The Causes of Synonymous Rate Variation in the Rodent Genome: Can Substitution Rates Be Used to Estimate the Sex Bias in Mutation Rate?

Nick G. C. Smith and Laurence D. Hurst

Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, United Kingdom

Manuscript received September 24, 1998
Accepted for publication February 16, 1999

ABSTRACT

Miyata *et al.* have suggested that the male-to-female mutation rate ratio (α) can be estimated by comparing the neutral substitution rates of X-linked (X), Y-linked (Y), and autosomal (A) genes. Rodent silent site X/A comparisons provide very different estimates from X/Y comparisons. We examine three explanations for this discrepancy: (1) statistical biases and artifacts, (2) nonneutral evolution, and (3) differences in mutation rate per germline replication. By estimating errors and using a variety of methodologies, we tentatively reject explanation 1. Our analyses of patterns of codon usage, synonymous rates, and nonsynonymous rates suggest that silent sites in rodents are evolving neutrally, and we can therefore reject explanation 2. We find both base composition and methylation differences between the different sets of chromosomes, a result consistent with explanation 3, but these differences do not appear to explain the observed discrepancies in estimates of α. Our finding of significantly low synonymous substitution rates in genomically imprinted genes suggests a link between hemizygous expression and an adaptive reduction in the mutation rate, which is consistent with explanation 3. Therefore our results provide circumstantial evidence in favor of the hypothesis that the discrepancies in estimates of α are due to differences in the mutation rate per germline replication between different parts of the genome. This explanation violates a critical assumption of the method of Miyata *et al.*, and hence we suggest that estimates of α, obtained using this method, need to be treated with caution.

IT has long been thought that, at least in humans, most mutations originate in males, ever since Haldane (1947) noted that most mutations causing Hemophilia A are paternally derived. The extent of any putative male mutation bias impacts on several areas including genetic counseling, understanding mutational processes, and many aspects of evolutionary biology (see Hurst and Ellegren 1998). Miyata *et al.* (1987) proposed to estimate the male-to-female mutation rate (α) by comparing nucleotide substitution rates in Y-linked, X-linked, and autosomal (A) genes. If the substitutions considered are selectively neutral (Kimura 1983) and if multiple substitutions can be properly accounted for, then substitution rates can provide unbiased estimates of the mutation rate.

Previous comparisons of synonymous substitution rates on the X chromosome and the autosomes of mouse and rat have yielded estimates of rodent $\alpha = \infty$ (Wolfe and Sharp 1993; McVean and Hurst 1997). In contrast, comparisons of substitution rates on the X and Y chromosomes have provided very different values for rodent α: about two when both synonymous and intronic substitution rates are used (Chang *et al.* 1994; Chang and Li 1995). A comparison of autosomal and Y-linked substitution rates in rodents has led to an estimate of α close to one (McVean and Hurst 1997). What causes these differences, and what do such differences tell us about the applicability of Miyata *et al.*'s (1987) method? We define three classes of explanations for the discrepancies between the X/A, Y/A, and X/Y estimates of α: statistical biases and artifacts, nonneutral evolution, and differences in mutation rate per germline replication.

Statistical biases and artifacts: The "errors" explanation supposes that there may be large standard errors in substitution rate estimates (Shimmin *et al.* 1993), and thus α values of two and ∞ may not be significantly different. An alternative "methodological" explanation suggests that the discrepancy between estimates of α may be due to different distance estimation methods yielding different substitution rates, as has been observed with other data sets (Smith and Hurst 1998b).

Nonneutral evolution: If synonymous substitutions in rodents are not selectively neutral, then different selective pressures on the different gene classes (X-linked, Y-linked, and autosomal) could lead to differences in K_S and thereby to discrepancies in estimates of α. For example, stronger selection could reduce synonymous substitution rates on the X-linked genes relative to the autosomal genes (Shimmin *et al.* 1993) because the X chromosome is exposed in males. Such exposure might strengthen selection against slightly deleterious synonymous mutations on the X chromosome relative to the

Corresponding author: Nick G. C. Smith, School of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom. E-mail: n.smith@bath.ac.uk

autosomes, thereby reducing X-linked synonymous substitution rates. In contrast, synonymous substitution rates are less likely to be reduced on the Y chromosome despite the continual hemizygous expression of Y-linked genes. Y-linked genes are weakly expressed and mutations on most Y-linked genes might well be masked by their X-linked paralogs (Hurst and Ellegren 1998). Both reasons suggest that selection on the Y chromosome against slightly deleterious mutations need be no stronger than on the autosomes.

This problem of nonneutrality might be circumvented by using intronic, rather than synonymous, substitution rates. If intronic substitutions are neutral in mammals, as is generally thought, then estimates of α based on intronic substitution rates should be safer than estimates based on synonymous substitution rates (Shimmin *et al.* 1993). But alignment difficulties make the intronic substitution rate estimation difficult (Smith and Hurst 1998b), and selection seems to affect introns as well as silent sites in Drosophila (Bauer and Aquadro 1997).

Differences in mutation rate per germline replication: The X/A and Y/A estimates of α are based on nonhomologous comparisons while the Y/X estimates have been based on homologous comparisons, which are presumably more reliable. The "compositional" explanation proposes that base composition effects on substitution rates could explain the discrepancy between X/A and Y/X estimates of α (Shimmin *et al.* 1993), because base frequency differences can lead to different mutational biases (Morton and Clegg 1995; Morton *et al.* 1997).

DNA methylation status is known to affect mutation rates strongly, with methylated CpG's mutating to TpG's at 10–20 times the rate of unmethylated CpG's (Kendrew 1994). If X-linked genes are methylated less than autosomal or Y-linked genes, then a low mutation rate on the X would be expected. This "methylation" explanation differs from the compositional explanation in that the mutation rate depends on base modifications rather than the bases themselves.

The "mutation rate selection" explanation proposes that selection might favor different optimal mutation rates for the different gene classes. The mutation rate on the X chromosome may be selectively lowered relative to that on the autosomes due to the exposure of highly deleterious mutations in males (McVean and Hurst 1997). Selection to reduce the mutation rate on the Y chromosome is expected to be weak because there are few genes on the Y chromosome, of which many are weakly expressed and/or masked by X-linked paralogs (Hurst and Ellegren 1998). A reduction of substitution rates on the X chromosome would lead to discrepancies between X/A, Y/A, and X/Y estimates of α for both intronic and synonymous substitution rates. Note that a modifier of mutation rate could act via methylation levels or composition. Thus the three explanations in the "differences in mutation rate per germline replication" class (methylation, composition, and mutation rate selection) are not necessarily in competition.

The "nonneutral evolution" and differences in mutation rate per germline replication classes of explanation both predict that hemizygously expressed genes should have a lower substitution rate than diploid expressed genes. We test this hypothesis by asking whether imprinted genes have low K_S values. Genomically imprinted genes are those for which expression is dependent upon the sex of the parent from which they are derived (Efstratiadis 1994). When reasoning similar to that employed for X-linked genes is used, both the nonneutral evolution and mutation rate selection arguments are consistent with reduced K_S values for imprinted genes relative to autosomal genes. However, X-linked genes and imprinted genes are not expected to yield identical substitution rates because there are several differences between the two classes of genes: the proportion of time hemizygously expressed, the effective population size, and potential expression level, composition, methylation, and recombination rate differences.

MATERIALS AND METHODS

Selection of protein coding sequences: A list of 470 mouse/rat mRNA pairs, with HOVERGEN 19 used to confirm orthology (Duret *et al.* 1994), was obtained from Makalowski and Boguski (1998). Mouse chromosome locations and gene names were obtained from the Mouse Genome Database (MGD; http://mgd.hgmp.mrc.ac.uk/). A total of 297 pairs of autosomal mouse-rat orthologs were determined. MGD was used to obtain a list of X-linked mouse mRNA sequences excluding those in the pseudoautosomal region. Gapped BLAST (Altschul *et al.* 1997) at the NCBI (http://www.ncbi.nlm.nih.gov/) was used to determine rat X-linked orthologs. A total of 37 pairs of X-linked mouse-rat orthologs were determined.

Fifteen mouse/rat imprinted orthologs were determined, with all mouse genes given in the Mammalian Genetics Unit Database (http://www.mgu.har.mrc.ac.uk/), with the exceptions of *Gabrb3* and *Mas* (see below). Gapped BLAST was used to find rat orthologs and MGD to provide gene names. A total of 15 pairs of imprinted mouse-rat orthologs were determined. The imprinted status of *Gabrb3* has yet to be fully demonstrated; however, the evidence in favor is now fairly compelling (Saitoh *et al.* 1994; Culiat *et al.* 1995; Odano *et al.* 1996; Delorey *et al.* 1998; Meguro *et al.* 1997). The status of *Mas* is ambiguous, with two reports supporting imprinting (Villar and Pedersen 1994; Miller *et al.* 1997) and two failing to find evidence of imprinting (Riesewijk *et al.* 1996; Schweifer *et al.* 1997). Given that the likelihood of a false positive is probably less than that of a false negative, we consider that the balance of evidence is adequate to allow its incorporation.

Selection of intron sequences: A list of intronic and exonic mouse/rat pairs of confirmed orthology was obtained from Hughes and Yeager (1997). MGD was used to determine gene names and chromosomal location, which yielded 20 complete coding sequences and 70 introns. The X-linked mouse/rat pairs described above were searched for intron sequences, which gave three exon and five intron X-linked pairs.

Selection of sequences for Y/X comparison: Alignments used in the Y/X comparisons were obtained from a previous

study (McVean and Hurst 1997), with both protein coding and intron alignments available for *Sry*, *Zfy*, and *Ube1y* (names as at MGD).

Sequence alignment: Alignments were performed using the GCG (Genetics Computer Group 1994) and EGCG (Rice 1997) sequence manipulation packages at HGMP (http://www.homepage.mrc.ac.uk/). FETCH was used to extract sequences from databases. GENETRANS was used to extract and combine exons automatically, while SEQED was used to extract introns manually. End-weighted PILEUP with the default gap penalties was used to perform exonic and intronic alignments. Exonic alignments were (if necessary) corrected by GAPFRAME to avoid frameshifts. In some cases, GAPFRAME could not correct the alignments. Then default PILEUP was used to produce protein alignments, and the program MRTRANS was used to recreate the DNA alignments from the protein alignments and the original DNA sequences (written by B. Pearson and available at HGMP). End-weighted CLUSTALW with the default gap penalties was also used to produce intronic alignments (Thompson *et al.* 1994), because it has been shown that different alignment packages with different default gap penalties produce significantly different intronic alignments (Smith and Hurst 1998b), and because there is no procedure for optimizing gap penalties (Altschul 1997).

Distance estimation: Several distance estimation methods were used. The default (and hopefully optimum) algorithmic estimation protocol used methods developed by Moriyama and Powell (1997) with Tamura's (1992) multiple hits correction method combined with Li's (1993) method for K_S and K_A estimation. Alternatively, Kimura's (1980) correction for multiple substitutions was applied to reduce the estimation error. Four different types of substitution rate were estimated: the synonymous rate (K_S), the nonsynonymous rate (K_A), the synonymous rate at fourfold degenerate sites (K_4), and the intronic rate (K_I). Estimates of K_4 are less sensitive to methodology than estimates of K_S (Li 1993).

The methods of Li *et al.* (1985) with Kimura's (1980) two parameter method for multiple substitution correction were also used to calculate K_S values. This enabled a more direct comparison with previous analyses that used such methods (Wolfe and Sharp 1993; McVean and Hurst 1997). Note that the method of Li *et al.* (1985) overestimates K_S by ~30% (Li 1993).

The maximum-likelihood package PAML (Yang 1997) was used to estimate substitution rates under a number of different assumptions. The program CODEML was used to estimate K_A and K_S under three settings: (1) "constant," a single rate for all sites; (2) "variable," a gamma distribution for variable substitution rates across sites; and (3) "correlated," a gamma distribution for variable substitution rates across sites and correlation of rates at adjacent codons. It should be noted, however, that the results obtained using the ''variable'' and ''correlated'' settings should be treated with caution since the gamma distribution may not apply to codon rates (Z. Yang, personal communication).

Calculating error limits of α estimates: The autosomal and X-linked synonymous and intronic substitution rates were compared to provide an estimate of α in a three-step procedure. First, substitution rate means and standard errors were calculated. Second, the X-to-autosome ratio of mean rates was calculated, and confidence limits for the ratio were calculated by adjusting both numerator and denominator by one standard error. Finally, Miyata *et al.*'s (1987) equation of $K_X/K_A = (2/3)(2 + \alpha)/(1 + \alpha)$ was used to calculate both α and its confidence limits.

Codon usage bias: One measure of the codon usage bias of a gene is given by the effective number of codons (ENC; Wright 1990). This statistic can vary from 20, with one codon used exclusively for each amino acid (high codon bias), to 61, with all synonymous codons used equally (low codon bias). Thus ENC is negatively correlated with codon usage bias.

We considered the possibility that ENC might be affected by both gene length and compositional bias. Random effects on smaller codon genes might reduce ENC, although it appears that ENC is highly robust to changes in gene length (Comeron and Aguade 1998; Moriyama and Powell 1998). Compositional bias makes codon usage less even, thereby reducing ENC (Wright 1990). To account for these effects, a simulated ENC value was calculated for each gene. The codon position-specific base frequencies (*e.g.*, frequency of C at the first codon position) of the gene were used to create 1000 simulated gene sequences of the same length as the original gene. Mean ENC was calculated for each set of simulated sequence, then the simulated ENC was divided by the real ENC to give the statistic CUBRE (codon usage bias relative to expected). This way of treating the data allows the CUBRE values to positively correlate with codon usage bias. Both ENC and CUBRE were used because the two statistics provide alternative null hypotheses of codon usage. ENC assumes no mutational biases between bases and hence equal use of all codons. CUBRE assumes that compositional biases are solely the result of mutational biases rather than selection on base composition.

Codon usage heterogeneity: A test of codon usage bias that accounts for mutational biases that might be caused by adjacent bases has been developed by Eyre-Walker (1991). Four nonoverlapping sets of codons are specified such that, within each set, all codons have the same base at the second codon position. For each set of codons, codon numbers are adjusted for fourth position base composition (*i.e.*, composition at the first base of the next codon). Then a chi-square test of heterogeneity is applied to determine whether the different amino acids within each set of codons show different codon usage patterns. Scaled chi-square values (similar to the codon usage statistic of Shields *et al.* 1988) are obtained by dividing chi-square values by the number of codons within each set (after adjustment for fourth-position base composition).

Using Eyre-Walker's notation, the four sets of codons are C (alanine, threonine, proline, and fourfold codons of serine; all codons have C at the second position), AGA (glutamic acid, lysine, and glutamine; all codons have A at the second position and either G or A at the third position), ATC (tyrosine, histidine, aspartic acid and asparagine; all codons have A at the second position and either T or C at the third position), and GTC (cysteine and the twofold degenerate codons of serine; all codons have G at the second position and T or C at the third position). The unscaled and scaled chi-square values for the AGA test, for example, are denoted χ^2_{AGA} and $sc\chi^2_{AGA}$, respectively.

If genes are too short, then codon numbers are too small, and the values obtained from the tests will not be chi-square distributed. Thus only suitably long genes were analyzed, with the adjusted number of codons required to be greater than the number of different codons multiplied by five. Too few X-linked and imprinted genes were long enough, so only the results obtained from autosomal genes are considered. Of the mouse autosomal genes, the numbers of genes used in the different tests were as follows: C test, 118; AGA test, 177; ATC test, 91; and GTC test, 85. Of the rat autosomal genes, the numbers were as follows: C test, 128; AGA test, 184; ATC test, 97; and GTC test, 84.

For these tests to be capable of detecting selection for optimal codon usage, it is required that amino acids within the same set of codons differ with respect to third position base of the optimal codon or at least with respect to the degree to

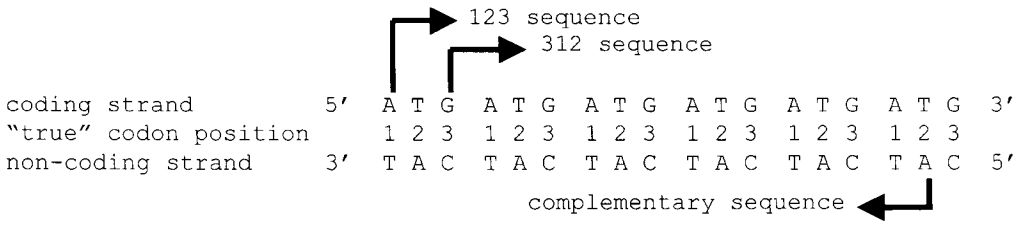

Figure 1.—Different sequences used in the analysis of codon usage heterogeneity (see materials and methods). Both the 123 and 312 strands are read 5′ to 3′ on the coding strand. The 123 sequence starts at position 1 of the first codon, while the 312 sequence starts at the third position of the first codon. The complementary strand is read 5′ to 3′ on the noncoding strand from the second position of the first codon.

which the optimal codon is favored. The data from Drosophila, with most amino acids having the same optimal third-position base (see Akashi 1994), suggest that Eyre-Walker's test may be weak.

Eyre-Walker (1991) details how comparisons of the codon usage heterogeneities of different reading frames can provide information on the nature of selection acting to cause such heterogeneities. He defines the 123 sequence as the complete original sequence, the 312 sequence as the original sequence read 5′ to 3′ from the third base, and the complementary sequence as the complementary strand read 5′ to 3′ from the second base (see Figure 1). Selection for optimal codon usage should give bias on the 123 frame only, while selection on RNA structure should give bias on the 123 and 312 reading frames, and selection for DNA structure should give bias on both reading frames and the complementary sequence.

Compositional analysis: A number of compositional characters were measured: all four base frequencies at fourfold degenerate sites and overall CpG frequency. Predicted CpG frequencies were calculated using gene length and C and G overall base frequencies. In addition, CpG/TpG and CpG/CpG orthologous pairs where the C/T change is silent (*i.e.*, C/T at third codon position) were counted. From this analysis, silent CpG mutability, defined as the ratio of mutated (CpG/TpG) to unmutated (CpG/CpG), could be calculated. Methylated CpG dinucleotides are known to mutate to TpG at high rates (Kendrew 1994), and thus the silent CpG mutability provides a measure of germline methylation density.

Statistical methods: We performed nonparametric tests of statistical significance. The Mann-Whitney *U*-test was used to compare sets of data (such as the substitution rates of X-linked and autosomal genes). Rank correlation statistics followed by the *z** transformation suggested by Hotelling (Sokal and Rohlf 1995), were used to examine the relationships between different gene characters. The Wilcoxon test was used to compare samples to expected values. As described above, we used chi-square tests to examine codon usage heterogeneity.

RESULTS

Statistical biases and artifacts

A number of different synonymous rate measures were used to provide estimates of α (Table 1). The finding that synonymous substitution rates are significantly lower for X-linked than autosomal genes holds for all measures (PAML correlated, $P = 0.032$; all other measures, $P < 0.0001$) and appears robust to the effects of outliers (data not shown). The eight different synonymous measures of α range from eight to ∞. In contrast with previous results (Miyata *et al.* 1987; Wolfe and Sharp 1993; McVean and Hurst 1997), our dataset estimates α to be less than (but not significantly different from) ∞ for six out of the eight measures. Differences in methodology do not appear to lead to qualitative differences in α estimation. The upper confidence limits for α are very high (∞ for six out of eight measures), while the lower confidence limits range between three and seven.

Four intronic measures of α were obtained (Table 1). As for the synonymous data, X-linked rates are lower than those on the autosomes, although not significantly so. The multiple substitutions correction method appeared to make little difference, but the two different alignment protocols gave rather different estimates of α. PILEUP estimated α to be ~17 and gave a minimum α estimate of 1.7, which is close to the values obtained from X/Y comparisons. CLUSTALW yielded an α estimate of ∞ and a lower limit of 4.4, results that are more in keeping with the synonymous measures. These intronic results should be treated with caution because of the small sample size ($N = 5$ on the X chromosome).

Imprinted genes do have low K_S values

If the discrepancies in α estimates are solely due to statistical biases and artifacts and the other explanations are incorrect, then hemizygously expressed genes should have the same synonymous substitution rates as diploid-expressed genes. However, we find that a low K_S may be a general feature of haploid-expressed genes, both X-linked and imprinted: imprinted genes have both a significantly lower mean K_S ($P = 0.0046$) and a significantly lower mean K_4 ($P = 0.0179$) when compared with autosomal genes when the default algorithmic estimation method is used (see Figure 2). When maximum-likelihood methods are used, autosomal K_S is higher than imprinted K_S (constant $P = 0.0034$, variable $P = 0.0017$, and correlated $P = 0.033$; see Figure

TABLE 1

A number of different methodologies are used for derivations of α estimates with X/A comparisons

Rate	Method	X mean	A mean	X/A	Mean α	Min. α	Max. α
K_S	PBL T	0.125	0.178	0.7	18.5	5.9	∞
K_S	PBL K	0.124	0.175	0.708	15.2	5.4	∞
K_S	LWL K	0.158	0.223	0.708	15.1	5.4	∞
K_4	T	0.132	0.18	0.735	8.7	3.9	119
K_4	K	0.132	0.178	0.741	7.95	3.7	60
K_S	PAML constant	0.125	0.183	0.686	33.2	7	∞
K_S	PAML variable	0.149	0.321	0.464	∞	∞	∞
K_S	PAML correlated	0.437	0.832	0.526	∞	∞	∞
K_I	PILEUP T	0.156	0.221	0.703	17.5	1.7	∞
K_I	PILEUP K	0.155	0.221	0.704	16.7	1.7	∞
K_I	CLUSTALW T	0.137	0.233	0.587	∞	4.4	∞
K_I	CLUSTRALW K	0.137	0.232	0.589	∞	4.3	∞

The rows show the steps in calculating α and its confidence limits (see also results section). The X chromosome and autosomal means are used to calculate the ratio of means (X/A). Then the ratio (and its upper and lower limits) is used to estimate α (and its lower and upper limits). The columns give different methodologies. K_S, K_4, and K_I are substitution rates at synonymous, fourfold degenerate, and intronic sites, respectively. The different methods are abbreviated as follows: PBL, Li (1993); LWL, Li *et al.* (1985); K, Kimura (1980); T, Tamura (1992). See materials and methods for details of the different PAML methods. The alignment program (PILEUP or CLUSTALW) is given fir the K_I data to demonstrate the strong effect of alignment protocol.

3). Similarly, autosomal K_S is higher than X-linked K_S when such methods are used (constant $P < 0.0001$, variable $P < 0.0001$, and correlated $P = 0.031$; see Figure 3). Both the nonneutral evolution and mutation rate selection arguments predict both X-linked and imprinted genes to have lower K_S rates than autosomal genes (see Introduction).

Nonneutral evolution of silent sites is not supported

We examine two issues concerning the proposition that nonneutral evolution of third-site mutations can account for the low K_S values seen in X-linked and imprinted genes. First we provide codon usage analyses to ask whether there is any evidence that selection has affected codon usage in rodents. Second, we ask whether, in principle, nonneutral evolution is adequate as a possible explanation.

Codon usage data: Although there exists a wealth of evidence to suggest that codon usage is selectively driven in Drosophila and bacteria (see Li 1997), several lines of evidence suggest that in mammals silent sites are neutral (see McVean and Hurst 1997). The "silent site selection" explanation belongs to the nonneutral evolution class of theories (see Introduction) and requires silent sites to be selectively constrained. We performed two sets of tests on rodent sequences to ask whether there is any evidence that codon usage is af-

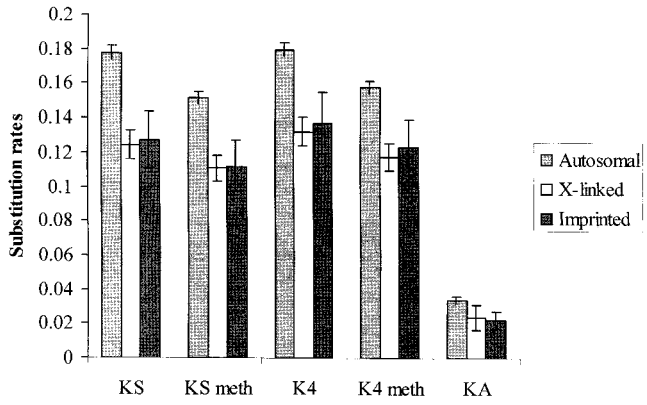

Figure 2.—Autosomal, X-linked, and imprinted genes compared with respect to synonymous substitution rate (K_S), nonsynonymous substitution rate (K_A), and fourfold substitution rate (K_4). The error bars give standard errors. The methods of Tamura (1992) and Li (1993) were used to calculate substitution rates. "meth" refers to rates estimated by excluding from consideration CpG ⇔ TpG substitutions at silent sites.

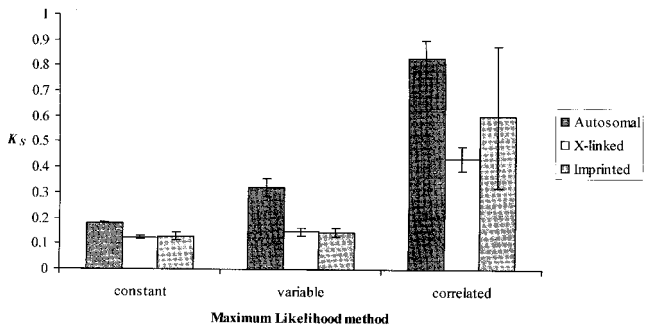

Figure 3.—Autosomal, X-linked, and imprinted genes compared with respect to synonymous substitution rates (K_S) obtained when different settings of the maximum-likelihood program PAML were used (see materials and methods).

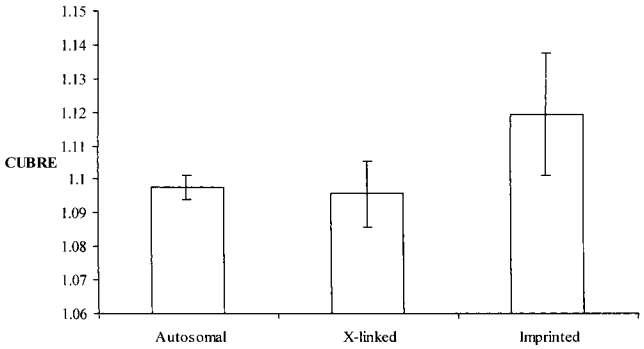

Figure 4.—Autosomal, X-linked, and imprinted genes compared with respect to CUBRE (codon usage bias relative to expected; see materials and methods).

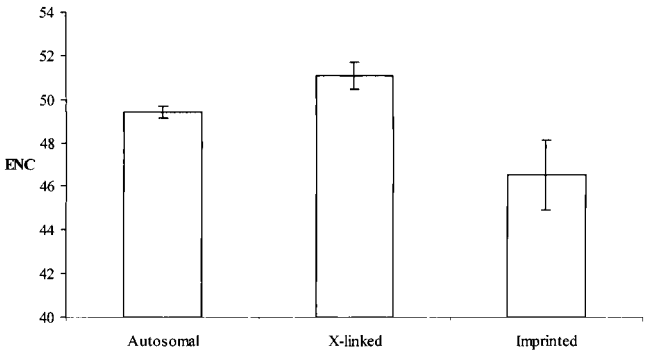

Figure 5.—Autosomal, X-linked, and imprinted genes compared with respect to ENC (effective number of codons; see materials and methods).

fected by selection and whether codon usage bias can explain the variation that we see in K_S.

CUBRE and ENC analysis: If synonymous codon usage were selectively driven, then one would expect to see a correlation between K_S and CUBRE and between K_S and ENC. If the synonymous codon usage of a gene is under strong selection, then CUBRE will be high and ENC low (high codon bias) because only optimal codons will be used, and K_S will be low because synonymous changes (from an optimal codon to a nonoptimal codon) will be highly constrained. Conversely, if the synonymous codon usage of a gene is under weak selection, then CUBRE will be low and ENC high (low codon bias) and K_S high. Such correlations have provided evidence for selection on codon bias in Drosophila (Sharp and Li 1989).

The predicted relationships between CUBRE and ENC values and K_S were tested in two different ways. First, if mean K_S is significantly higher for autosomal genes than for X-linked genes, do mean CUBRE and mean ENC show the expected relationships (significantly higher and lower, respectively, on the X chromosome)? Second, is there a significant correlation between K_S and CUBRE and between K_S and ENC among either autosomal or X-linked genes?

Both K_S and K_4 are significantly higher for autosomal than X-linked genes under the default algorithmic estimation method ($P < 0.0001$ in both cases, see Figure 2). K_S is also significantly higher for autosomal than X-linked genes under maximum-likelihood estimation (constant $P < 0.0001$, variable $P < 0.0001$, and correlated $P = 0.03$).

In conflict with the predictions of the silent site selection argument, mean CUBRE is lower ($P = 0.99$; see Figure 4) and mean ENC is higher ($P = 0.09$; see Figure 5) on the X-linked genes. No significant correlations were found between K_S and CUBRE (autosomal and X-linked, $P > 0.1$ and $P > 0.1$) or K_S and ENC ($P > 0.1$ and $P > 0.05$).

Considering imprinted genes allows a further test of the silent site selection argument. Given that imprinted genes have significantly lower synonymous substitution rates than autosomal genes (see above), the silent site selection argument predicts that imprinted genes should show greater codon usage bias; *i.e.*, imprinted genes should have a higher mean CUBRE and lower mean ENC than autosomes. These predictions hold for both CUBRE ($P = 0.35$) and ENC ($P = 0.09$), although the results are not significant.

Codon usage heterogeneity: Although the results of the codon bias analysis presented above provide no evidence that silent sites in rodents are under selection, CUBRE is significantly greater than unity in each of the three classes of autosomal, X-linked, and imprinted genes ($P < 0.001$ in all three cases). This result means that either silent sites are subject to selective pressures (and for some reason our previous tests did not have the power to demonstrate selection) or the scheme used to produce the expected ENC values is incorrect. The latter explanation seems plausible because simulated ENC values were calculated using only position-specific base frequencies and gene length. It is known that neighboring bases can affect mutation rates (Bulmer 1986), and thus simply combining position-specific base frequencies is unlikely to give reliable codon frequency estimates.

To control for mutational biases caused by adjacent bases, the codon usage heterogeneity tests developed by Eyre-Walker (1991) were applied to those autosomal genes that fulfilled the gene length criteria (see materials and methods). Restricting our analysis to autosomal genes, we aimed to determine whether the result of mean CUBRE greater than one is a methodological artifact.

For each of the four different tests (χ^2_{GTC}, χ^2_{C}, χ^2_{AGA}, and χ^2_{ATC}), two statistics were calculated: the proportion of genes showing significant heterogeneity at the 5% level and the overall χ^2 for all the genes. For both mouse and rat, only one of the overall χ^2 tests, overall χ^2_{AGA}, showed significant heterogeneity ($P < 0.001$ in both species). Only the rat χ^2_{AGA} test gave significantly more genes with significant bias than expected by chance ($P <$

0.025 with Yates' continuity correction), though the mouse χ^2_{AGA} test showed a similar tendency ($P < 0.1$).

Those genes that showed significant heterogeneity in any of the four tests in either mouse or rat were classed as the high χ^2 group, while the remaining genes with nonsignificant heterogeneities in both mouse and rat for all four tests formed the low χ^2 group. If the silent site selection argument is correct, then the high χ^2 group (more codon bias) should have lower K_S (silent sites more constrained) than the low χ^2 group. In conflict with this prediction, the high χ^2 group had a higher mean K_S than the lower χ^2 group, though the two means were not significantly different ($P = 0.21$).

The χ^2_{AGA} test was repeated with analysis of the 123, 312, and complementary sequences (see materials and methods). All autosomal genes that fulfilled the χ^2_{AGA} test length criteria for the 123, 312, and complementary sequences were included, thereby increasing the sample size. For both mouse and rat, overall χ^2_{AGA} was significant for the 123 sequence ($P < 0.001$ for both species) and complementary sequence ($P < 0.001$ for both species) but not for the 312 sequence ($P > 0.25$ for both species).

As explained above, the silent site selection argument predicts a negative correlation between codon bias and K_S. The four tests of heterogeneity were used to test this prediction, with scaled values used to control for gene length (see materials and methods). None of the four scaled heterogeneities correlated significantly with K_S ($P > 0.1$ in all four cases).

Nonneutral evolution, the problem of advantageous recessives, and the K_A-K_S correlation: One assumption of the notion that a low K_S in imprinted and X-linked genes might be the result of nonneutral evolution is that advantageous recessive silent mutations are rare. If advantageous recessives were common, then the exposure of X-linked and autosomal genes could lead to them having higher substitution rates than autosomal genes (Charlesworth *et al.* 1987), a result that would conflict with the low K_S values reported here.

We cannot test the frequency of advantageous recessive mutations directly. However, we can ask whether nonsynonymous substitutions show any evidence for the presence of advantageous recessive mutations. If purifying selection had acted on synonymous mutations to reduce X-linked K_S, one would expect an even greater reduction in X-linked K_A if nonsynonymous mutations were also under purifying selection, because nonsynonymous mutations are likely to be affected by stronger selection than synonymous ones.

K_A is significantly higher on the autosomes than on the X chromosome ($P = 0.003$; see Figure 2). Although the relationship is not significant, K_A is higher on autosomal genes than imprinted genes ($P = 0.52$; see Figure 2). These K_A relationships cannot, however, be considered independently of the K_S relationships because of the well-known K_A-K_S correlation (*e.g.*, Mouchiroud *et al.* 1995). Therefore, following McVean and Hurst (1997), we ask if X-linked K_A values are lower than one might expect given the low K_S values on the X chromosome. If the dominant mode of selection on the X were stabilizing selection then X-linked genes should have low K_A values given their K_S values. If instead directional selection on advantageous recessives were common then both might have K_A values higher than or comparable with autosomal genes with comparable K_S values.

We compared two rank orders, one for K_S and one for K_A, both of which gave the ranks of X-linked genes among autosomal genes. The K_A ranks were higher than the K_S ranks but not significantly so ($P = 0.14$). Thus we can reject the notion that X-linked K_A values are lower than expected on the basis of K_S values, which is a result that provides evidence against nonneutral evolution reducing the K_S on the X chromosome.

A similar analysis was applied to the imprinted genes. As for the X-linked genes, if the dominant mode of selection affecting imprinted genes is stabilizing selection then imprinted genes should have low K_A values given their K_S values. For imprinted genes among autosomal genes there is a tendency toward higher K_A ranks ($P = 0.08$). Consistent with the similarly high K_A ranks of both X-linked and imprinted genes, the comparison of imprinted and X-linked genes suggests no difference between the two classes ($P = 0.69$ for X-linked among imprinted and $P = 0.56$ for imprinted among X-linked). This result is consistent with elevated K_A ranks being an effect of hemizygous expression. If the hemizygously expressed genes are pooled, then the K_A ranks among the autosomal genes are significantly higher than the K_S ranks ($P = 0.041$).

The X and autosomes differ in sequence composition

The composition and methylation explanations for the discrepancies between X/Y, Y/A, and X/A estimates of α both invoke differences in the mutation rate per germline replication, which are the result of sequence differences. We tested the idea that the observed differences between X-linked, imprinted, and autosomal synonymous substitution rates might be due to sequence differences.

Most of the compositional features considered differ significantly between the autosomal genes and the X-linked genes (see Table 2). Most notably the mean GC4% on the X chromosome was only 53% compared with 61% on autosomes. This effect cannot account for all the variation in K_S, because imprinted genes, which also have a low K_S, have a GC4% of 65%, which is higher than that of the autosomal genes. No compositional feature differs significantly between the autosomal and imprinted genes (Table 2). We do find, however, that imprinted genes tend to be more GC rich than autosomal genes ($P = 0.053$). The fact that the difference is close to significance may help to explain the different

TABLE 2

Compositional statistics for autosomal (A), X-linked (X), and imprinted (I) genes

	Autosomal	X-linked	Imprinted	*P* (A = X)	*P* (A = I)
A4	0.184	0.204	0.164	0.06	0.2
C4	0.345	0.307	0.359	0.006	0.43
G4	0.267	0.221	0.289	<0.0001	0.22
T4	0.204	0.268	0.187	<0.0001	0.28
GC4	0.612	0.529	0.649	0.0001	0.053
TpG/CpG	1.184	1.502	0.688	0.52	0.075
CpG O/E	0.462	0.404	0.485	0.0037	0.92

Means are given, along with the Mann-Whitney *P* values for the AX and AI comparisons. A4, C4, G4, T4, and GC4 are the base frequencies at fourfold degenerate sites. TpG/CpG gives the ratio of silent Cpg ⇔ TpG substitutions to conserved silent CpG sites (see materials and methods). CpG O/E is the observed-over-expected ratio of CpG sites (equation for calculating expected values given in materials and methods).

results obtained by previous compositional comparisons of imprinted genes and autosomal genes (Neumann *et al.* 1995; McVean *et al.* 1996).

In addition to compositional differences, we have found evidence of methylation differences (see materials and methods). On the X chromosome not only are C's and G's rare, but also the frequency of CpG's (with C at a silent site), when controlled for GC%, is lower than that found on the autosomes ($P = 0.037$; see Table 3). But is there evidence that such differences lead to synonymous substitution rate differences?

Autosomal genes were used to examine correlations between sequence characters and synonymous substitution rates. No significant correlations were observed between K_4 and any of the fourfold base frequencies ($P > 0.5$ in all four cases). The number of silent CpG ⇔ TpG substitutions correlates positively with K_S ($P < 0.001$), which is not surprising given that such changes contribute directly to the number of synonymous substitutions. However, the ratio of silent CpG ⇔ TpG changes to the number of conserved silent CpG sites, which gives an indication of the mutability of silent CpG sites (see materials and methods), also correlates strongly with K_S ($P < 0.001$). In other words, the synonymous substitution rate at a class of site known to be strongly influenced by methylation status correlates strongly with the whole gene synonymous substitution rate. Such an effect has already been shown to explain at least some of the K_S variation within *Igf2r*, one of the imprinted genes in this study (Smith and Hurst 1998a). The ratio of observed to expected CpG (see materials and methods) also shows a positive correla-

TABLE 3

Estimates of α obtained using a variety of methodologies that use all three comparisons of X/A, Y/A, and X/Y

	K_S	K_S meth	K_4	K_I	PAML K_S constant	PAML K_S variable	PAML K_S correlated
X mean	0.125	0.111	0.132	0.137	0.125	0.149	0.4377
A mean	0.178	0.152	0.18	0.233	0.182	0.321	0.8327
Y mean	0.183	0.171	0.238	0.194	0.182	0.238	0.1997
Zfx	0.0833	0.0785	0.0867	0.144	0.0771	0.0771	0.335
Xfy	0.208	0.208	0.232	0.231	0.205	0.246	0.354
Ube1x	0.0861	0.0784	0.0796	0.167	0.085	0.0876	0.0335
Ube1y	0.138	0.119	0.223	0.175	0.121	0.13	0.125
α X/A	18.5	9.59	8.69	∞	33.2	∞	∞
α X/Y	1.93	2.12	2.99	1.79	1.88	2.28	0.358
α Y/A	1.07	1.29	1.95	0.711	0.994	0.588	0.136
α *Zfx/Zfy*	9.96	15.2	16.7	2.31	15.6	∞	1.09
α *Ube1x/Ube1y*	2.31	2.06	28.8	1.07	1.82	1.95	∞

Substitution rates (K_S, synonymous; K_4, fourfold degenerate; and K_I, intronic) were estimated either by maximum likelihood using the program PAML (see materials and methods for the settings used) or by the algorithmic methods of tamura (1992) and Li (1993). The "K_S meth" data were obtained by excluding CpG ⇔ TpG substitutions at silent sites from consideration. The default settings of the CLUSTALW were used to prepare intronic alignments. Values of α for the X/A, X/Y, and Y/A comparisons were calculated according to Miyata *et al.*'s (1987) equations.

tion with K_S ($P < 0.02$). This positive correlation is surprising, given that heavily methylated genes might be expected to have both high rates of silent substitution and low ratios of observed over expected CpG (both due to the methylation-induced mutation of CpG dinucleotides). It might be that methylated genes are under particularly strong stabilizing selection for some reason.

Could these links between methylation and K_S provide an explanation for the variation in K_S between genes? Silent site methylation-induced mutability is higher on the X chromosome than on the autosomes, which is the wrong direction for explaining the low K_S of X-linked genes. To test the methylation explanation more rigorously, K_S and K_4 values were calculated for all gene pairs with all silent CpG ⇔ TpG changes ignored. Autosomal K_S remained significantly higher than both X-linked K_S ($P < 0.0001$) and imprinted K_S ($P < 0.01$). Autosomal K_4 was significantly higher than X-linked K_4 ($P < 0.0001$) but was not significantly higher than imprinted K_4 ($P = 0.058$). Furthermore, the discrepancies between X/A and X/Y estimates of α remain even after removal of such methylation-induced changes (see below and Table 3).

X/Y and Y/A estimates of α

To reject an explanation for the discrepancies between estimates of α it is not enough to simply show that the explanation is unable to yield X/A α estimates close to the values of about two given by X/Y comparisons. It must also be demonstrated that the explanation is unable to raise the α estimates provided by X/Y and Y/A comparisons. Therefore we have used Y-linked sequences to give X/Y and Y/A estimates of α. The methodological explanation can be tested by observing the impact of changes in methodology on the relative α estimates of the X/A, X/Y, and Y/A comparisons.

Using mean chromosomal class substitution rates, the relative values of α estimates are conserved across methodologies (see Table 3), which suggests a rejection of the methodology explanation. For both algorithmic estimates of K_S, K_4, and K_I rates and for maximum-likelihood estimates of K_S, the X/A comparisons give much higher estimates of α than the X/Y and Y/A comparisons. However, the homologous X/Y comparisons of *Zfx*/*Zfy* and *Ube1x*/*Ube1y* do yield some α estimates greater than 2 (Table 3). The K_S *Zfx*/*Zfy* comparison gives an α estimate of 10. A previous study found $\alpha = \infty$ in this comparison (Shimmin *et al.* 1994), and although we concur with the result of a high α estimate, we disagree with the conclusion of the authors of that study that such a result was due to silent site constraints affecting *Zfx* being relaxed on *Zfy*. *Zfx* shows no evidence of codon usage bias [CUBRE = 1.017; *c.f.*, mean X-linked CUBRE = 1.0958; furthermore, none of Eyre-Walker's (1991) codon usage tests showed significant heterogeneity], and therefore silent sites on *Zfx* are unlikely to be under selective constraints. Using K_4 rates, both the *Zfx*/*Zfy* and *Ube1x*/*Ube1y* comparisons give high α estimates. This appears to be due to K_4 being greater than K_S on the Y chromosome, presumably due to a low rate at twofold degenerate sites.

The methylation explanation predicts that differences between chromosomes in methylation density could lead to differences between chromosomes in rates. Although this explanation does not appear to be able to reduce the X/A estimate of α to about two, might it be able to increase the X/Y and Y/A α estimates? To test this hypothesis we recalculated K_S values while ignoring CpG ⇔ TpG changes at silent sites (see Table 3). The X/A estimate decreased and the X/Y and Y/A estimates increased, but there remained large discrepancies between α estimates. The *Ube1x*/*Ube1y* comparison gave a lower α estimate on the removal of methylation-induced changes, but the *Zfx*/*Zfy* α estimate increased to a value actually above the X/A estimate.

Even if selection does act to reduce mutation rates on the X chromosome, it does not then automatically follow that mutation rates should also be selectively reduced on the Y chromosome (see Introduction). In keeping with this analysis K_S, K_4, and K_I are not nearly as low on the Y chromosome as they are on the X chromosome, although the sample size is very low (see materials and methods).

DISCUSSION

We considered three explanations for the discrepancies between X/A, Y/A, and Y/X estimates of α (see Introduction).

Statistical biases and artifacts: Are the errors in estimates of α so large that values of 2 and ∞ are not significantly different? Or was it the use of biased distance estimation measures in previous analyses that led to the discrepancy in α estimates? Using both synonymous and intronic substitutions and a variety of methodologies, we used the X/A comparison to obtain estimates of α (see Table 1). The intronic and synonymous estimates differ with respect to expected value of α (finite but large for synonymous, infinite for intronic), but the large errors mean that both estimates provide 95% confidence intervals for α between 0.5–7 and ∞ when a range of different methods is used. These lower limits approach (and for the PILEUP intronic estimate fall below) estimates of α of ~2 obtained from Y/X comparisons (intronic and synonymous rates show similar results). However, we do not feel it is appropriate to calculate a statistic for the difference between the X/A and Y/X estimates of α, because the confidence limits previously obtained for Y/X estimates (Chang *et al.* 1994; Chang and Li 1995) are of a different nature than the confidence limits we have obtained.

The confidence limits provided in this study are based on the variation observed in synonymous substitution

rates between many genes (37 X-linked and 297 autosomal genes), which thereby reduce any bias caused by substitution rate variation (Shimmin *et al.* 1993). In contrast, studies that use Y/X comparisons to estimate α have ignored variation in substitution rates between genes, and have instead used the errors inherent in estimating substitution rates (Chang *et al.* 1994; Chang and Li 1995). This method might be defended on the basis that homologous genes should show little variation in substitution rates, but an assumption of absolutely no variation in rates between homologous genes (other than that caused by rate estimation) seems unlikely to be correct. Thus we regard the Y/X 95% confidence intervals of 1.0–3.2 (Chang *et al.* 1994) and 1.0–3.9 (Chang and Li 1995) as perhaps too narrow.

Because high estimates of α are obtained when both biased and unbiased methods of K_S and K_I estimation are used, the methodological explanation can be rejected. But unless X/A, Y/A, and X/Y estimates of α can be statistically compared, it is impossible to discount fully the errors argument. That only one of nine X/A α estimates has a 95% confidence limit under two (and that estimate was based on the smallest sample) does suggest rejection of the errors hypothesis. Furthermore, estimates of α that use the X/A, X/Y, and Y/A comparisons with the three distance measures of K_S, K_4, and K_I all yielded the same result of $\alpha_{X/A}$ greater than $\alpha_{X/Y}$ and $\alpha_{Y/A}$, which supports rejection of both the errors and methodological arguments.

It should be noted that the relatively small expansion in dataset size from McVean and Hurst's (1997) study to the present one appears to have caused a considerable increase in the K_S estimate of X/A. McVean and Hurst obtained an X/A ratio of 0.62. Even when the same methods as McVean and Hurst's are used, the dataset used here still gives an X/A ratio of 0.708. If further expansions in dataset size lead to similar increases in the X/A ratio, the X/A and Y/X estimates of α may converge further. Further evidence for larger datasets yielding higher X/A ratios comes from the human and rodent comparison. With 35 autosomal and only 4 X-linked genes, Miyata *et al.* (1987) obtained an X/A ratio of 0.60, while a recent study of 559 autosomal and 71 X-linked genes (Terada *et al.* 1997) gave an X/A ratio of 0.78.

Nonneutral evolution: The discrepancies between synonymous X/A, Y/X, and Y/A estimates of α may be due to stronger selection against silent mutations on the X chromosome (for further details see Introduction). If this is a necessary and sufficient explanation of α estimate discrepancies, then two requirements must be fulfilled. The first requirement is that there be no α estimate discrepancy when intronic data are used. The second requirement is that silent substitutions must be selectively constrained.

The initial requirement seems to be contravened by the intronic X/A, X/Y, and Y/A estimates of α. X/A gave $\alpha = \infty$, while X/Y and Y/A both gave α less than three (Table 3). However, it is not possible to determine whether the discrepancies in α estimates are significant.

Concerning the second requirement, our codon usage analyses provide no support for the hypothesis that silent substitutions in rodents are selectively constrained. Neither X-linked nor imprinted CUBRE is significantly higher than autosomal CUBRE, despite X-linked and imprinted K_S being significantly lower than autosomal K_S. Furthermore, neither autosomal nor X-linked genes demonstrate significant negative correlations between K_S and CUBRE.

Analysis of codon usage heterogeneities for autosomal genes reveals that only one of the four codon sets (A^G_A) shows significant levels of heterogeneity. Comparisons of A^G_A tests for different reading frames provide evidence that the significant heterogeneity for the 123 sequences is not the result of selection, because significant heterogeneity is found for the 123 and complementary sequences but not for the 312 sequences. Such results are similar to those of Eyre-Walker (1991, p. 447), and we concur with this author's appraisal that "*It is difficult to think of any form of selection that would lead to significant bias appearing on the complementary strand and in other reading frames, but which is attenuated in the 312 reading frame.*"

Selection for optimal codon usage should give bias on the 123 frame only, while selection on RNA structure should give bias on the 123 and 312 reading frames, and selection for DNA structure should give bias on both reading frames and the complementary sequence. One possible explanation for the attenuation of bias in the 312 reading frame is the interaction of neighboring base mutational biases and longer range mutational biases (Eyre-Walker 1991). The latter class of biases can act at distances of several bases (Bulmer 1990), and are thus not accounted for by the method of second- and fourth-codon-position correction employed here. The base position examined for codon usage bias for both the 123 reading frame and the complementary sequence is the "true" third position (see Figure 1). Because the neighboring nucleotides (true first and second positions) are constrained, mutational biases will be able, over time, to generate significant heterogeneities at third positions. But the base position examined for codon usage bias for the 312 sequence is the true second position, which has the true first and third positions for neighbors. Because the third-codon position is only weakly constrained and thus rapidly evolving, mutational biases may be unable to build up sufficiently before the third site, and thus the direction of bias, changes.

The lack of significant heterogeneity in three tests out of four, the demonstration that the A^G_A bias is unlikely to be the result of selection for codon usage, and the fact that so few of the genes showed significant bias when considered individually all argue against the nonneutral

evolution theory. Furthermore, no negative correlation between codon bias and K_S was observed, and the "high bias" group of genes did not have a significantly lower K_S than the "low bias" group of genes.

Furthermore, it is unclear whether nonneutral evolution at third sites can explain the low K_S values that we observe. If the silent site selection explanation is correct, and if nonsynonymous mutations are as likely to be deleterious as synonymous ones, then the K_A values for the X-linked and imprinted genes should be lower relative to the autosomal genes than their K_S values. In fact, for both X-linked and imprinted genes, K_A values are higher than would be expected on the basis of their K_S values but not to a significant extent.

Differences in mutation rate per germline replication: *Composition and methylation explanations:* If the composition or methylation arguments are to explain low K_S on the X chromosome, there must exist significant sequence differences between the X chromosome and the autosomes, and those compositional characters that differ significantly must be shown to correlate (and in the required direction) with K_S. The former requirement is fulfilled by most of the compositional characters investigated (overall and fourfold base frequencies and CpG statistics); but only one sequence character was observed to correlate significantly with K_S or K_4. This was silent CpG mutability, which can be interpreted as a measure of methylation density.

Methylation at CpG sites appears to have a large effect on mutation rates. CpG ⇔ TpG substitutions constitute a reasonably large proportion of silent substitutions (the autosomal, X-linked, and imprinted estimates of K_S fell by 15, 11, and 12%, respectively, when such substitutions were ignored), and methylation levels can alter mutability without affecting the protein-coding sequence of a gene. Unfortunately for the methylation theory, X-linked genes actually showed greater silent CpG mutability than autosomal genes, and the removal of CpG ⇔ TpG substitutions from estimates of K_S had little effect on the significant difference between X-linked and autosomal K_S means. Such adjusted K_S values provide a lower X/A estimate of mean α than the unadjusted K_S values, although X/Y and Y/A estimates of α remain much lower than the X/A estimate.

It is surprising that CpG mutability is higher for X-linked genes than for autosomal genes, when one considers that male germline DNA is far more heavily methylated than female germline DNA (Monk *et al.* 1987). For example, Ketterling *et al.* (1993) found a male-to-female mutation bias of 11 when considering CpG dinucleotides in the factor IX gene. The reduction in the X/A estimate of α upon removal of silent CpG ⇔ TpG substitutions does fit in with a higher male bias for such mutations than for other sorts of point mutations.

Methylation status appears to be a crucial mechanism in genomic imprinting, and the fact that many imprinted genes possess CpG-rich regions implies that imprinted genes are relatively unmethylated in the germline (Constancia *et al.* 1998). Such a methylation profile is consistent with low K_S values for imprinted genes. Indeed, although the differences are not significant, the imprinted genes show a lower silent CpG mutability and a higher observed-over-expected CpG ratio (indicative of lower overall CpG mutability) when compared to the autosomal genes. However, after the removal of CpG ⇔ TpG substitutions imprinted genes still have a significantly lower K_S than autosomal genes (Figure 2), and so such substitutions cannot provide a complete explanation of the low K_S of imprinted genes.

It is impossible to refute conclusively either the composition or methylation explanations because in principle almost any sequence character could affect mutation rates, and one cannot examine every possibility.

Mutation rate selection explanation: The mutation rate selection explanation is supported by some circumstantial evidence. Both intronic and synonymous estimates of α appear to show discrepancies between the X/A and Y/X comparisons. Imprinted genes show the low K_S predicted on the basis of their hemizygous expression. The finding that K_S on the Y is higher than on the X is hardly supportive, but at least the result is consistent with selection on mutation rates.

Detailed analysis of imprinted genes provides further suggestive evidence in favor of the existence of modifiers of mutation rates. If imprinted genes have low K_S because of modifiers, then clusters of imprinted genes should have lower K_S than imprinted genes on their own, because the selective pressure favoring modifiers of mutation rate should increase with the number of genes affected by the modifier. So the more imprinted genes there are in a cluster, the better the chance of a modifier of the mutation rate becoming fixed. All of the imprinted genes in the dataset exist in clusters except for *Htr2a* (Yang *et al.* 1992), *Ins1* (Wentworth *et al.* 1986), and *Rasgrf1* (Plass *et al.* 1996). Nonclustered imprinted genes have higher K_S than clustered imprinted genes, and although the difference is not significant ($P = 0.22$) this may well be a result of limited sample size. We feel this result is suggestive and worthy of further analysis.

The status of Miyata's method for assessing α: How do the above results relate to the validity of Miyata *et al.*'s (1987) proposal that the male-to-female mutation rate ratio (α) can be estimated by comparing the substitution rates of X-linked, Y-linked, and autosomal genes? Their method (hereafter Miyata's method) depends on two assumptions. The first assumption is that mutation rates are, at least in principle, measurable. For molecular evolutionary methods this requires the presence of neutral sites and the use of appropriate multiple substitution correction methods. The second assumption is that all of the differences between the mutation rates of genes are because of the amount of time spent in the two germlines.

The statistical biases and artifacts explanation does not affect the logic underpinning Miyata's method but is certainly unsatisfying and probably incorrect. The nonneutral evolution hypothesis impinges on the first assumption that mutation rates are measurable. The evidence we have presented against this hypothesis lends credence to the first assumption of Miyata's method.

The differences in mutation rate per germline replication explanations suggest violations of the second assumption of Miyata's method. Different mutation rates on different chromosomes might be the result of different patterns of base composition or methylation or hemizygosity rather than the result of the amount of time spent in the two germlines. Neither the composition nor methylation explanations can be rejected (but given that both theories appear unfalsifiable, nothing can be read into this), and the evidence in favor of the mutation rate selection argument is only circumstantial. However, we consider the finding of a connection between low K_S and hemizygosity to be best interpreted as a violation of the second assumption of Miyata's method. Therefore we feel that caution should be exercised when interpreting estimates of α obtained when Miyata's method is used.

The authors thank Nick Britton, John Barrett, Gil McVean, Adam Eyre-Walker, Etsuko Moriyama, and two anonymous referees.

LITERATURE CITED

Akashi, H., 1994 Synonymous codon usage in Drosophila melanogaster—natural selection and translational accuracy. Genetics **136:** 927–935.

Altschul, S. F., 1997 Sequence comparison and alignment, pp. 137–168 in *DNA and Protein Sequence Analysis*, edited by M. J. Bishop and C. J. Rawlings. Oxford University Press, Oxford.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. H. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Bauer, V. L., and C. F. Aquadro, 1997 Rates of DNA sequence evolution are not sex-biased in Drosophila melanogaster and D-simulans. Mol. Biol. Evol. **14:** 1252–1257.

Bulmer, M., 1986 Neighboring base effects on substitution rates in pseudogenes. Mol. Biol. Evol. **3:** 322–329.

Bulmer, M., 1990 The effects of context on synonymous codon usage in genes with low codon usage bias. Nucleic Acids Res. **18:** 2869–2873.

Chang, B. H. J., and W.-H. Li, 1995 Estimating the intensity of male-driven evolution in rodents by using X-linked and Y-linked *Ube-1* genes and pseudogenes. J. Mol. Evol. **40:** 70–77.

Chang, B. H. J., L. C. Shimmin, S. K. Shyue, D. Hewett-Emmett and W.-H. Li, 1994 Weak male-driven molecular evolution in rodents. Proc. Natl. Acad. Sci. USA **91:** 827–831.

Charlesworth, B., J. A. Coyne and N. H. Barton, 1987 The relative rates of evolution of sex chromosomes and autosomes. Am. Nat. **130:** 113–146.

Comeron, J. M., and M. Aguade, 1998 An evaluation of measures of synonymous codon usage bias. J. Mol. Evol. **47:** 268–274.

Constancia, M., B. Pickard, G. Kelsey and W. Reik, 1998 Imprinting mechanisms. Genet. Res. **8:** 881–900.

Culiat, C. T., L. J. Stubbs, R. P. Woychik, L. B. Russell, D. K. Johnson *et al.*, 1995 Deficiency of the beta-3 subunit of the type-a gamma-aminobutyric-acid receptor causes cleft-palate in mice. Nat. Genet. **11:** 344–346.

Delorey, T. M., A. Handforth, S. G. Anagostaras, G. E. Homanics, B. A. Minassian *et al.*, 1998 Mice lacking the beta3 subunit of the GABAA receptor have the epilepsy phenotype and many of the behavioural characteristics of Angelman syndrome. J. Neurosci. **18:** 8505–8514.

Duret, L., D. Mouchiroud and M. Gouy, 1994 Hovergen—a database of homologous vertebrate genes. Nucleic Acids Res. **22:** 2360–2365.

Efstratiadis, A., 1994 Parental imprinting of autosomal mammalian genes. Curr. Opin. Genet. Dev. **4:** 265–280.

Eyre-Walker, A., 1991 An analysis of codon usage in mammals: selection or mutation bias? J. Mol. Evol. **33:** 442–449.

Genetics Computer Group, 1994 Program Manual for the Wisconsin Package, Version 8, September 1994, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin 53711.

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.

Hughes, A. L., and M. Yeager, 1997 Comparative evolutionary rates of introns and exons in murine rodents. J. Mol. Evol. **45:** 125–130.

Hurst, L. D., and H. Ellegren, 1998 Sex biases in the mutation rate. Trends Genet. **14:** 446–452.

Kendrew, J. (Editor), 1994 *The Encyclopedia of Molecular Biology*. Blackwell, Oxford.

Ketterling, R. P., 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.

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., 1983 *The Neutral Theory of Evolution*. Cambridge University Press, Cambridge, United Kingdom.

Li, W. H., 1993 Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36:** 96–99.

Li, W. H., 1997 *Molecular Evolution*. Sinauer Associates, Sunderland, MA.

Li, W. H., C. I. Wu and C. C. Luo, 1985 A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2:** 150–174.

Makalowski, W., and M. S. Boguski, 1998 Evolutionary parameters of the transcribed mammalian genome: An analysis of 2,820 orthologous rodent and human sequences. Proc. Natl. Acad. Sci. USA **95:** 9407–9412.

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.

McVean, G. T., L. D. Hurst and T. Moore, 1996 Genomic evolution in mice and men—imprinted genes have little intronic content. Bioessays **18:** 773–775.

Meguro, M., K. Mitsuya, H. Sui, K. Shigenami, H. Kugoh *et al.*, 1997 Evidence for uniparental, paternal expression of the human GABA(A) receptor subunit genes, using microcell-mediated chromosome transfer. Hum. Mol. Genet. **6:** 2127–2133.

Miller, N., A. H. McCann, D. O'Connell, I. S. Pedersen, V. Spiers *et al.*, 1997 The *MAS* proto-oncogene is imprinted in human breast tissue. Genomics **46:** 509–512.

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.

Monk, M., M. Boubelik and S. Lehnert, 1987 Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineage during mouse embryo development. Development **99:** 371–386.

Moriyama, E. N., and J. R. Powell, 1997 Synonymous substitution rates in Drosophila: Mitochondrial versus nuclear genes. J. Mol. Evol. **45:** 378–391.

Moriyama, E. N., and J. R. Powell, 1998 Gene length and codon usage bias in Drosophila melanogaster, Saccharomyces cerevisiae and Escherichia coli. Nucleic Acids Res. **26:** 3188–3193.

Morton, B. R., and M. T. Clegg, 1995 Neighboring base composition is strongly correlated with base substitution bias in a region of the chloroplast genome. J. Mol. Evol. **41:** 597–603.

Morton, B. R., V. M. Oberholzer and M. T. Clegg, 1997 The

influence of specific neighboring bases on substitution bias in noncoding regions of the plant chloroplast genome. J. Mol. Evol. **45:** 227–231.

Mouchiroud, D., C. Gautier and G. Bernardi, 1995 Frequencies of synonymous substitutions in mammals are gene-specific and correlated with frequencies of nonsynonymous substitutions. J. Mol. Evol. **40:** 107–113.

Neumann, B., P. Kubicka and D. P. Barlow, 1995 Characteristics of imprinted genes. Nat. Genet. **9:** 12–13.

Odano, I., T. Anezaki, H. Ohkubo, Y. Yonekura, Y. Onishi *et al.*, 1996 Decrease in benzodiazepine receptor-binding in a patient with Angelman Syndrome detected by I-123 iomazenil and single-photon emission tomography. Eur. J. Nucl. Med. **23:** 598–604.

Plass, C., H. Shibata, I. Kalcheva, L. Mullins, N. Kotelevtseva *et al.*, 1996 Identification of Grf1 on mouse chromosome 9 as an imprinted gene by RLGS-M. Nat. Genet. **14:** 106–109.

Rice, P., 1997 *Program Manual for the EGCG Package.* The Sanger Centre, Hinxton Hall, Cambridge, CB10 1RQ, England.

Riesewijk, A. M., M. T. Schepens, E. M. Mariman, H. H. Ropers and V. M. Kalscheuer, 1996 The *MAS* protooncogene is not imprinted in humans. Genomics **35:** 380–382.

Saitoh, S., N. Harada, Y. Jinno, K. Hashimoto, K. Imaizumi *et al.*, 1994 Molecular and clinical-study of 61 angelman-syndrome patients. Am. J. Med. Genet. **52:** 158–163.

Schweifer, N., P. J. M. Valk, R. Delwel, R. Cox, F. Francis *et al.*, 1997 Characterization of the C3 YAC contig from proximal mouse chromosome 17 and analysis of allelic expression of genes flanking the imprinted *Igf2r* gene. Genomics **43:** 285–297.

Sharp, P. M., and W. H. Li, 1989 On the rate of DNA sequence evolution in *Drosophila*. J. Mol. Evol. **28:** 398–402.

Shields, D. C., P. M. Sharp, D. G. Higgins and F. Wright, 1988 Silent sites in Drosophila genes are not neutral—evidence of selection among synonymous codons. Mol. Biol. Evol. **5:** 704–716.

Shimmin, L. C., B. H. J. Chang, D. Hewett-Emmett and W. H. Li, 1993 Potential problems in estimating the male-to-female mutation rate ratio from DNA sequence data. J. Mol. Evol. **37:** 160–166.

Shimmin, L. C., B. H. J. Chang and W. H. Li, 1994 Contrasting rates of nucleotide substitution in the X-linked and Y-linked zinc-finger genes. J. Mol. Evol. **39:** 569–578.

Smith, N. G. C., and L. D. Hurst, 1998a Molecular evolution of an imprinted gene: repeatability of patterns of evolution within the mammalian insulin-like growth factor type II receptor. Genetics **150:** 823–833.

Smith, N. G. C., and L. D. Hurst, 1998b Sensitivity of patterns of molecular evolution to alterations in methodology: a critique of Hughes and Yeager. J. Mol. Evol. **47:** 493–500.

Sokal, R. R., and F. J. Rohlf, 1995 *Biometry*. W. H. Freeman and Company, New York.

Tamura, K., 1992 Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases. Mol. Biol. Evol. **10:** 512–526.

Terada, N., K. Kuma and T. Miyata, 1997 Verification of male-driven molecular evolution. Genes Genet. Syst. **72:** 428.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 ClustalW—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Villar, A. J., and R. A. Pedersen, 1994 1994 parental imprinting of the *Mas* protooncogene in mouse. Nat. Genet. **8:** 373–379.

Wentworth, B. M., I. M. Schaefer, L. Villa-Komaroff and J. M. Chirgwin, 1986 Characterization of the two nonallelic genes encoding mouse preproinsulin. J. Mol. Evol. **23:** 305–312.

Wolfe, K. H., and P. M. Sharp, 1993 Mammalian gene evolution—nucleotide-sequence divergence between mouse and rat. J. Mol. Evol. **37:** 441–456.

Wright, F., 1990 The effective number of codons used in a gene. Gene **87:** 23–29.

Yang, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13:** 555–556.

Yang, W., K. Chen, N. C. Lan, T. K. Gallaher and J. C. Shih, 1992 Gene structure and expression of the mouse 5-HT2 receptor. J. Neurosci. Res. **33:** 196–204.

Communicating editor: R. R. Hudson