.0108
Ferritin	X	551	0.588	0.0029	0.0019	0.0577	0.0000	0.0082	0.0000	0.0082
Desp	X	854	0.316	0.0053	0.0000	0.2326	0.0000	0.0065	0.0000	0.0065
Mean X values ± standard errors		877	0.478	0.0056	0.0025	0.1023	0.0087	0.0102	0.0006	0.0108
		69	0.042	0.0010	0.0007	0.0314	0.0088	0.0014	0.0003	0.0014

See legend to Table 1.

divergence were calculated for all sites, for CpG dinucleotides separately, and for all sites other than CpG dinucleotides. Under a neutral model, the amount of divergence, k, between sequences drawn from two species is given by

$$k = 2\mu t + 4N_e\mu,$$

where t is the time since the species have diverged measured in generations, μ is the mutation rate, and N_e is the ancestral effective population size (Kimura 1983a). The mutation rate was estimated from $\mu = k/(2t + 4N_e)$, assuming different values of t and N_e taken from the literature. Most of the uncertainty in the estimate of the total mutation rate derives from uncertainty concerning divergence time, ancestral population size, and generation length (rather than from sampling variance in estimates of rates of molecular evolution). Thus, a range of values for population size, divergence time, and generation length was used to provide a range of values for mutation rates. Comparison of divergence on the X chromosome to divergence on autosomes (k_X/k_A) was used to estimate the ratio of the male mutation rate to the female mutation rate ($\alpha = \mu_m/\mu_f$) following Miyata *et al.* (1987):

$$(k_X/k_A) = (2/3)(2 + \alpha)/(1 + \alpha).$$

This comparison was used to evaluate the strength of male-driven molecular evolution and the evidence for a lower mutation rate on the X (which would inflate the estimate of α). Minimum and maximum values of α were obtained by adjusting both numerator and denominator of k_X/k_A by one standard error (Smith and Hurst 1999). Estimates of the deleterious mutation rate were calculated from consideration of the total mutation rate and the fraction of the genome subject to constraint (Kimura 1983a,b; Kondrashov and Crow 1993; Eyre-Walker and Keightley 1999). The range of values for the total mutation rate was used to provide a reasonable range of values for the deleterious mutation rate. Throughout this article, mutation rates are expressed per site per generation unless stated otherwise.

RESULTS

Rates and patterns of molecular evolution: We observed a total of 199 differences between the human and chimpanzee sequences: 131 transitions (66%), 52 transversions (26%), and 16 insertion-deletion variants (8%). Insertion-deletion variants were less than one-tenth as common as nucleotide substitutions and consisted of changes of 1 bp (8 mutations), 2 bp (5 mutations), 3 bp (1 mutation), and 4 bp (2 mutations). Thus, 15/16 of these insertion-deletion variants would have resulted in frameshift mutations in coding regions. Approximately one-fifth of all single nucleotide mutations were transitions at CpG dinucleotides.

Rates of divergence for each of the 12 autosomal pseudogenes are given in Table 1 and rates of divergence for each of the 6 X-linked pseudogenes are given in Table 2. Rates are given separately for transitions and transversions at CpG sites and at non-CpG sites, for all single nucleotide substitutions, for insertion-deletion (length) variants, and for all changes together. Mean values for each of these categories are given separately for autosomal and X-linked pseudogenes.

The average level of divergence for autosomal pseu-

TABLE 3

Estimates of mutation rate assuming different divergence times (t) and different ancestral population sizes (N_e)

t (10^6 yr)	N_e	μ
4.5	10^4	2.7×10^{-8}
4.5	10^5	1.6×10^{-8}
5.0	10^4	2.5×10^{-8}
5.0	10^5	1.5×10^{-8}
5.5	10^4	2.3×10^{-8}
5.5	10^4	1.4×10^{-8}
6.0	10^4	2.1×10^{-8}
6.0	10^5	1.3×10^{-8}

Calculations are based on a generation length of 20 years and average autosomal sequence divergence of 1.33% (Table 1).

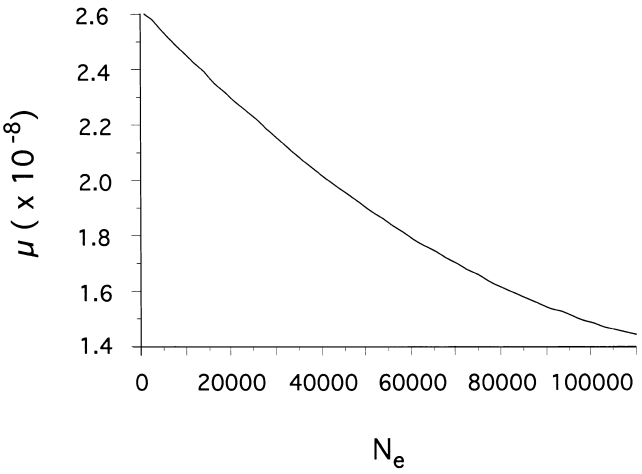

FIGURE 1.—Estimated mutation rate as a function of ancestral population size. Mutation rate calculated from $\mu = k/(2t + 4N_e)$, where $k = 0.0133$ and $t = 2.5 \times 10^5$, corresponding to a Homo-Pan divergence time of 5 mya and a generation time of 20 years.

dogenes was 1.33 ± 0.11% and ranged from a low of 0.4% to a high of 2.56% (Table 1). The average level of divergence for X-linked pseudogenes was 1.08 ± 0.14% and ranged from a low of 0.65% to a high of 1.47% (Table 2). While the average level of divergence was lower on the X chromosome than on autosomes, this difference was not significant (Mann-Whitney $U = 29.5$, $P > 0.5$). Substitution rates among loci varied by a factor of six and there was a slight trend toward higher rates in pseudogenes with intermediate GC content (WOLFE *et al.* 1989; Tables 1 and 2), although there was no statistical association between GC content and rate of evolution among all loci. To test for heterogeneity in substitution rates among loci, a chi-square test was performed on a 2×18 contingency table with columns corresponding to counts of variant and invariant sites between human and chimpanzee and rows corresponding to each of the 18 loci. This test was significant for all loci ($\chi^2 = 39.12$, d.f. = 17, $P = 0.002$) and for autosomal loci alone ($\chi^2 = 33.51$, d.f. = 11, $P = 0.0004$), but not for X-linked loci alone ($\chi^2 = 2.82$, d.f. = 5, $P = 0.73$). These results suggest that there may be underlying differences in mutation rates among some autosomal loci.

For autosomal loci, rates of divergence at CpG sites for transitions (8.75%) and transversions (2.32%) were approximately one order of magnitude higher than corresponding values at non-CpG sites (transitions 0.63%, transversions 0.29%; Table 1), consistent with the notion that CpG dinucleotides are mutational hotspots in mammalian genomes (COOPER and KRAWCZAK 1993; SOMMER 1995). A similar pattern is seen for X-linked loci (Table 2). For both autosomal and X-linked pseudogenes, divergence for single nucleotide changes was an order of magnitude larger than for insertion-deletion changes.

Mutation rates: The average mutation rate was calculated from the average autosomal rate of evolution assuming a generation time of 20 years (Table 3). Recent estimates of the time since humans and chimpanzees diverged (T) include 4.5 mya (TAKAHATA and SATTA 1997), 5.5 mya (KUMAR and HEDGES 1998), and 6.0 mya (GOODMAN *et al.* 1998). ARNASON *et al.* (1998) estimated the Homo-Pan divergence at 10–13 mya; however, their estimate is based on a calibration using distant, nonprimate species and is at odds with most other recent estimates. Mutation rates were calculated for a range of different human-chimpanzee divergence times and for two different ancestral population sizes. Mutation rate estimates vary from 1.3×10^{-8} (assuming $T = 6$ mya and $N_e = 10^5$) to 2.7×10^{-8} (assuming $T = 4.5$ mya and $N_e = 10^4$). If the average generation time is assumed to be 25 years (*e.g.*, EYRE-WALKER and KEIGHTLEY 1999), then mutation rates are estimated to be between 1.6×10^{-8} and 3.4×10^{-8}. Figure 1 shows the estimated mutation rate as a function of ancestral population size assuming $T = 5$ mya and a generation time of 20 years.

Underlying the average mutation rate is heterogeneity in rates for different sites and for different classes of mutations. We calculated rates for different types of mutations on the basis of a divergence time of 5 mya, ancestral population size of 10^4, and generation time of 20 years (Table 4). Rates vary from a high of $\sim 2 \times 10^{-7}$ (for transitions at CpG sites) to a low of $\sim 2 \times 10^{-9}$ (for length mutations).

Comparison of divergence values for X-linked and autosomal pseudogenes allows us to estimate α, the ratio of the male mutation rate to the female mutation rate. The mean observed divergence ($k_X/k_A = 0.0108/0.0133 = 0.81$) leads to an estimate of $\alpha = 3.6$ [$k_X/k_A = (2/3)(2 + \alpha)/(1 + \alpha)$] (MIYATA *et al.* 1987). Minimum and maximum values of α can be conservatively calculated using the standard errors in Tables 1 and 2 following SMITH and HURST (1999); $\alpha_{min} = 1$ and $\alpha_{max} = \infty$. Despite this enormous confidence interval, the mean

value of $\alpha = 3.6$ is in general agreement with or slightly smaller than previous estimates. For example, Ketterling *et al.* (1993) estimated $\alpha = 3.5$ from the germ line origin of mutations in the factor IX gene, and Huang *et al.* (1997) estimated $\alpha = 5$ from comparison of rates of molecular evolution of X- and Y-linked sequences. If the X chromosome mutation rate was lower than the autosomal mutation rate independent of sex-specific effects (McVean and Hurst 1997; Smith and Hurst 1999), then we would expect our estimate of α to be inflated relative to estimates derived from X-Y comparisons. For example, McVean and Hurst observed $k_X/k_A = 0.62$ for synonymous substitutions between mouse and rat, which is below the theoretical minimum value of $k_X/k_A = 0.66$ under male-driven molecular evolution and implies $\alpha = \infty$. In contrast, our estimate of $\alpha = 3.6$ is slightly smaller than estimates from X-Y comparisons ($\alpha = 5$, Huang *et al.* 1997; $\alpha = 6$, Shimmin *et al.* 1993).

DISCUSSION

Average mutation rate per nucleotide site: Mutation rates estimated for a range of divergence times and ancestral population sizes fall between 1.3×10^{-8} and 2.7×10^{-8} assuming a generation time of 20 years (Table 3) or between 1.6×10^{-8} and 3.4×10^{-8} assuming a generation time of 25 years. We suggest that 2.5×10^{-8} is a reasonable estimate of the average mutation rate per nucleotide site (but caution that the actual rate may be between 1.3×10^{-8} and 3.4×10^{-8}). The human diploid genome contains 7×10^9 bp (Marshall 1999) and thus ~175 new mutations per generation (range 91–238). It is clear that the accuracy of the estimate of mutation rate depends more on the uncertainty in divergence time, ancestral population size, and generation time than on the estimates of molecular substitution rates, which have standard errors approximately one-tenth of the mean values (Table 1).

How does this estimate compare with previous measures of the human mutation rate? Vogel and Motulsky (1997) summarize the average rate from a variety of autosomal dominant or X-linked phenotypes as $\sim 10^{-5}$ per locus (range 10^{-6}–10^{-4}). They also summarize the rate per nucleotide as estimated from disease genes as $\sim 10^{-8}$–10^{-9} per site. The range of values reported here (1.3–3.4×10^{-8}) is considerably higher than estimates that rely on screens of disease phenotypes. This probably reflects the fact that many mutations have only minor or no effects and thus go undetected in screens based on disease phenotypes.

The average autosomal pseudogene divergence reported here for nucleotide substitutions (1.21%) is not higher than previous estimates of synonymous site divergence between human and chimpanzee, suggesting that silent sites in human and chimpanzee lineages are evolving at the neutral rate (Li and Graur 1991). For example, Hammer (1995) reported 1.4% divergence at 1826 silent sites in 21 genes spread throughout the genome. Other estimates of human-chimpanzee synonymous divergence are slightly higher. Li *et al.* (1987) reported a divergence of 1.9% for 921 silent sites in 7 globin genes. Globin pseudogene divergence between human and chimpanzee is also slightly higher than reported here. Miyamoto *et al.* (1987) observed 1.6% divergence for 6974 bp of the ψη-globin region. The observations of higher silent site divergence for globin genes than for the genes in Hammer (1995) and of higher ψη-globin gene divergence than for the pseudogenes reported here raise the possibility that the β-globin gene region may have a higher-than-average local mutation rate. Alternatively, some of these differences may be due to the stochastic error associated with the small number of sites surveyed in some studies or due to sequencing errors.

Heterogeneity in mutation rates: There are clear differences in rates of mutation for different sites (CpG *vs.* non-CpG) and for different types of mutation (transitions, transversions, length variants; Table 4). In mammalian genomes, CpG sites are hotspots for transition mutations because of methylation-mediated deamination of 5-methylcytosine (Cooper and Krawczak 1993). The molecular mechanism responsible for increased transversion rates at CpG sites is unknown, although high CpG transversion rates are also seen in the factor IX gene (Sommer 1995; Sommer and Ketterling 1996). In mammalian genomes the high mutation rate observed at CpG sites causes these sites to be quickly lost. Thus while CpG transitions occur at a rate that is 10 times higher than the average mutation rate, they account for only one-fifth of all mutations. The absence of an outgroup sequence in our study precludes assigning polarity to the changes we observed; some unknown fraction of the mutations assigned to CpG sites may be creating CpG dinucleotides rather than eliminating them.

If methylation-mediated deamination of 5-methylcytosine is a strictly time-dependent process and not de-

TABLE 4

Estimates of mutation rate for different sites and different classes of mutation

Mutation type	Mutation rate
Transition at CpG	1.6×10^{-7}
Transversion at CpG	4.4×10^{-8}
Transition at non-CpG	1.2×10^{-8}
Transversion at non-CpG	5.5×10^{-9}
All nucleotide substitutions	2.3×10^{-8}
Length mutations	2.3×10^{-9}
All mutations	2.5×10^{-8}

Rates calculated on the basis of a divergence time of 5 mya, ancestral population size of 10^4, generation length of 20 yr, and rates of molecular evolution given in Table 1.

pendent on the number of germ-line cell divisions, then CpG transitions may not exhibit the X-autosome difference characteristic of male-driven molecular evolution (*e.g.*, Vogel and Motulsky 1997, Table 9.9). This hypothesis is supported by our observation that while non-CpG transition and transversion substitution rates are higher on the autosomes than on the X chromosome (consistent with male-driven molecular evolution), transitions at CpG sites do not follow this pattern. In fact, there is a slightly higher CpG transition rate on the X chromosome than on autosomes, but this difference is not significant. A similar result is found at silent sites in rodents (Smith and Hurst 1999). However, Anagnostopoulos *et al.* (1999) observe higher CpG transition rates on the Y chromosome than on the X chromosome, suggesting that the CpG mutation rate may depend on the number of cell divisions. A further test of the hypothesis that CpG mutation rates are time dependent rather than replication dependent could be provided by sequencing pseudogenes on the Y chromosome. Under this hypothesis, CpG substitution rates on the Y are predicted to be the same as on the X and autosomes. Also, if CpG mutations are strictly time dependent, they should provide an accurate molecular clock across groups with different generation times. Confounding these simple predictions is the possibility that methylation patterns may differ, either among chromosomes or over time. In our study, nothing is known of the methylation status of these pseudogenes.

X-chromosome-autosome comparisons: In principle, there are several reasons why the X chromosome might exhibit a lower substitution rate than the autosomes. First is male-driven molecular evolution. If most mutations arise in the male germ line (Haldane 1947), then we expect the X chromosome to have a lower substitution rate than the autosomes because the X chromosome spends only one-third of its time in males (Miyata *et al.* 1987). Second is a lower mutation rate on the X chromosome, independent of sex-specific effects (McVean and Hurst 1997). Mutation rates may reflect an adaptive balance between the benefits of reducing the deleterious mutation rate and the costs of increased fidelity. McVean and Hurst (1997) suggested that selection has reduced the X chromosome mutation rate because recessive X-linked deleterious mutations are exposed every generation in males. Third is nonneutral evolution (Smith and Hurst 1999). This hypothesis suggests that different substitution rates reflect the fact that the sites being surveyed are themselves subject to selection and that these selective forces operate differently on the X chromosome and on autosomes.

In our study, only pseudogenes were surveyed, so we can confidently reject the nonneutral explanation for X-autosome differences in substitution rate. Distinguishing between the other two hypotheses is tricky. If male-driven evolution is the only force operating, then we expect α to be roughly equivalent when estimated from X-autosome or X-Y comparisons. If the X chromosome has a lower mutation rate, independent of sex-specific effects, then we expect α to be larger when estimated from X-autosome comparisons than when estimated from X-Y comparisons. In rodents, α from X-autosome comparisons is larger than α from X-Y comparisons (Smith and Hurst 1999). Moreover, imprinted genes exhibit evolutionary rates similar to those on the X chromosome, suggesting that functionally hemizygous genes have distinct patterns from other genes, irrespective of their mode of inheritance (Smith and Hurst 1999). Both of these observations are consistent with the hypothesis that the X chromosome has a lower mutation rate. In contrast, the estimate of α derived here from X-autosomal comparisons ($\alpha = 3.6$) is somewhat lower than estimates derived from X-Y comparisons ($\alpha = 5$, Huang *et al.* 1997; $\alpha = 6$, Shimmin *et al.* 1993) and thus provides no evidence for a lower mutation rate on the X chromosome in primates. In virtually all of these studies, including the present one, the confidence intervals on α are enormous, making it difficult to distinguish among competing hypotheses. The central difficulty is that the male-driven-evolution hypothesis and the lower-mutation-on-the-X hypothesis both predict a lower level of divergence for the X relative to autosomes. Two other observations lend support to male-driven molecular evolution. First, there is good evidence for a bias in the sex ratio of disease mutations in families (reviewed in Vogel and Motulsky 1997; Hurst and Ellegren 1998). Second, birds also exhibit a higher mutation rate in males despite the fact that females are the heterogametic sex (Ellegren and Fridolfsson 1997).

Deleterious mutation rate: Kondrashov and Crow (1993) suggested that it is possible to calculate the genomic deleterious mutation rate (U) from an estimate of the neutral mutation rate and an estimate of the fraction of the genome that is subject to constraint. Our estimate of the neutral mutation rate is 175 mutations per genome per generation (range 91–238). As a minimum estimate of the fraction of the genome under constraint, we consider only coding sequences. The human genome contains ~70,000 genes (Bird 1995) with an average length of ~1500 bp (*e.g.*, Dunham *et al.* 1999; Eyre-Walker and Keightley 1999) for a total of 2.1×10^8 bp of coding sequences and 1.6×10^8 nonsynonymous sites. What proportion of nonsynonymous changes are neutral and what proportion are deleterious? The fraction that are neutral, f_o, can be calculated by comparing the total mutation rate, μ_t, with the substitution rate, $\nu_o = f_o\mu_t$ (Kimura 1983a,b). The proportion that are deleterious is $1 - f_o$. Using this approach, Kimura (1983b) estimated that 86% of nonsynonymous substitutions are deleterious. A more conservative estimate is obtained by assuming that silent substitutions are entirely neutral and thus reflect the total mutation rate. Then the ratio of nonsynonymous to silent substitutions (K_a/K_s) estimates f_o. This will be an underestimate to

the extent that silent mutations are deleterious. Data from Ohta indicate that the average $K_a/K_s = 0.27$ among 49 genes in primates (Ohta 1995). This suggests that 1.7% of the genome is subject to constraint [$(1.6 \times 10^8)(0.73)/(7 \times 10^9 \text{ bp}) = 0.017$]. The estimated genomic deleterious mutation rate, U, is thus ~ 3 ($U = 175 \times 0.017$), with a minimum value of 1.5 ($U = 91 \times 0.017$) and a maximum value of 4 ($U = 238 \times 0.017$), based on differences in divergence time, generation length, and ancestral effective population size. In fact, this range is likely to be biased downward because we have considered only nonsynonymous sites as potential targets for deleterious mutations. For example, a recent comparison of 100 kb of mostly noncoding DNA surrounding T cell receptor loci revealed striking conservation between humans and mice (Koop and Hood 1994), suggesting that much of this noncoding DNA may be functional and therefore includes targets for deleterious mutations.

Our estimate of $U = 3$ is slightly higher than another recent estimate in humans based on a similar approach ($U = 1.6$; Eyre-Walker and Keightley 1999). The difference between these estimates of U is due in part to the different estimates of constraint ($1 - K_a/K_s$). Eyre-Walker and Keightley's estimate of $1 - K_a/K_s = 0.38$ is considerably lower than the value of 0.73 obtained by Ohta (1995) for a different set of genes. The genes analyzed by Eyre-Walker and Keightley (1999) appear to have an unusually low level of constraint and may not be representative of the genome as a whole. Our estimate of $U = 3$ is considerably higher than recent estimates from mutation accumulation experiments in *Escherichia coli* ($U = 0.0002$; Kibota and Lynch 1996), *Caenorhabditis elegans* ($U = 0.005$; Keightley and Caballero 1997), and *Drosophila melanogaster* ($U = 0.02$–1, Mukai *et al.* 1972; Keightley 1996; Fry *et al.* 1999). However, mutations of small effect may go undetected in these experiments. In general, organisms with larger genomes appear to have a greater number of deleterious mutations, although it does not appear that the deleterious mutation rate is constant per base pair across these organisms.

The high deleterious mutation rate in humans presents a paradox. If mutations interact multiplicatively, the genetic load associated with such a high U would be intolerable in species with a low rate of reproduction (Muller 1950; Wallace 1981; Crow 1993; Kondrashov 1995; Eyre-Walker and Keightley 1999). The reduction in fitness (*i.e.*, the genetic load) due to deleterious mutations with multiplicative effects is given by $1 - e^{-U}$ (Kimura and Moruyama 1966). For $U = 3$, the average fitness is reduced to 0.05, or put differently, each female would need to produce 40 offspring for 2 to survive and maintain the population at constant size. This assumes that all mortality is due to selection and so the actual number of offspring required to maintain a constant population size is probably higher. The problem can be mitigated somewhat by soft selection (Wallace 1991) or by selection early in development (*e.g.*, *in utero*). However, many mutations are unconditionally deleterious and it is improbable that the reproductive potential *on average* for human females can approach 40 zygotes. This problem can be overcome if most deleterious mutations exhibit synergistic epistasis; that is, if each additional mutation leads to a larger decrease in relative fitness (Kondrashov 1995; Crow 1997; Eyre-Walker and Keightley 1999). In the extreme, this gives rise to truncation selection in which all individuals carrying more than a threshold number of mutations are eliminated from the population. While extreme truncation selection seems unrealistic, the results presented here indicate that some form of positive epistasis among deleterious mutations is likely.

We thank M. F. Hammer, A. S. Kondrashov, and N. A. Moran for discussions. We thank A. G. Clark, P. Keightley, and one anonymous reviewer for comments on the manuscript. This work was supported by the National Science Foundation.

LITERATURE CITED

Anagnostopoulos, T., P. M. Green, G. Rowley, C. M. Lewis and F. Giannelli, 1999 DNA variation in a 5-Mb region of the X chromosome and estimates of sex-specific/type-specific mutation rates. Am. J. Hum. Genet. **64:** 508–517.

Arnason, U., A. Gullberg and A. Janke, 1998 Molecular timing of primate divergences as estimated by two nonprimate calibration points. J. Mol. Evol. **47:** 718–727.

Bird, A. P., 1995 Gene number, noise reduction and biological complexity. Trends Genet. **11:** 94–100.

Cooper, D. N., and M. Krawczak, 1993 *Human Gene Mutation.* Bios Scientific Publishers, Oxford.

Crow, J. F., 1993 Mutation, mean fitness, and genetic load. Oxf. Surv. Evol. Biol. **9:** 3–42.

Crow, J. F., 1997 The high spontaneous mutation rate: Is it a health risk? Proc. Natl. Acad. Sci. USA **94:** 8380–8386.

Drake, J. W., B. Charlesworth, D. Charlesworth and J. F. Crow, 1998 Rates of spontaneous mutation. Genetics **148:** 1667–1686.

Dunham, I., A. R. Hunt, J. E. Collins, R. Bruskiewich, D. M. Beare *et al.*, 1999 The DNA sequence of human chromosome 22. Nature **402:** 489–495.

Ellegren, H., and A. K. Fridolfsson, 1997 Male-driven evolution of DNA sequences in birds. Nat. Genet. **17:** 182–184.

Eyre-Walker, A., and P. D. Keightley, 1999 High genomic deleterious mutation rates in hominids. Nature **397:** 344–347.

Fry, J. D., P. D. Keightley, S. L. Heinsohn and S. V. Nuzhdin, 1999 New estimates of the rates and effects of mildly deleterious mutation in *Drosophila melanogaster.* Proc. Natl. Acad. Sci. USA **96:** 574–579.

Goodman, M., C. A. Porter, J. Czelusniak, S. L. Page, H. Schneider *et al.*, 1998 Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. Mol. Phylogenet. Evol. **9:** 585–598.

Haldane, J. B. S., 1927 A mathematical theory of natural and artificial selection. Part V. Selection and mutation. Proc. Camb. Philos. Soc. **23:** 838–844.

Haldane, J. B. S., 1932 *The Causes of Evolution.* Longmans, Green, & Co., London.

Haldane, J. B. S., 1935 The rate of spontaneous mutation of a human gene. J. Genet. **31:** 317–326.

Haldane, J. B. S., 1947 The mutation rate of the gene for hemophilia, and its segregation ratios in males and females. Ann. Eugen. **13:** 262–271.

Hammer, M. F., 1995 A recent common ancestry for human Y chromosomes. Nature **378:** 376–378.

Huang, W., B. H. J. Chang, X. Gu, D. Hewett-Emmett and W. H.

Li, 1997 Sex differences in mutation rate in higher primates estimated from AMG intron sequences. J. Mol. Evol. **44:** 463–465.
Hurst, L. D., and H. Ellegren, 1998 Sex biases in the mutation rate. Trends Genet. **14:** 446–452.
Keightley, P. D., 1996 Nature of deleterious mutation load in Drosophila. Genetics **144:** 1993–1999.
Keightley, P. D., and A. Caballero, 1997 Genomic mutation rates for lifetime reproductive output and lifespan in Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA **94:** 3823–3827.
Ketterling, R. H., E. Vielhaber, C. D. K. Bottema, D. J. Schaid, M. P. Cohen *et al.*, 1993 Germ-line origins of mutation in families with Hemophilia B: the sex ratio varies with the type of mutation. Am. J. Hum. Genet. **52:** 152–166.
Kibota, T. T., and M. Lynch, 1996 Estimate of the genomic mutation rate deleterious to overall fitness in E. coli. Nature **381:** 694–696.
Kimura, M., 1968 Evolutionary rate at the molecular level. Nature **217:** 624–626.
Kimura, M., 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.
Kimura, M., 1983a *The Neutral Theory of Molecular Evolution.* Cambridge University Press, Cambridge, United Kingdom.
Kimura, M., 1983b Rare variant alleles in the light of the neutral theory. Mol. Biol. Evol. **1:** 84–93.
Kimura, M., and T. Moruyama, 1966 The mutational load with epistatic gene interactions in fitness. Genetics **54:** 1337–1351.
Kondrashov, A. S., 1995 Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over? J. Theor. Biol. **175:** 583–594.
Kondrashov, A. S., and J. F. Crow, 1993 A molecular approach to estimating the human deleterious mutation rate. Hum. Mutat. **2:** 229–234.
Koop, B. F., and L. Hood, 1994 Striking sequence similarity over almost 100 kilobases of human and mouse T-cell receptor DNA. Nat. Genet. **7:** 48–53.
Kumar, S., and B. Hedges, 1998 A molecular timescale for vertebrate evolution. Nature **392:** 917–920.
Li, W. H., and D. Graur, 1991 *Fundamentals of Molecular Evolution.* Sinauer Associates, Sunderland, MA.
Li, W. H., M. Tanimura and P. M. Sharp, 1987 An evaluation of the molecular clock hypothesis using mammalian DNA sequences. J. Mol. Evol. **25:** 330–342.
Marshall, E., 1999 A high-stakes gamble on genome sequencing. Science **284:** 1906–1909.
McVean, G. T., and L. D. Hurst, 1997 Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. Nature **386:** 388–392.
Miyamoto, M. M., J. L. Slightom and M. Goodman, 1987 Phylogenetic relations of human and African apes from DNA sequences in the ψη-globin region. Science **238:** 369–373.
Miyata, T., H. Hayashida, K. Kuma, K. Mitsuyasu and T. Yasunaga, 1987 Male-driven molecular evolution: a model and nucleotide sequence analysis. Cold Spring Harbor Symp. Quant. Biol. **52:** 863–867.
Mukai, T., S. I. Chigusa, L. E. Mettler and J. F. Crow, 1972 Mutation rate and dominance of genes affecting viability in Drosophila melanogaster. Genetics **72:** 335–355.
Muller, H. J., 1950 Our load of mutations. Am. J. Hum. Genet. **2:** 111–176.
Ohta, T., 1995 Synonymous and non-synonymous substitutions in mammalian genes and the nearly neutral theory. J. Mol. Evol. **40:** 56–63.
Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi *et al.*, 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:** 487–491.
Shimmin, L. C., B. H. J. Chang and W. H. Li, 1993 Male-driven evolution of DNA sequences. Nature **362:** 745–747.
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 rates? Genetics **152:** 661–673.
Sommer, S. S., 1995 Recent human germ-line mutation: inferences from patients with hemophilia B. Trends Genet. **11:** 141–147.
Sommer, S. S., and R. P. Ketterling, 1996 The factor IX gene as a model for analysis of human germline mutations: an update. Hum. Mol. Genet. **5:** 1505–1514.
Takahata, N., and Y. Satta, 1997 Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. Proc. Natl. Acad. Sci. USA **94:** 4811–4815.
Vogel, F., and A. G. Motulsky, 1997 Human genetics: problems and approaches. Springer-Verlag, Berlin.
Wallace, B., 1981 *Basic Population Genetics.* Columbia University Press, New York.
Wallace, B., 1991 *Fifty Years of Genetic Load.* Cornell University Press, Ithaca, NY.
Wolfe, K. H., P. M. Sharp and W. H. Li, 1989 Mutation rates differ among regions of the mammalian genome. Nature **337:** 283–285.

Communicating editor: A. G. Clark