Patterns of Transitional Mutation Biases Within and Among Mammalian Genomes

Michael S. Rosenberg, Sankar Subramanian, and Sudhir Kumar

Center for Evolutionary Functional Genomics and Department of Biology, Arizona State University

Significant transition/transversion mutation bias is a well-appreciated aspect of mammalian nuclear genomes; however, patterns of bias among genes within a genome and among species remain largely uncharacterized. Understanding these patterns is important for understanding similarities and differences in mutational patterns among genomes and genomic regions. Therefore, we have conducted an analysis of 7,587 pairs of sequences of 4,347 mammalian protein-coding genes from seven species (human, mouse, rat, cow, sheep, pig, and macaque) and from the introns of 51 gene pairs and multiple intergenic regions (37 kbp, 52 kbp and 65 kbp) from the human, chimpanzee, and baboon genomes. Our analyses show that genes and regions with widely varying base composition exhibit uniformity of transition mutation rate both within and among mammalian lineages, as long as the transitional mutations caused by CpG hypermutability are excluded. The estimates show no relationship to potential intrachromosomal or interchromosomal effects. This uniformity points to similarity in point mutation processes in genomic regions with substantially different GC-content biases.

Introduction

A fundamental aspect of DNA point mutation is the observation that transitional nucleotide changes (purine to purine or pyrimidine to pyrimidine) occur with greater frequency than transversional changes (purine to pyrimidine or vice versa). This bias is due primarily to the biochemical structure of the nucleotide bases and the chemical properties of complementary base pairing (Topal and Fresco 1976*a*, 1976*b*). Reliable estimates of the bias are important for understanding the mechanism of nucleotide substitution, reconstruction of phylogenies, estimation of distances among sequences, and assessing the mode and strength of natural selection (Wakeley 1996). In particular, significant differences in transition bias may point to different mutational mechanisms among genomic regions. Furthermore, the relative abundance of transitional and transversional mutations has important consequences in epidemiological research as each class is associated with different diseases (Hollstein et al. 1991; Blanck, Tolbert, and Hoppin 1999; Martínez-Arias et al. 2001).

Often, transition/transversion mutation rate biases are assumed to vary considerably among genes and genomic segments (Jukes 1987; Wakeley 1996; Yang and Yoder 1999). This perception is commonly based on individual gene analysis, as the extent of transition bias among genomic regions and among species has not been characterized. Therefore, we have examined transition bias in the introns of 51 gene pairs and three orthologous intergenic regions from human, chimpanzee, and baboon, and 4,347 protein-coding genes (11,428 sequences) from seven mammal species. To use as much data as possible, we have taken a pairwise species approach to estimating the instantaneous transition rate bias (κ) for genes, genomic regions, and species.

Key words: transition/transversion bias, substitution patterns, mammalian genomes.

E-mail: s.kumar@asu.edu.

Mol. Biol. Evol. 20(6):988–993. 2003
DOI: 10.1093/molbev/msg113

Materials and Methods

Sequence Data

Phylogenetic trees of 8,627 nuclear vertebrate gene families in the Hovergen database release 36 (Duret, Mouchiroud, and Gouy 1994) were constructed from amino acid sequence alignments using the Neighbor-Joining method in MEGA2 (Kumar et al. 2001). The cDNA alignments for sequence sets were then generated, using amino acid sequence alignments as guides. Neighbor-Joining trees were manually inspected to identify orthologous sequence sets. We enforced strict orthology definitions by considering sequences to be orthologous only if no gene duplication events were detected since their divergence from the most recent common ancestor. Analyses were restricted to fourfold-degenerate sites of sequence pairs in order to minimize any effect of selection on transition bias estimation. A stringent approach to identifying fourfold-degenerate sites was taken by selecting only those sites that were fourfold-degenerate in both sequences. Genes were extracted for seven species (human, mouse, rat, cow, sheep, pig, and macaque) and the following species pairs were assembled: mouse–rat, cow–sheep, human–macaque, cow–pig, human–mouse, human–cow, and mouse–cow. These species pairs were chosen solely on the basis of gene availability; sequences for at least 75 homologous genes were present in our data set for each pair. Sequence pairs showing no transitional or no transversional differences were dropped, because further calculations become undefined.

Variation in coding and noncoding regions were examined through analysis of fully concatenated exon and concatenated intron sequences from 51 gene pairs from human–chimpanzee (26 pairs) and human–baboon (25 pairs). Orthologous intergenic regions from human, chimpanzee, and baboon were culled from five large contigs (human: AC002066, AC002080; chimpanzee: AC087253, AC087512; baboon: AC084730). Repeat elements were identified using RepeatMasker (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker) and removed prior to sequence alignment. Using human genome annotations, all coding genes and 5 kb of flanking region from the upstream and downstream regions around each

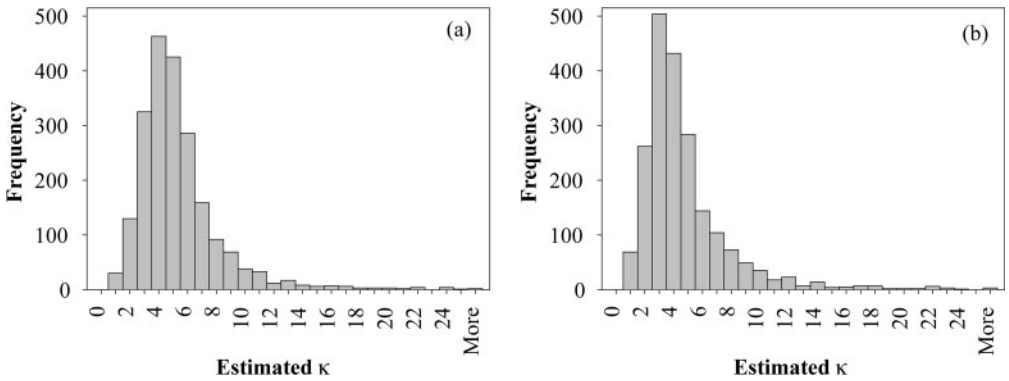

FIG. 1.—Frequency histograms of transition bias (κ) in mouse–rat gene comparisons (*a*) including all fourfold-degenerate sites (mean = 4.26, median = 4.24, variance = 9.65) and (*b*) after the removal of fourfold-degenerate sites involved in CpG dinucleotides (mean = 3.45, median = 3.40, variance = 10.21).

gene (to exclude potential regulatory regions) were removed from the contigs.

Estimating Transition Bias

There are numerous approaches to estimating the transition bias, κ (Kimura 1980; Wakeley 1994, 1996; Pollock and Goldstein 1995; Ina 1998; Yang and Yoder 1999). These approaches fall into two basic categories: those based on paired sequences and those based on greater numbers of sequences and phylogenetic trees. We use a paired sequences approach because it allows us to use substantially more genes and sequences then would be available if we restricted ourselves to phylogeny-based estimation. In any case, efficient estimators of κ can be constructed in pairwise sequence analysis under the Hasegawa, Kishino, and Yano (HKY) (1985) model. Under this model, the frequency of change is dependent not only on transition and transversion substitution rates (α and β, respectively), but also on equilibrium nucleotide frequencies (π). The instantaneous rate of change of a nucleotide i to nucleotide j is $\alpha\pi_j$ for transitional substitutions and $\beta\pi_j$ for transversional substitutions. The transition bias $\kappa = \alpha/\beta$. This is the instantaneous ratio of transitional change to transversional change and differs from the actual expected number of transitions and transversions. For the HKY model, the ratio of the expected number of transitions ($\hat{s}$) and transversions ($\hat{v}$) is

$$\frac{\hat{s}}{\hat{v}} = \frac{2\alpha(\pi_A\pi_G + \pi_C\pi_T)}{2\beta(\pi_A\pi_C + \pi_A\pi_T + \pi_G\pi_C + \pi_G\pi_T)} = \kappa\frac{(\pi_A\pi_G + \pi_C\pi_T)}{(\pi_A + \pi_G)(\pi_C + \pi_T)} = \kappa\frac{(\pi_A\pi_G + \pi_C\pi_T)}{\pi_R\pi_Y}, \quad (1)$$

where π_R and π_Y are the frequencies of purines and pyrimidines, respectively. For each pair of sequences, we estimated $\hat{s}$ and $\hat{v}$ using the Tamura-Nei (1993) method to correct for multiple substitutions at a site, taking into account the base composition bias and the transition/transversion rate bias.

For a set of genes from a specific species pair, we calculated an overall estimate of κ as the weighted average of the natural logarithms of the individual κ estimates, where the weight for each estimate is equal to the sequence length (n) of the associated sequences, i.e.,

$$\kappa = e^{\frac{\sum n_i \ln \kappa_i}{\sum n_i}} \quad (2)$$

This average is more appropriate than the standard arithmetic mean because κ is a ratio, and because longer sequences are expected to give more accurate estimates than shorter sequences.

Spatial Patterns of Transition Bias

The spatial pattern of κ estimates in the human genome was determined from a correlogram (Sokal and Oden 1978; Cliff and Ord 1981) of 3,023 human–mouse homologous genes. Chromosomal locations for each gene were as in the human genome. Physical distance between two genes was measured as the number of nucleotides between the end of one gene and the beginning of the next (overlapping genes were given distances of zero). The autocorrelation coefficient, Moran's I (Moran 1950), was determined for pairs of genes located on the same chromosome within specific distances of each other; successive distance classes represented 500-kb windows. Spatial analyses were performed using PASSAGE (Rosenberg 2001).

Results

Transition bias estimates based on paired sequences from closely related species (species within the same suborder) shows a broad range (fig. 1*a*), explaining why many authors believe that the true value of κ varies substantially among genes (e.g., Jukes 1987; Wakeley 1996; Yang and Yoder 1999). However, while a few genes do show rather high transition bias, most of the κ estimates cluster tightly around the median value. In fact, if we compare κ estimates for the same gene ($n = 38$) obtained from independent species pairs from rodent and artiodactyl lineages, no correspondence ($r = 0.21$, $P = 0.22$) is seen between the resulting values (fig. 2). That is, there is essentially no correlation between transition bias estimates derived from identical genes for mouse–rat and cow–sheep sequences.

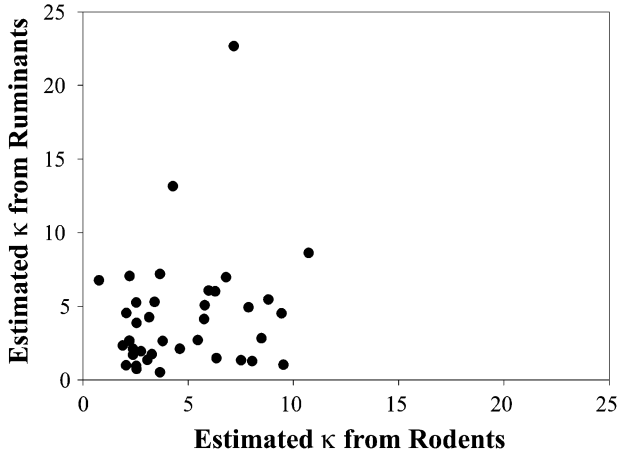

FIG. 2.—Plot of the estimated transition/transversion rate ratio for orthologous genes ($n = 38$) from ruminants (cow–sheep gene pairs) and rodents (mouse–rat gene pairs). The estimates show a weak (non-significant) correlation, $r = 0.21$ (after removal of the two outliers with $\kappa > 10$, $r = 0.19$).

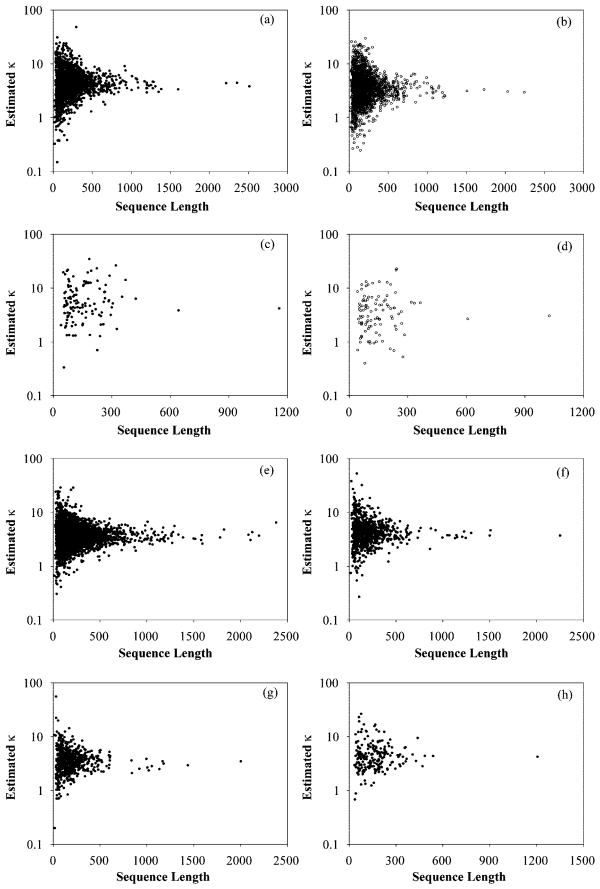

FIG. 3.—Plots of estimated transition/transversion rate ratio (κ) versus sequence length for species pairs with at least 100 genes. Mouse–rat comparison (*a*) including all fourfold-degenerate sites (2,128 genes) and (*b*) after removal of fourfold-degenerate CpG sites (2,060). Cow–sheep comparison (*c*) including all fourfold-degenerate sites (120 genes) and (*d*) after the removal of fourfold-degenerate CpG sites (99 genes). (*e*) Human–mouse (3,712 genes). (*f*) Human–cow (844 genes). (*g*) Cow–mouse (607 genes). (*h*) Cow–pig, (205 genes).

Instead, there is a striking relationship between estimated κ and length of the sequence used in the estimation (fig. 3*a*). This result is not dependent on a specific set of genes or taxa because the pattern is found in six different data sets (fig. 3*a*, *c*, and *e–h*). The exhibited funnel pattern (Light and Pillemer 1984) shows that as the number of sites used to estimate κ increases, the estimate converges to the mean value. This means that the transition bias for genes within a genome may be very similar, and that the observed variance in κ among genes is a result of the large estimation error associated with short sequence length. This should be the null hypothesis as there is no a priori reason to assume that fundamental aspects of mutation pattern should vary within a single genome.

The overall κ estimate from closely related species pairs are shown in table 1. They vary substantially from 4.0 to 5.5, even though they show similar funnel patterns (see fig. 3*a* and *c*). One complication with these estimates is the hypermutability of CpG dinucleotides in mammals (Bird 1980). Methylation of the cytosine base is followed by spontaneous deamination of methyl-C to give rise to a thymine residue. This transitional substitution occurs approximately 10 times faster than other point mutations in mammals (Krawczak, Ball, and Cooper 1998; Giannelli, Anagnostopoulos, and Green 1999) and has a profound effect on estimates of transition bias, as CpG content varies among genes. Not only do CpG sites add significantly

Table 1
Estimated Transition/Transversion Rate Ratio for Species Pairs

	All Neutral Sites		CpG Sites Removed	
	No. of Genes/bp	κ	No. of Genes/bp	κ
Protein-Coding Sequences				
Mouse–rat (rodents)	2,128	4.26	2,060	3.45
Cow–sheep (ruminants)	120	5.40	99	3.37
Human–macaque (primates)	77	5.54	65	3.68
Introns				
Human–chimp	26	4.73	26	3.73
Human–baboon	25	4.63	25	4.14
Intergenic Regions				
AC002066 vs. AC087512 (human–chimp)	65,015 bp	4.31	64,194 bp	3.72
AC002080 vs. AC087253 (human–chimp)	52,627 bp	4.01	51,805 bp	3.58
AC002066 vs. AC084730 (human–baboon)	37,937 bp	4.38	37,268 bp	4.07

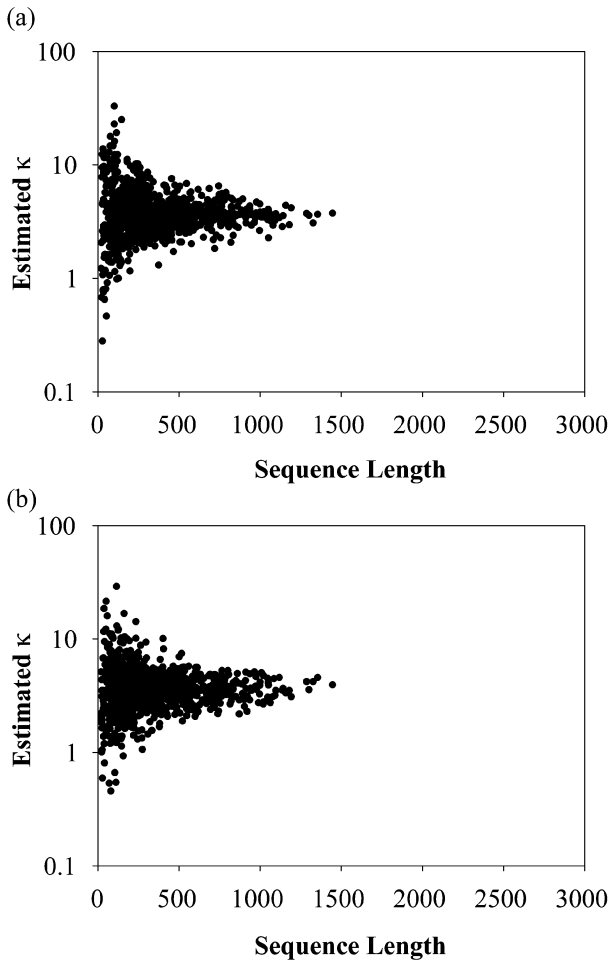

FIG. 4.—Plot of the estimated transition/transversion rate ratio versus sequence length for the simulated mouse–rat gene pairs. (*a*) fixed substitution rate (mean = 3.65, variance = 5.57); (*b*) variable substitution rate (mean = 3.59, variance = 4.36).

more transitional changes than other dinucleotides, the differential rate of mutation violates the HKY substitution model used to estimate κ. Not surprisingly, there can be a strong correlation between CpG content of a gene and the κ estimate (human–macaque, $n = 77$, $r = 0.34$, $P = 0.003$; cow–sheep, $n = 120$, $r = 0.25$, $P = 0.007$; mouse–rat, $n = 2128$, $r = 0.08$, $P < 0.001$). Removal of all fourfold-degenerate sites involved in CpG dinucleotides as well as CpG dinucleotides in intergenic sequences leads to a baseline estimate of the transition bias, excluding the hypermutable sites. These new κ estimates (table 1) are very similar among mammalian lineages (ranging from 3.4 to 3.7) and between intergenic regions, introns, and the fourfold-degenerate sites of exons. The exception is the slightly higher value of 4.1 found for the human–baboon comparison (in both introns and the intergenic region).

We conducted computer simulations to examine if the funnel pattern of κ estimates from different genes can be produced under the assumption of a single true κ among genes; a secondary purpose of the simulations was to examine whether equations (1) and (2) could be used to provide unbiased estimates of κ. We simulated the evolution of 1,000 genes across an 18-taxa model tree of vertebrate evolution (Rosenberg and Kumar 2001). For each gene, the substitution rate was set to 2.2×10^{-3} substitutions per site per Myr, the observed rate at neutral sites (Kumar and Subramanian 2002), and κ was set to 3.6 (approximately the observed average from this study; table 1). Nucleotide frequencies were chosen randomly for each gene, such that the G + C content was between 40% and 80% and the frequencies of both G and C (and A and T) were within 2% to 8% of each other (following observations from real data). To mimic the distribution seen in the real data, sequence lengths for each simulated gene were chosen randomly from a gamma distribution with a gamma parameter of 1.5 (resultant range: 21 to 1,448 sites). A second simulation was conducted where substitution rates were chosen from a normal distribution with observed (Kumar and Subramanian 2002) mean (2.2×10^{-3}) and variance (7.1×10^{-7}).

The average κ's estimated from species pairs of the simulated data were quite close to the true simulated value (e.g., mouse–rat, $\kappa = 3.65$; cow–pig, $\kappa = 3.63$) when we use equations (1) and (2). Use of the arithmetic mean rather than equation (2) leads to overestimation of κ (mouse–rat, $\kappa = 4.01$; cow–pig, $\kappa = 3.89$). Furthermore, figure 4 shows the simulated distribution of κ estimates for mouse–rat paired genes (we chose these species as being the most closely related species with a large number of observed gene pairs). Both the observed (fig. 3*b*) and simulated (fig. 4) distributions have similar ranges and shapes. The spread (variance) of the observed estimates is larger than that of the simulated data, indicating that the relatively simplistic nature of the simulation cannot capture the full stochastic variation present in the observed data.

We used two approaches to examine whether the excess variance could be due to extreme local effects on transition bias. First, for 51 gene pairs (human–chimp and human–baboon) we compared κ estimates derived from exons and introns of the same genes. Because the number of fourfold-degenerate sites in the exons was quite small (median = 188 sites), these κ estimates are expected to have large stochastic variation, whereas the concatenated introns were quite long (median = 9,147 sites) and should produce κ estimates with much less variance. Before removal of CpG sites, there was a moderate (although nonsignificant) correlation between κ estimates from introns and exons of the same genes ($n = 31$, $r = 0.29$, $P = 0.12$); after removal of the CpG sites, this correlation disappeared ($n = 24$, $r = 0.02$, $P = 0.91$). This result implies no local chromosomal effect on κ estimation beyond the correlated effects of GC and CpG content.

The second approach was to calculate a correlogram for κ estimates for 3,023 human genes (fig. 5). This analysis is motivated by the idea that if there is regional genomic variation in the fundamental mutation pattern (Matassi, Sharpa, and Gautier 1999; Williams and Hurst 2000; Lercher, Williams, and Hurst 2001), genes located near each other on a chromosome might have more similar transition biases than those located farther apart. This correlogram reveals no spatial patterning; genes located near each other on a chromosome show variation similar to that among those located farther apart. Furthermore, the average κ was essentially the same whether taken across

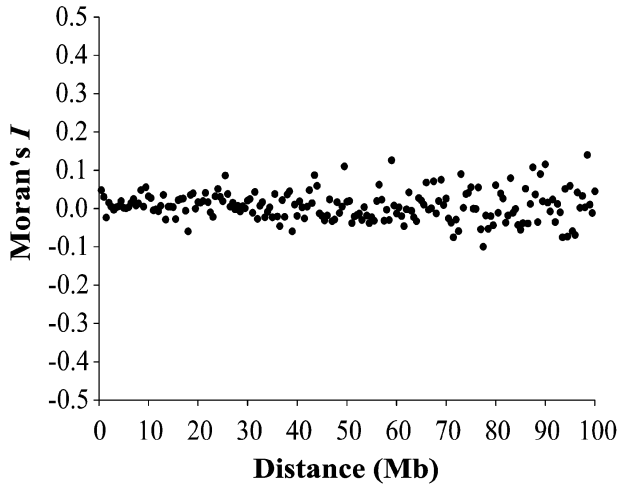

FIG. 5.—Correlogram of κ estimates for 3,023 human genes with known chromosomal locations. Distance classes represent 500-kb windows. The degree of autocorrelation is quite minimal at all distances (rarely exceeding 0.1).

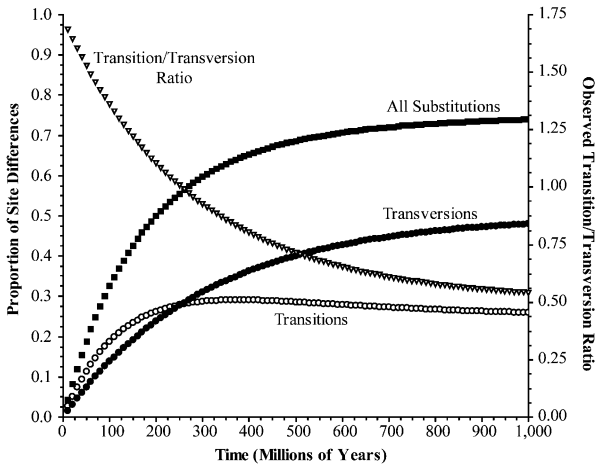

FIG. 6.—Plot of expected transitional differences, transversional differences, all differences, and transition/transversion difference ratio versus time for two sequences (with equilibrium nucleotide frequencies of $\pi_A = 0.2$, $\pi_C = 0.3$, $\pi_G = 0.3$, and $\pi_T = 0.2$) evolving under an HKY model with the observed mean rate of substitution (Kumar and Subramanian 2002) and instantaneous transition/transversion rate ratio ($\kappa = 3.6$). The left axis indicates the proportion of sites expected to differ by a specific substitution type (transitional, transversional, or any), and the right axis indicates the expected observed ratio of transitions to transversions. This latter ratio differs from κ; it is simply the ratio of observed transitional and transversional differences at a specified time.

the entire genome or determined for individual chromosomes (results not shown).

Discussion

Two critical factors distinguish this study from previous work (Jukes 1987; Wakeley 1996; Yang and Yoder 1999) which found large variation in transition bias among genes and lineages. By using a large numbers of genes, we were able to discover the relationship between κ estimation and sequence length (fig. 3), heretofore hidden under the presumption of true variation in mutational transition bias among genes. In addition, the importance of CpG dinucleotides on transition bias in estimation has previously been under-appreciated, particularly with respect to interspecific comparisons. These sites mutate in a fundamentally different way than other positions, and their removal from the comparison reveals much higher uniformity of transition bias among different mammalian lineages. Generally, CpG sites can only be removed from closely related species, because the footprints of CpG mutations are erased with higher frequency as the evolutionary divergence increases. This means that the effect of CpG on κ estimation decreases as divergence increases; the increase in transition number owing to the hypermutability of CpG sites will be partially masked by secondary substitutions at these sites. While one cannot completely ignore the potential effects of CpG hypermutability on transition bias estimation from distantly related species pairs, κ estimation from these distantly related species pairs are expected to yield values similar to those of closely related pairs after the removal of CpG sites (table 1). This is indeed the case: human–mouse (3,712 genes), $\kappa = 3.66$; human–cow (844 genes), $\kappa = 4.06$; mouse–cow (607 genes), $\kappa = 3.47$; cow–pig (205 genes), $\kappa = 4.32$.

Other factors may have large effects on transition bias. Rate variation among sites can complicate the estimate of κ (Wakeley 1994); however, use of only fourfold-degenerate (neutral) sites should yield a subset of sites with relatively constant rates of substitution (Kumar and Subramanian 2002). Using 5 taxa and 14 genes from Springer et al. (2003), we found the gamma parameter (using ML) for fourfold-degenerate sites to be approximately 2. Use of this value in the computation of κ produces essentially the same values already reported. Codon usage bias will often have large effects on substitution pattern and transition bias because all mutations at fourfold-degenerate sites may not be neutral. In mammals, however, codon usage bias is extremely weak (Akashi 2001; Urrutia and Hurst 2001) and is likely to have no effect on our estimates.

The results we have shown are for the neutral mutation transition bias. Sites under selection pressure are expected to show widely varying transition biases. Within coding sequences, transitional changes are often synonymous when transversional changes are not; furthermore, when both transitions and transversions lead to a change in the protein sequence, the transitional change is often less severe with respect to the chemical properties of the original and mutant amino acids (Grantham 1974; Zhang 2000). Furthermore, regional effects on transition bias cannot be completely discounted. We have already shown the effect of CpG sites (and thus, the correlated effect of GC content) on transition bias. There may easily be additional local variation as yet uncharacterized. More extensive analyses of mutational patterns among the genes of closely related species are needed, and it is important to exercise caution in using the universal average we report for analyses of individual genes.

Knowledge of the mutational transition/transversion rate bias allows a general prediction of time to saturation of substitutions at fourfold-degenerate sites (fig. 6). Given these observed mutational parameters, transversions become more common than transitions after 250 Myr, the

time about which transitions become saturated (at ~25% of sites). Transversions become saturated much more slowly, asymptotically beginning to reach 50% after about 750 Myr. We find that the observed number of transitional substitutions accumulates approximately linearly for about 100 Myr, while the transversional substitutions accumulate linearly for about 250 Myr.

Significant similarity of mutational transition/transversion rate bias has important implications for understanding the point mutational mechanism and genome evolution. Furthermore, the consistency of κ among genes and lineages would allow us to more easily model the rate of duplicate gene silencing (Lynch and Conery 2000), employ the same, but sophisticated, model in evolutionary inference when analyzing large numbers of genes or large genomic regions, and begin to tease apart the contribution of selection in maintaining genomic structural attributes, including the isochores.

Acknowledgments

This research was supported by grants from the National Science Foundation, the National Institutes of Health, and the Burroughs Wellcome Fund to S.K. Thanks to T. Dowling and K. Tamura for comments on early versions of this manuscript, and to B. Ramanujam for computational assistance.

Literature Cited

Akashi, H. 2001. Gene expression and molecular evolution. Curr. Opin. Genet. Dev. **11**:660–666.

Bird, A. P. 1980. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. **8**:1499–1504.

Blanck, H. M., P. E. Tolbert, and J. A. Hoppin. 1999. Patterns of genetic alterations in pancreatic cancer: a pooled analysis. Environ. Mol. Mutagen. **33**:111–122.

Cliff, A. D., and J. K. Ord. 1981. Spatial Processes. Pion, London.

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

Giannelli, F., T. Anagnostopoulos, and P. M. Green. 1999. Mutation rates in humans. II. Sporadic mutation-specific rates and rate of detrimental human mutations inferred from hemophilia B. Am. J. Hum. Gen. **65**:1580–1587.

Grantham, R. 1974. Amino acid difference formula to help explain protein evolution. Science **185**:862–864.

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**:160–174.

Hollstein, M., D. Sidranksy, B. Vogelstein, and C. C. Harris. 1991. p53 mutations in human cancers. Science **253**:49–53.

Ina, Y. 1998. Estimation of the transition/transversion ratio. J. Mol. Evol. **46**:521–533.

Jukes, T. H. 1987. Transitions, transversions, and the molecular evolutionary clock. J. Mol. Evol. **26**:87–98.

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.

Krawczak, M., E. V. Ball, and D. N. Cooper. 1998. Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. Am. J. Hum. Gen. **63**:474–488.

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

Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17**:1244–1245.

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

Light, R. J., and D. B. Pillemer. 1984. Summing up: the science of reviewing research. Harvard University Press, Cambridge.

Lynch, M., and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. Science **290**:1151–1155.

Martínez-Arias, R., E. Mateu, J. Bertranpetit, and F. Calafell. 2001. Profiles of accepted mutation: From neutrality in a pseudogene to disease-causing mutation on its homologous gene. Hum. Genet. **109**:7–10.

Matassi, G., P. M. Sharpa, and C. Gautier. 1999. Chromosomal location effects on gene sequence evolution in mammals. Curr. Biol. **9**:786–791.

Moran, P. A. P. 1950. Notes on continuous stochastic phenomena. Biometrika **37**:17–23.

Pollock, D. D., and D. B. Goldstein. 1995. A comparison of two methods for constructing evolutionary distances from a weighted contribution of transition and transversion differences. Mol. Biol. Evol. **12**:713–717.

Rosenberg, M. S. 2001. PASSAGE: pattern analysis, spatial statistics, and geographic exegesis. Version 1.0. Department of Biology, Arizona State University, Tempe, AZ.

Rosenberg, M. S., and S. Kumar. 2001. Incomplete taxon sampling is not a problem for phylogenetic inference. Proc. Natl. Acad. Sci. USA **98**:10751–10756.

Sokal, R. R., and N. L. Oden. 1978. Spatial autocorrelation in biology 1. Methodology. Biol. J. Linn. Soc. **10**:199–228.

Springer, M. S., W. J. Murphy, E. Eizirik, and S. J. O'Brien. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. Proc. Natl. Acad. Sci. USA **100**:1056–1061.

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

Topal, M. D., and J. R. Fresco. 1976*a*. Base pairing and fidelity in codon-anticodon interaction. Nature **263**:289–293.

———. 1976*b*. Complementary base pairing and the origin of substitution mutations. Nature **263**:285–289.

Urrutia, A. O., and L. D. Hurst. 2001. Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. Genetics **159**:1191–1199.

Wakeley, J. 1994. Substitution-rate variation among sites and the estimation of transition bias. Mol. Biol. Evol. **11**:436–442.

———. 1996. The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. Trends Ecol. Evol. **11**:158–163.

Williams, E. J. B., and L. D. Hurst. 2000. The proteins of linked genes evolve at similar rates. Nature **407**:900–903.

Yang, Z., and A. D. Yoder. 1999. Estimation of the transition/transversion rate bias and species sampling. J. Mol. Evol. **48**:274–283.

Zhang, J. 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. J. Mol. Evol. **50**:56–68.

Mark Springer, Associate Editor

Accepted March 1, 2003