Copyright © 1998 by the Genetics Society of America

Molecular Evolution of an Imprinted Gene: Repeatability of Patterns of Evolution Within the Mammalian Insulin-Like Growth Factor Type II Receptor

Nick G. C. Smith and Laurence D. Hurst

Centre for Mathematical Biology, School of Biology and Biochemistry, University of Bath, Bath BA2 7AY, United Kingdom

Manuscript received March 2, 1998
Accepted for publication June 30, 1998

ABSTRACT

The repeatability of patterns of variation in K_a/K_s and K_s is expected if such patterns are the result of deterministic forces. We have contrasted the molecular evolution of the mammalian insulin-like growth factor type II receptor (*Igf2r*) in the mouse-rat comparison with that in the human-cow comparison. In so doing, we investigate explanations for both the evolution of genomic imprinting and for K_s variation (and hence putatively for mutation rate evolution). Previous analysis of *Igf2r*, in the mouse-rat comparison, found K_a/K_s patterns that were suggested to be contrary to those expected under the conflict theory of imprinting. We find that K_a/K_s variation is repeatable and hence confirm these patterns. However, we also find that the molecular evolution of *Igf2r* signal sequences suggests that positive selection, and hence conflict, may be affecting this region. The variation in K_s across *Igf2r* is also repeatable. To the best of our knowledge this is the first demonstration of such repeatability. We consider three explanations for the variation in K_s across the gene: (1) that it is the result of mutational biases, (2) that it is the result of selection on the mutation rate, and (3) that it is the product of selection on codon usage. Explanations 2 and 3 predict a K_a–K_s correlation, which is not found. Explanation 3 also predicts a negative correlation between codon bias and K_s, which is also not found. However, in support of explanation 1 we do find that in rodents the rate of silent C $\Rightarrow$ T mutations at CpG sites does covary with K_s, suggesting that methylation-induced mutational patterns can explain some of the variation in K_s. We find evidence to suggest that this CpG effect is due to both variation in CpG density, and to variation in the frequency with which CpGs mutate. Interestingly, however, a GC4 analysis shows no covariance with K_s, suggesting that to eliminate methyl-associated effects CpG rates themselves must be analyzed. These results suggest that, in contrast to previous studies of intragenic variation, K_s patterns are not simply caused by the same forces responsible for K_a/K_s correlations.

IN mammals there exists much variation in synonymous substitution rates (K_s) and the ratio of nonsynonymous to synonymous substitution rates (K_a/K_s), both within (Alvarez-Valin *et al.* 1998) and between (Wolfe and Sharp 1993) genes. There are two alternative explanations for such variation: that it is simply a consequence of chance (*e.g.*, some genes, or parts of them, might just happen to mutate faster than others) or that it is due to deterministic effects. Possible deterministic forces include selection and repeatable mutational biases. Analysis of the repeatability of patterns of molecular evolution in independent comparisons is one means to distinguish between stochastic and deterministic explanations (Mouchiroud *et al.* 1995). Here we ask about the repeatability of patterns of intragene variation in K_a/K_s and K_s. To the best of our knowledge this is the first example where repeatability of intragene variation in both of these parameters has been examined. The gene that we examine, the insulin-like growth factor type II receptor (*Igf2r*), is of interest as it is one of a small class of mammalian genes that are imprinted. It has also recently been implicated in the determination of IQ (Chorney *et al.* 1998).

Genomically imprinted genes are those for which expression is dependent upon the sex of the parent from which they are derived (Efstratiadis 1994). Well-described examples include murine insulin-like growth factor II (*Igf2*), a growth factor expressed from the paternally derived genome in fetuses, and the murine insulin-like growth factor type II receptor (*Igf2r*) expressed from the maternally derived genome. In eutherians and marsupials, but not in birds (Zhou *et al.* 1995), the protein product of this gene binds to Igf2 and takes it to the lysosome for digestion. The gene is also known as the cation-independent mannose-6-phosphate (M6P) receptor as it is also known to bind M6P (Kornfeld 1992). The latent complex of transforming growth factor beta (TGFβ) also binds to the Igf2r protein via the M6P binding domains (Dennis and Rifkin 1991). Additionally the Igf2r protein has the ability to bind retinoic acid (Kang *et al.* 1998). The binding activities of Igf2

Corresponding author: Nick Smith, Centre for Mathematical Biology, School of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK. E-mail: n.smith@bath.ac.uk

and M6P are known to occur at different sites within the Igf2r protein. In rats *Igf2r* is known to be imprinted in the liver (Mills *et al.* 1998) and in humans *IGF2R* appears to be imprinted in some but not all individuals (Ogawa *et al.* 1993; Xu *et al.* 1993). We are unaware of data pertinent to the imprinting status of the gene in cow. However, from our knowledge that the mouse, rat, and human *IGF2R* genes are now imprinted, we can be sure that some of the changes revealed by human-cow and mouse-rat comparisons of *Igf2r* will have occurred in an imprinted gene.

Aside from allowing us to ask whether patterns of K_a/K_s and of K_s within genes are repeatable and hence deterministically based, the patterns of molecular evolution in this gene are of interest for two reasons. First, the pattern of variation in K_s across the gene may shed light on the evolution of mutation rates or of codon usage patterns. Second, the variation in K_a/K_s is potentially informative as to the evolution of imprinting. One theory for the evolution of imprinting would find discriminating support from evidence for antagonistic coevolution between imprinted genes (see McVean and Hurst 1997b; Trivers and Burt 1998; but see also Haig 1997).

K_s variation, mutation rate evolution, and imprinted genes: We consider four hypotheses to explain variation in K_s across *Igf2r*.

Hypothesis 1: K_s variation may be caused by stochastic forces, in which case K_s will not be repeatable. The remaining three hypotheses suppose that K_s variation is the result of deterministic forces that would cause K_s repeatability.

Hypothesis 2: If silent sites are considered neutral then K_s is an unbiased estimator of the mutation rate (Kimura 1983), and so K_s variation may be the result of deterministic mutational processes, with mutational biases dependent on perhaps sequence composition or regional effects. The two final hypotheses, (3) and (4), both suggest that K_s variation is molded by selection, but differ with respect to the character that selection is acting upon.

Hypothesis 3: Decreasing mutation rates is expensive both in time and energy, while the benefits of a lower mutation rate depend on the functional importance and mutational sensitivity of the region concerned. Such reasoning has been used to explain the constant per genome mutation rate across many unicellular organisms (Drake 1991). If this trade-off theory of mutation rate evolution is correct, and if selection is strong enough, one might imagine that selection would favor imprinted genes to have low mutation rates. This prediction derives from imprinted genes having haploid expression and hence recessive mutations in imprinted genes being unmasked. Hence it may be optimal for an organism to invest more into reducing the mutation rate of imprinted genes below the mutation rate of comparable nonimprinted ones. It should, however, be noted that selection on the modifier is likely to be very weak.

From the same logic, one might expect that the most important parts of an imprinted gene should be provided with especially low mutation rates. Thus if silent sites in the mammalian *Igf2r* gene are neutral, then the variation in K_s (and hence mutation rate) across the gene may be the result of selection acting to optimize mutation rates. We term this explanation of K_s variation the "selected mutation rate" hypothesis.

Hypothesis 4: Alvarez-Varin *et al.* (1998) have found intragenic correlations between synonymous and nonsynonymous substitution rates in mammalian genes. They conclude that these correlations might well be due to common selective constraints, such as selection for translational accuracy, between synonymous and nonsynonymous sites. This argument suggests that K_s in mammals is not the mutation rate (as often assumed), but instead reflects the local strength of selection. We term this explanation of K_s variation the "common constraints" hypothesis.

Just as for the selected mutation rate argument, the common constraints hypothesis would predict that imprinted genes, due to their haploid expression, would be under greater selective pressures, and thus should have lower K_s values than nonimprinted genes. The empirical data on this point are equivocal. Previous analyses have found *Igf2* to have one of the lowest K_s values in the rodent genome (McVean and Hurst 1997b). Imprinted genes as a class have a K_s (K_s = 19.87, N = 7) between that of autosomes (K_s = 22.9, N = 238) and *X*-linked genes (K_s = 14.63, N = 33; McVean and Hurst 1997a,b). However, on nonparametric analysis the imprinted genes are found to be not significantly different from autosomal genes. Given sample size limitations, however, the issue cannot be considered fully resolved.

K_a/K_s variation and the evolution of genomic imprinting: Numerous theories have been proposed for the evolution of genomic imprinting (for a review see Hurst 1997). The "conflict" hypothesis (Moore and Haig 1991) proposes that imprinting is an intraindividual manifestation of classic parent-offspring conflict (Trivers 1974). According to this model, paternally derived genes are under selection to extract resources from mothers whereas maternally expressed genes are under selection to oppose this (for models see Mochizuki *et al.* 1996; Spencer *et al.* 1998). This hypothesis makes predictions about the growth effects of paternally and maternally expressed genes and it is presently unclear whether these predictions are borne out (Hurst and McVean 1997).

Given that maternal and paternal genes have antagonistic interests, the conflict theory is unique in predicting that an arms race may develop between maternally and paternally derived genes for the control of fetal growth demands. Such antagonistic coevolution may be mediated through changes in the structure of

the proteins concerned. Were one to find evidence for rapid evolution of imprinted genes (*i.e.*, a high K_a/K_s ratio) this, it has been argued, could then reasonably be taken as evidence favoring the conflict hypothesis (McVean and Hurst 1997b; see also Trivers and Burt 1998). Unfortunately this is not a falsifying prediction as the finding of slow evolution need not be considered evidence against the hypothesis (Haig 1997; McVean and Hurst 1997b) for a variety of reasons.

Comparable maternal-fetal conflict is a good candidate explanation for the rapid evolution of numerous genes, such as the placental lactogens (Wallis 1993), prolactin (Wallis 1981), the homeobox gene *Pem* (Sutton and Wilkinson 1997), growth hormones (Wallis 1994), the pregnancy-associated glycoproteins (Xie *et al.* 1997), and numerous others (see McVean and Hurst 1997b). Antagonistic coevolution is similarly the most likely explanation for the rapid changes seen in antigenic components of parasites and antigen recognition components of immune systems (for example see Hughes *et al.* 1994), and may be invoked to explain rapid changes in a gene putatively involved in sexual conflict (Tsaur and Wu 1997).

Perhaps surprisingly then, an analysis of the molecular evolution of seven imprinted genes revealed them to be no faster evolving at the protein level (controlling for K_s, and hence possibly for mutation rate, variation) than ordinary receptors and significantly slower evolving than immune system genes (McVean and Hurst 1997b). In the same study, an analysis of molecular evolution within *Igf2r*, compared between mouse and rat, showed that where Igf2r binds to Igf2 a low K_a/K_s ratio is found [indicating, contrary to initial expectations (McVean and Hurst 1997b), stabilizing rather than disruptive selection]. This analysis also shows two high peaks in K_a/K_s that were not commented on. One of these was at the signal sequence and one at the position at which Igf2r binds M6P.

Neither of these peaks in the pattern of protein evolution rates are expected. Signal sequences direct protein transport within the cell, and possess the conserved structure of a hydrophobic core flanked by a polar basic region and a hydrophilic region (Kendrew 1994). Since artificial signal sequence constructs lacking such features fail to perform their normal functions (Izard *et al.* 1996), one would expect strong amino acid conservation at signal sequences and hence low K_a/K_s values. Igf2r assists the transport of acid hydrolases to the lysosomes by binding the enzymes at their phosphomannosyl residues using the M6P binding site (Kornfeld 1992). Because the phosphomannosyl residues are high-affinity ligands there should be high sequence specificity at the M6P binding site, and hence strong amino acid conservation and low K_a/K_s values.

Here we ask whether the peaks and troughs in the K_a/K_s pattern across the gene revealed in the mouse-rat comparison are also observed in the human-cow comparison or whether they are statistical artifacts and hence not worth further investigation. That the mouse-rat patterns might be artifacts is possible since the sliding window used was small and hence the expected error in estimates of K_a/K_s per window is high.

MATERIALS AND METHODS

Extraction and analysis of genes: NCBI Entrez (http://www.ncbi.nlm.nih.gov/Entrez/) was used to search for complete coding mammalian *Igf2r* sequences. Four were obtained: human, cow, mouse, and rat (accession numbers J03528, J03527, U04710, and U59809, respectively). The orthologies of these sequences were confirmed using BLASTN (Altschul *et al.* 1990) and the HOVERGEN database (Duret *et al.* 1994).

Sequence manipulation and alignment was performed using programs in the Wisconsin package (Genetics Computer Group 1994). All four mammalian Igf2r protein-coding DNA sequences were aligned using the default DNA alignment parameters of the program PILEUP. The program GAPFRAME was used to move gaps to codon boundaries, and the resultant alignment was very similar to the protein alignment obtained using the default protein alignment parameters of PILEUP. All four gapped sequences were then cut into 24 orthologous nonoverlapping sections each of 300 bp, or 74 orthologous overlapping sections (300 bp every 100 bp). Synonymous and nonsynonymous nucleotide substitution rates (K_s and K_a) were estimated for all mouse-rat and human-cow orthologous pairs using the program KESTIM (Comeron 1995) with Kimura's (1980) two-parameter model for multiple hits correction. The mouse-rat and human-cow species pairs, which share no evolutionary history, were chosen to ensure independence and thus a valid test of repeatability.

A nonparametric statistical test of repeatability was used to avoid assumptions about possible distributions of rates of molecular evolution. For both comparisons each window was ranked across the gene. Then the two patterns of ranks across the gene were compared using a rank correlation test. Thus a statistic was obtained (P_{rank}) to describe the probability of the ranking patterns being so similar through chance alone.

Analysis of similarity: For all of the nonsynonymous substitutions in the two comparisons, we characterized the extent to which these changes were conservative. This provides an indication as to whether it is likely that the replacements seen might be owing to weak stabilizing selection (in which case conservative changes are likely) or due to positive selection (in which case nonconservative changes are more likely; *cf.* Tucker and Lundrigan 1995). To characterize changes as being conservative or not we employed PAM matrices. Such matrices use data on the difference between the expected number of replacements of one amino acid with another (from knowledge of relative positions in the genetic code), and data on the actual rate at which such substitutions occur. The difference in the two rates is thought to indicate some measure of physicochemical similarity between the two amino acids under consideration such that similar amino acids are more likely to be interchangeable.

Different PAM matrices are appropriate for genes of different levels of divergence. We obtained appropriate matrices from the Amino Acid Index Database (http://www.cbi.pku.edu.cn/srs5bin/cgi-bin/wgetz?-fun+Pagelibinfo+-info+AAINDEX/). As the mouse-rat protein comparison shows 93% identity, we employed the BENS940101 PAM6-8 matrix (Benner *et al.* 1994). The human-cow protein comparison shows 81% identity, so we used the BENS940102 PAM22-29 matrix (Benner *et al.* 1994). For each segment in the gene each amino acid

difference was assigned the relevant value from the appropriate PAM matrix. The mean value per amino acid change per segment could then be calculated. We can then compare the regions of interest with the evolution in the rest of the gene.

Signal sequence evolution: To ask whether the signal sequence of Igf2r has an unusual mode of evolution we compared its evolution with that of comparable signal sequences in other genes. We used NCBI Entrez (http://www.ncbi.nlm.nih.gov/Entrez/) and the HOVERGEN database (Duret *et al.* 1994) to find nine other genes for which signal sequence information was available for mouse and rat orthologues. Alignments were prepared using the Wisconsin package (Genetics Computer Group 1994). DNA alignments were prepared for both signal sequences and entire protein-coding sequences using the default settings of PILEUP. The program GAPFRAME was used to move all gaps to codon boundaries. Nucleotide substitution rates were estimated with the Moriyama package (Moriyama and Powell 1997), using the synonymous/nonsynonymous method of Li (1993) and the multiple substitution correction method of Tamura and Nei (1993).

The genes (accession numbers in brackets, mouse then rat) were as follows: interleukin 4 receptor (M29854, X69903), low density lipoprotein receptor (X64414, X13722), tumor necrosis factor receptor 1 (M60468, M63122), calreticulin (X14926, X53363), cholesterol esterase (U33169, X16054), glutamate dehydrogenase (X57024, X14223), procathepsin E (X97399, D38104), immunoglobulin light chain V-λ-1 (J00590, M17092), and oxytocin-neurophysin I (M88355, K01701).

RESULTS

Repeatability of K_a/K_s: With the 24 nonoverlapping sections, K_a/K_s patterns across *Igf2r* are significantly repeatable ($P_{rank} < 0.01$, see Figure 1). It appears that the large-scale fluctuations in K_a/K_s rather than the small-scale fluctuations are responsible for the observed repeatability (our unpublished results). K_a/K_s is an indicator of both the intensity and type of selection. Stronger stabilizing selection reduces K_a/K_s, while both weaker stabilizing selection and stronger positive selection increase K_a/K_s. A K_a/K_s ratio significantly above one is the strict requirement for evidence of positive selection, but K_a/K_s ratios less than one do not preclude positive selection because different sites within the same segment might be under different selection pressures. The repeatability of patterns of selection across *Igf2r* means that selection is able to respond differently to various small subdomains across the *Igf2r* protein. This result is consistent with the conservation of the various functions of the different parts of the Igf2r molecule across the cow, human, mouse, and rat species.

The significant repeatability of K_a/K_s across the *Igf2r* gene means that the unexpected K_a/K_s patterns are not artifacts, but are representative of *Igf2r*'s molecular evolution. We therefore need to explain the low rate of protein evolution at the Igf2 binding site, and the high rates of protein evolution both at the signal sequence and at one of the M6P binding sites.

Similarity of amino acid substitutions: To examine the selection pressures acting on the fast-evolving signal sequence and M6P binding site, we compared the amino

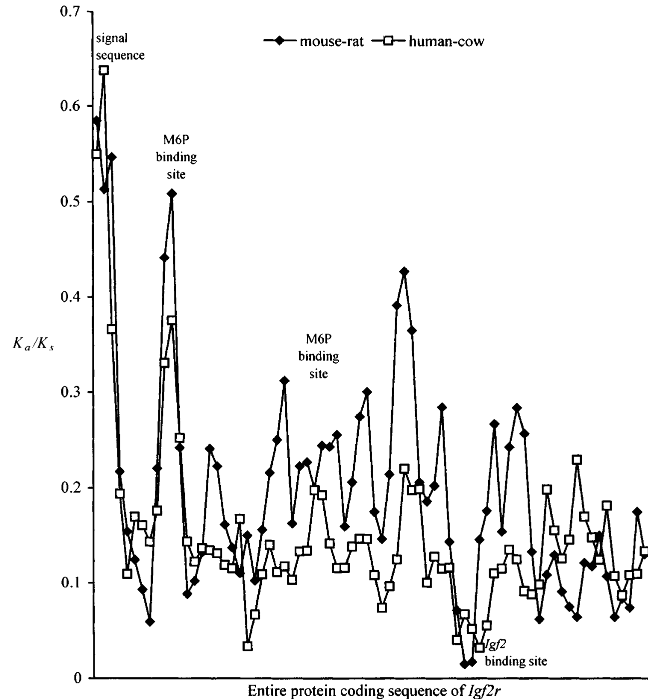


Figure 1.—K_a/K_s patterns are repeatable across *Igf2r*. The nonsynonymous over synonymous substitution rates (K_a/K_s) across the *Igf2r* gene are shown for both the mouse-rat and human-cow comparisons. The data are for the 74 overlapping windows. Note the repeated peaks at the signal sequence and the first M6P binding site, and the repeated troughs at the Igf2 binding site.

acids that had changed in each nonoverlapping window to ask how conservative the changes were, using data from the appropriate PAM matrices to ascribe similarity.

In the human-cow comparison the changes in the signal sequence were on average exceptionally nonconservative (see Figure 2). Indeed, none of the windows shows a lower level of similarity and the average level

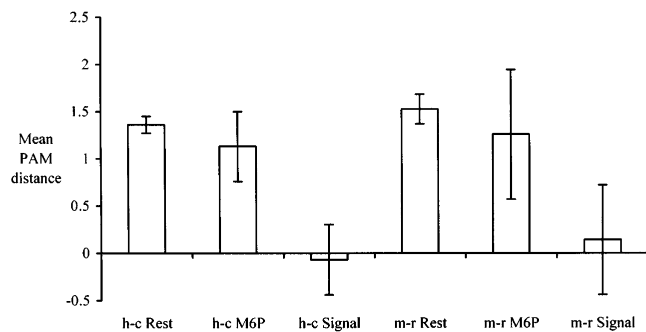


Figure 2.—Mean PAM distances per amino acid change for different regions of *Igf2r*. Data are presented for both the human-cow (h-c) and mouse-rat (m-r) comparisons. The human-cow and mouse-rat data were obtained using the PAM22-29 and PAM6-8 matrices, respectively (see materials and methods for references). Signal, the window containing the signal sequence; M6P, the two windows containing the rapidly evolving mannose-6-phosphate binding site; and Rest, all other parts of the gene. Bars show SE.

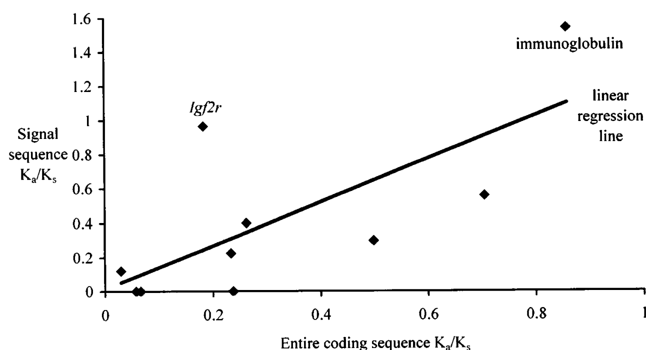


Figure 3.—K_a/K_s of signal sequences *vs.* K_a/K_s of entire genes. A comparison is shown of the mouse-rat K_a/K_s values of the signal sequence and of the entire protein coding sequence for ten genes (full gene names and accession numbers in materials and methods). Both *Igf2r* and the immunoglobin gene show higher signal sequence K_a/K_s values than the other genes.

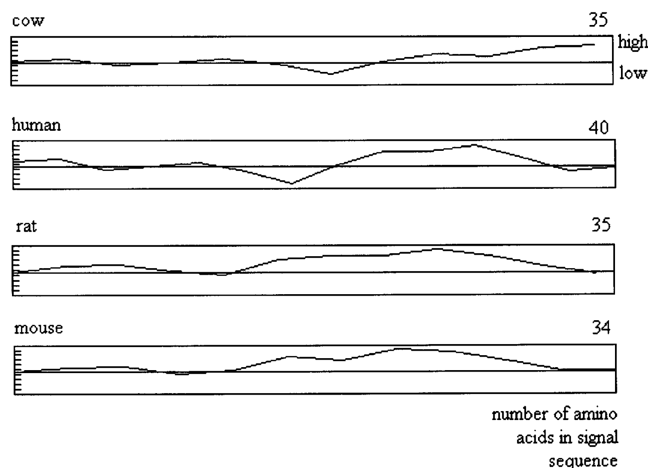


Figure 4.—Hydrophobicity plots across *Igf2r* for four species. Sliding window plots of hydrophobicity values [as produced by Staden's (1996) package] across the signal sequences of the *Igf2r* gene in mouse, rat, cow, and human are shown. The signal sequence of cow is the only one that does not show the prototypical pattern of a central peak. The numbers give the lengths of the different signal sequences in amino acids (showing that the plots are not aligned).

of similarity in the gene as a whole is significantly higher ($P = 0.0017$, two-tailed Mann-Whitney *U*-test). This suggests that positive selection might be acting on this sequence. In the mouse-rat comparison the picture is not so clear. The signal sequence does show a low similarity but three other windows have lower values. Compared with the rest of the gene the signal sequence does have a lower value ($P = 0.0387$, two-tailed Mann-Whitney *U*-test).

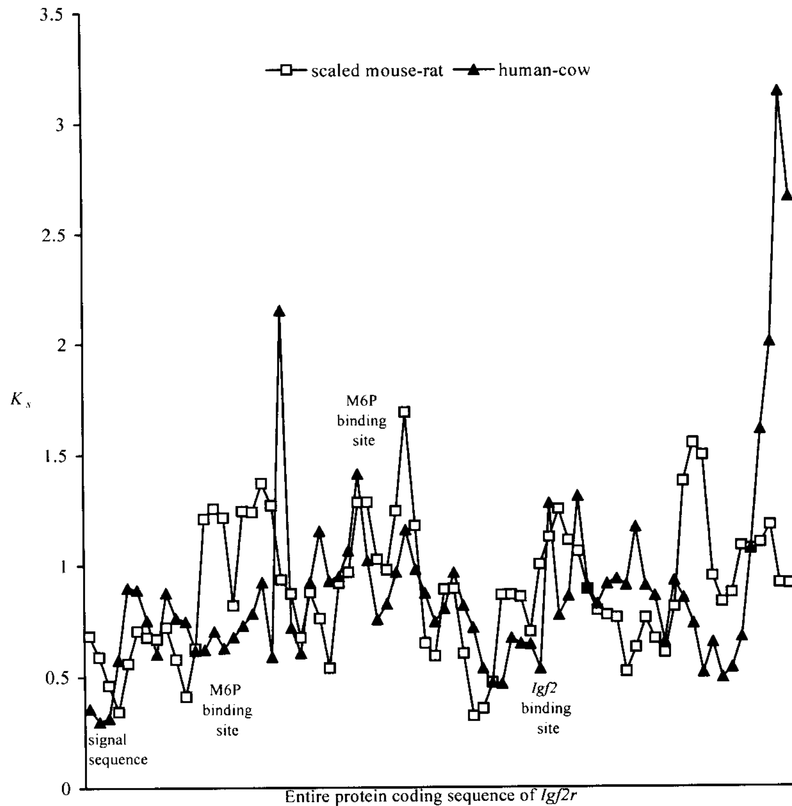
At the M6P binding site there appears to be nothing unusual about the amino acids being replaced in both mouse-rat and human-cow, suggesting that this region is under weak stabilizing selection (two-tailed Mann-Whitney *U*-tests between the M6P site and the rest of the gene give $P = 0.6248$ and $P = 0.6829$ for the human-cow and mouse-rat comparisons, respectively).

The signal sequence of Igf2r is unusually fast evolving: One would expect signal sequences to be generally slow evolving (see Introduction), but is this actually true; *i.e.*, Is the signal sequence of *Igf2r* unusually fast evolving? When signal sequence K_a/K_s is plotted against the entire coding sequence K_a/K_s for the nine genes and for *Igf2r* it becomes clear that *Igf2r*'s signal sequence is evolving unusually fast (see Figure 3). The only gene with a comparable signal sequence rate is immunoglobulin light chain V-λ-1. However, the rate of signal sequence evolution in this gene is comparable to the rate of evolution in the rest of the gene. *Igf2r* is hence unusual in that the signal sequence shows an unusually high ratio of K_a/K_s given the rate of evolution of the rest of the sequence. Compared with a large sample of genes (unpublished data) the K_a/K_s ratio of the majority of *Igf2r* is not unusual. The value in the signal sequence is high in these terms. Hence the discrepancy between the signal sequence and the rest of the gene is better understood as unusually fast evolution of the signal sequence rather than unusually slow evolution of the majority of the gene.

Signal sequence hydrophobicity: Is the rapid evolution of the *Igf2r* signal sequence at the expense of its functionality? We examined the hydrophobicity plots of the four species' signal sequences using the Staden package (Staden 1996). The conserved signal sequence structure (see Introduction) predicts a central hydrophobic peak. The signal sequences of human, rat, and mouse fulfilled this prediction but the signal sequence of cow showed no such feature (Figure 4). This observation could explain the high divergence in the human-cow comparison, but offers no clues as to why both the human-cow and mouse-rat signal sequence divergences are so high.

Repeatability of K_s: With the 24 nonoverlapping sections, K_s variation is significantly repeatable ($P_{rank} = 0.05$; see Figure 5). With the rank of size change between neighboring overlapping sections as a potentially more powerful test, K_s patterns remained significantly repeatable ($P_{rank} < 0.05$). The repeatability of K_s across the *Igf2r* gene thus implies that the variation in K_s within the gene is being shaped by deterministic forces. We have performed several tests to attempt to distinguish between the potential deterministic forces described above (see Introduction). First, however, we ask whether the repeatability might be an artifact.

K_s repeatability is probably not a methodological artifact: Unless multiple substitution correction methods take account of variation in sequence composition, nucleotide substitution rate estimates can be biased (Pesole *et al.* 1995). Furthermore, the use of different substitution rate estimation methods can lead to different conclusions using the same dataset (Smith and Hurst 1998). Could this effect be responsible for our observation of K_s repeatability, since Kimura's (1980) two parameter



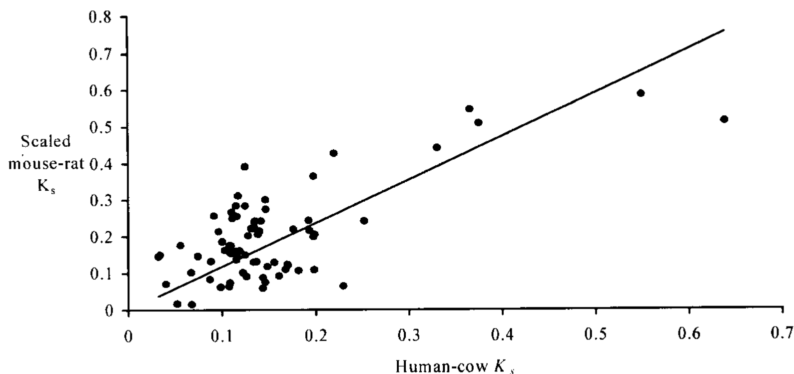


Figure 5.—K_s patterns are repeatable across *Igf2r*. The mouse-rat and human-cow synonymous substitution rates (K_s) across the *IGF2R* gene are shown in the upper plot. The data are of the 74 overlapping windows. The actual human-cow values are used, but the mouse-rat values are scaled up so that the means of both sets of K_s values are the same, to allow easier comparison. The lower plot shows the correlation between the mouse-rat and human-cow orthologous overlapping window K_s values.

model does not take account of variation in sequence composition? Under this hypothesis our finding of K_s repeatability would be an artifact caused by compositional repeatability leading to repeatable bias in K_s estimates. To test this idea, we reanalyzed the section alignments with Moriyama's package (Moriyama and Powell 1997) using the synonymous/nonsynonymous method of Li (1993) and the multiple substitution correction method of Tamura and Nei (1993; that uses all four nucleotide frequencies, as well as a transversion substitution rate and two transition substitution rates). A significant K_s repeatability was observed ($P_{rank} < 0.05$), and thus the repeatability of K_s is probably not a methodological artifact.

K_s repeatability is not due to the majority of Igf2r consisting of repeated units: The large extracellular domain of IGF2R consists almost entirely of 15 conserved repeat units, between which there is roughly 20% amino acid identity in the human gene (Morgan *et al.* 1987). If compositional features affect K_s values, one might expect to see each repeat giving similar K_s patterns. Using the Wisconsin package (Genetics Computer Group 1994), alignments of the repeats were prepared using the same alignment gaps as for the preparation of overlapping and nonoverlapping sections. For each repeat, and for both species pairs, three alignments were prepared: one of the 5′ three-fifths of the repeat (section 1), one of the middle three-fifths (section 2), and one of the 3′

three-fifths (section 3). Overlapping sections were used because of the small size of the repeats (~150 bp). K_s was then estimated for each alignment using Moriyama's package (Moriyama and Powell 1997).

It was then determined for each repeat in both species comparisons which of the three sections in each repeat (first, second, or third) had the highest K_s. The null hypothesis that there is no link between repeats and mutation rates predicts that in any given repeat all three thirds of the sequence are equally likely to have the fastest rate of evolution. The alternative hypothesis that the repeat structure does influence K_s would suppose that repeatably one of the three sections will have the highest K_s. Thus for both species comparisons the null expectation was that of the 15 repeats, five sections should have the first third having the highest K_s, five should have the second with highest K_s, etc. The human-cow comparison gave (6, 5, 4) while the mouse-rat comparison gave (7, 3, 5). Neither result was significantly different from the null expectation (the chi-squared test gave $P > 0.5$ for both comparisons). Thus the null hypothesis of no link between repeats and K_s patterns cannot be rejected, and so we conclude that the repeatability of K_s patterns is probably not due to the repeatability of molecular evolution in repeats. We note, however, that this is a relatively weak test.

Having established that the repeatability in K_s is probably not an artifact, we now turn to the question of why this repeatability is found. We consider three hypotheses (see Introduction for details): either (1) the variation is due to differences in composition along the gene allied with some form of mutational bias, or (2) the variation is due to selection favoring important sites to have low mutation rates (selected mutation rate hypothesis), or (3) K_s variation is the result of varying selection on codon usage (common constraints hypothesis). Hypotheses 1 and 2 assume that silent site mutations are neutral. Hypotheses 2 and 3 propose that the repeatability is selection-driven rather than mutation-driven. We shall start with these.

Adaptive explanations for K_s repeatability: Here we evaluate the extent to which adaptive explanations for K_s variability and repeatability can explain the data. The common constraints argument predicts that codon bias should be negatively correlated with K_s [because both depend on the strength of selection, for Drosophila and bacterial data see Li (1997)], and that there should be a correlation between K_a and K_s (because both are influenced by similar selective pressures). The selected mutation rate argument predicts that regions of the gene under strong stabilizing selection (low K_a) should have low mutation rates (low K_s), and conversely regions under weak stabilizing selection or even positive selection (high K_a) should have high mutation rates (high K_s). For both adaptive hypotheses, K_s repeatability follows from selection patterns being conserved across all four species. Thus the selected mutation rate argument

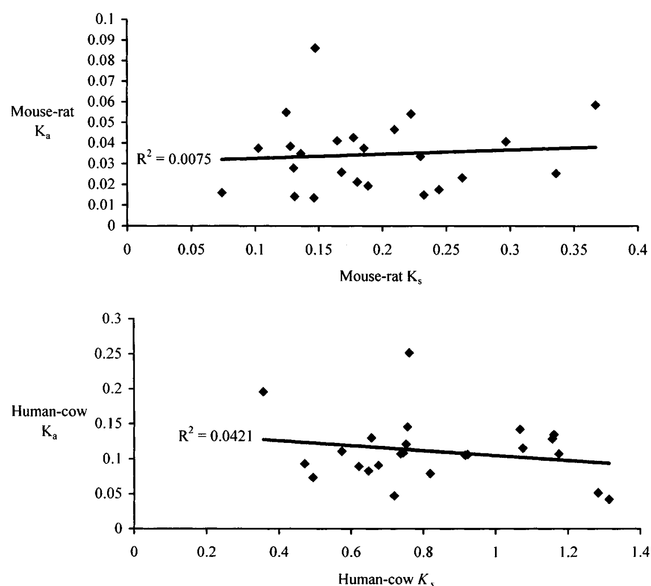


Figure 6.—No intragenic correlation between K_a and K_s. The K_a value of each nonoverlapping section is plotted against the corresponding K_s value for both the mouse-rat and the human-cow comparisons. The linear regression lines for both comparisons indicate no significant positive correlation between K_a and K_s.

agrees with the common constraints argument in predicting K_s repeatability and a correlation between K_a and K_s, but differs in predicting no negative correlation between codon bias and K_s.

Evidence from codon bias data: To differentiate between the selected mutation rate and common constraints hypotheses, we looked for a relationship between codon bias and K_s within the *Igf2r* gene. We used the Moriyama package (Moriyama and Powell 1997) to determine effective numbers of codons (*ENC*; Wright 1990) for the sections. For both species comparisons the *ENC* data of the two species were averaged. There was no correlation between mouse-rat *ENC* and K_s ($P_{rank} \gg 0.1$), but the human-cow comparison showed a negative correlation on the edge of significance ($P_{rank} = 0.05$). A negative correlation between *ENC* and K_s is equivalent to a positive correlation between codon bias and K_s (the lower the effective number of codons the stronger the codon bias), and so both species comparisons rejected the prediction of the common constraints hypothesis. The codon bias data show us that silent sites do not appear to be affected by selection, and thus K_s can be taken as an indicator of mutation rate.

No correlation between K_s and K_a: Both selective explanations of K_s repeatability predict a positive correlation between K_s and K_a. However, neither the mouse-rat nor the human-cow comparison give a significant positive correlation ($P_{rank} > 0.5$ for both; see Figure 6). This result provides evidence against both adaptive explanations of K_s repeatability. We note additionally, that presence of a significant positive correlation between K_a and

K_s could not have been taken as conclusive evidence in favor of the selected mutation rate argument because there are several other possible explanations for such a correlation (see Li 1997).

No correlation between K_s and functional importance: We separated the gene into sections according to predicted functional importance, and then examined whether regions likely *a priori* to be important (such as binding sites) showed a systematically lower K_s than the rest of the gene. Both the common constraints and selected mutation rate hypotheses would predict lower K_s values for more important sections of the gene because in such regions selection is likely to be more powerful. However, both the mouse-rat and human-cow comparisons gave a higher mean K_a/K_s for the extracellular ligand-binding regions than the extracellular regions not associated with ligand binding (0.27 *vs.* 0.17 and 0.20 *vs.* 0.11, respectively). This result contradicts the prediction of greater functional importance of binding sites (with stabilizing selection, the stronger the selection the lower the K_a/K_s values), and thus renders this test inapplicable. However, the observation that the K_s at the Igf2 binding site is not low despite the K_a/K_s being unusually low (see Figures 1 and 5) provides further, albeit circumstantial, evidence against both the common constraints and selected mutation rate hypotheses.

Mutational explanations for K_s repeatability: Here we consider two possible compositional correlates to K_s variation. First we ask whether GC composition might explain the patterns. Second we ask more particularly whether methylation at CpG sites might explain some of the variance.

GC composition: If local base composition influences mutation rates (and hence K_s), then the repeatability of K_s between highly related (and hence compositionally similar) sequences follows. For example, a correlation between GC content and K_s has been proposed (Wolfe *et al.* 1989). However, we found no significant correlation between fourfold site GC content and K_s for either the mouse-rat ($P_{rank} > 0.1$) or human-cow ($P_{rank} > 0.1$) comparisons. Wolfe *et al.* (1989) is now discredited due to (1) a small sample size effect (Wolfe and Sharp 1993) and (2) a biased estimating protocol (Bernardi *et al.* 1997). Correction for both finds no correlation between chromosomal site (defined by GC content) and mutation rate given by K_s (Bernardi *et al.* 1997), as we have also found (Smith and Hurst 1998). The finding that the repeats within the *Igf2r* gene do not show similar K_s plots (see above) also suggests that composition might not explain the patterns of K_s variation.

The influence of methylation: Even though local GC content seems not to influence local K_s, more complicated compositional characters may well influence local K_s (for example, see Morton *et al.* 1997), and thus the above result does not falsify the compositional hypothesis. One compositional feature that might well influence K_s patterns, and be responsible for the K_s repeatability under the compositional hypothesis, is the methylation pattern. Methylation is known to affect mutation rates with a methylated cytosine in a CpG pair 10 to 20 times more mutable than an unmodified cytosine (Kendrew 1994).

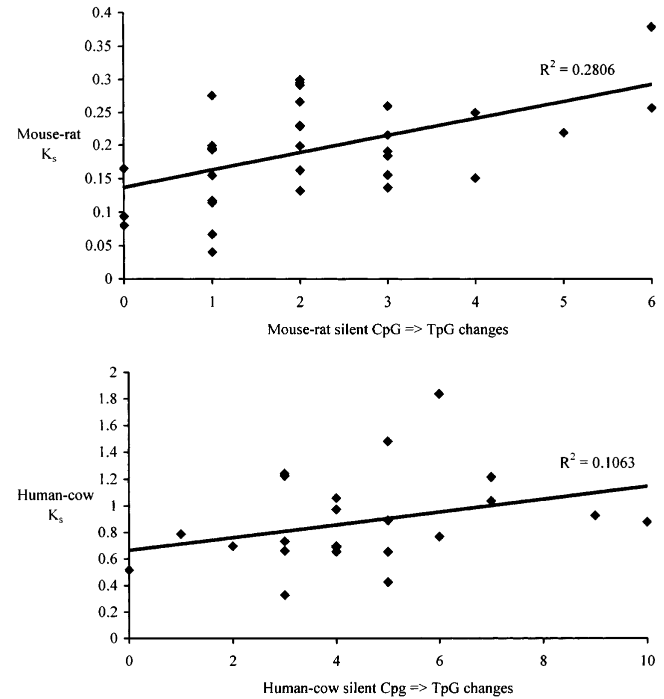


Figure 7.—Silent CpG ⇒ TpG changes correlate with K_s. For both comparisons of mouse-rat and human-cow, and for each nonoverlapping window, K_s is plotted against the number of silent CpG ⇒ TpG changes. The correlations show that such methylation-induced mutations provide a possible explanation for repeatable K_s patterns.

To test this possibility we examined the frequency of CpG ⇒ TpG mutations. For each segment we found the number of times a CpG was found in one of the two sequences and TpG was found in the other, where the C existed at a third site. This protocol gives all possible synonymous CpG ⇒ TpG changes (we assume all mutational changes to be in this direction). We find that the variation in K_s between segments is correlated to the frequency of such mutations (see Figure 7). The mouse-rat comparison shows a significant correlation ($P_{rank} = 0.02$), while the human-cow comparison shows a tendency in the same direction ($0.1 > P_{rank} > 0.05$). We also find that this class of mutations accounts for a reasonable proportion of all mutations (15% in the human-cow comparison, and 21% in the mouse-rat comparison). We conclude that methyl-induced mutations can account for some of the repeatability that we observe.

Two hypotheses might be considered to account for this pattern. At one extreme, all segments might have an equal number of CpG sites that might change, but these segments differ in the density of methylation. Alternatively, the segments might have the same density of methylation per CpG, but differ in the total number of CpG sites.

We find that variation in total number of mutable CpGs partially explains the pattern in K_s. We calculated the number of times CpG was found in both sequences (*i.e.*, the CpG/CpG frequency) again with the C at the third site. Adding the CpG/CpG figure to the CpG/TpG figure we arrive at a figure for the total number of mutable or mutated silent CpG sites. This total correlates well with K_s ($P_{rank} < 0.05$ for the mouse-rat comparison, though $P_{rank} > 0.1$ for the human-cow comparison), indicating that the variance in K_s explained by methylation is to some extent a result of differences between segments in CpG content.

Conversely, we can ask whether segments also differ in their propensity to methylate CpG sites. The methylation of *Hpa*II sites of the mouse E15 embryo *Igf2r* gene has been shown to vary from 50% to 90% throughout the locus (Stoger *et al.* 1993), which at least suggests that there might be enough variation in methylation patterns to account for the highly variable mutation rate patterns (Figure 5). Asking whether segments differ significantly in the proportion of CpGs that become TpGs we find heterogeneity between segments (chi-squared test for heterogeneity gives $P < 0.05$ for the mouse-rat comparison, and $0.5 > P > 0.1$ for the human-cow comparison). This supports the view that the observed variation in CpG > TpG changes (and hence K_s) is at least partially due to differences in the methylation density of potentially methylated sites.

This finding contrasts with human polymorphism data. The M6P/IGF2R Information Core (http://www.radonc.duke.edu/~jirtle/homepage.html) provides a human mutation database for *IGF2R*. Six substitutional mutations resulting in disease symptoms are described, with no C to T mutations. However, this small dataset suffers from an ascertainment bias, in that not all substitutional mutations are equally likely to cause disease symptoms.

DISCUSSION

The mammalian *Igf2r* gene is multifunctional, and plays an important role in glycoprotein transport via M6P binding (Kornfeld 1992) and growth and development via IGF2 and RA binding (Kang *et al.* 1998), and also tumor suppression (Desouza *et al.* 1997). In keeping with this functional complexity mammalian *Igf2r* displays complex patterns of molecular evolution. The significant repeatability of both K_a/K_s and K_s means that these patterns are driven by deterministic forces, and we have tested a number of hypotheses concerning the nature of these forces.

K_a/K_s variation and the conflict theory of imprinting: A simple prediction of the conflict theory of imprinting is that the Igf2 binding site should exhibit a high rate of molecular evolution driven by positive selection (McVean and Hurst 1997b; Trivers and Burt 1998). Instead we find evidence from K_a/K_s patterns for strong stabilizing selection at the Igf2 binding site in both the mouse-rat and human-cow comparisons. This finding cannot then be dismissed as a statistical artifact. Fast evolution at this site might have provided good evidence supporting the conflict hypothesis (McVean and Hurst 1997b; Trivers and Burt 1998). We have looked for positive selection in other parts of the gene, and have found evidence from PAM matrix data that the signal sequence might be under positive selection. We have shown that signal sequence evolution in *Igf2r* is exceptional when compared to other examples of signal sequence evolution. This is the first evidence for any form of molecular evolution (*sensu stricto*) in an imprinted gene that is different from the form of molecular evolution in most nonimmune system/nonplacental genes [note, however, that imprinted genes have a few other unusual properties, such as intronic dimensions (Hurst *et al.* 1996), sex specific recombination rates (Paldi *et al.* 1995), and repeat structures (Neumann *et al.* 1995)].

The evolution of *Igf2r*'s signal sequence might be a result of antagonistic coevolution, in which case we can speculate that the signal sequence might be coevolving with the gene product of an imprinted gene (or genes) expressed from the paternally derived genome. The antagonism might concern the location of Igf2r. The putative paternally derived product could, in theory, attempt to remove the protein to a cellular location where its effects on Igf2 levels are minimal. It would be interesting to see the effect of site-directed mutagenesis of the signal sequence on the cellular location of Igf2r (no mutations in the *IGF2R* signal sequence causing human disease are listed at the M6P/IGF2R Information Core at http://www.radonc.duke.edu/~jirtle/homepage.html). Alternatively, the unusual evolution at this site might be a property of *Igf2r* independent of imprinting. It would then be informative to analyze the molecular evolution of *Igf2r* in birds for example, assuming that avian *Igf2r* is not imprinted (as yet, chicken is the only bird for which the *Igf2r* gene has been sequenced).

Both this study and a previous analysis of the K_a/K_s values of imprinted genes (McVean and Hurst 1997b) assume that the conflict theory of imprinting predicts rapid protein evolution at the sites of functional antagonism. This assumption is in keeping with the results of studies of classic parent-offspring conflict (see Introduction), though such fast rates of evolution might be due to the acquisition of new roles. However, Haig (1997) has suggested an alternative prediction, that the mutual dependence of the roles of Igf2r and Igf2 leads to an ESS rather than a continuing arms race. Haig suggests that the *a priori* expectation is unclear. This failure to provide a testable prediction of the rates of evolution at sites of functional antagonism might be considered a weakness of the current models of the conflict theory of imprinting.

K_s variation and the evolution of mutation rates: The result of K_s repeatability in mammalian *Igf2r*, along with

evidence from codon bias data that the silent sites appear to be neutral, extends our understanding of mutation rate variation: to quote Casane *et al.* (1997) "it is still not clear whether the mutation pattern is constant over the entire genome of an organism or is variable among regions of the genome." Large scale variation in mammalian mutational patterns has been previously demonstrated on two levels: between different chromosomes [K_s used as a measure of mutation rate in McVean and Hurst (1997a)] and between pseudogenes inserted into different genomic regions (Casane *et al.* 1997). The repeatability of K_s patterns within *Igf2r* provides strong evidence that there is deterministic variation in mutation rates at a fine scale, between different regions within the same gene.

We have tested various hypotheses to explain the K_s repeatability. Since the K_s repeatability does not appear to be a methodological artifact, the remaining hypotheses can be divided into two groups, mutational and adaptive explanations. The selected mutation rate argument holds that local mutation rates, given by K_s, are the results of conflicting selection pressures governing mutation rates (see Introduction). Alternatively, the common constraints argument proposes that silent sites may be under selection, and so K_s is not the mutation rate but instead depends on selective constraints (see Introduction). The lack of a negative correlation between codon bias and K_s provides evidence against the common constraints hypothesis, and the lack of a correlation between K_s and K_a provides evidence against both adaptive explanations. The comparison of local mutation rates and local functional importance provided further, but circumstantial, evidence against both adaptive hypotheses.

Repeatable composition of repeats does not cause the repeatable K_s patterns. We did, however, find that silent C $\Rightarrow$ T mutations at CpG sites do covary with K_s, suggesting that the variation in K_s is somehow connected with methylation-induced mutational patterns. The variation in silent C $\Rightarrow$ T mutations at CpG sites appears to be the result of variation both in the number of mutable sites, and in the density of methylation.

It has previously been reported that, at least for GC-rich genes, the variation in K_s does covary with the variation in K_a (Alvarez-Valin *et al.* 1998). This was interpreted as evidence for selection differentially affecting codon usage across the gene. Were this the only force affecting K_s variation, then one would not have predicted K_s repeatability in a gene, such as *Igf2r*, in which K_a and K_s do not covary. We hence conclude that there must be some other forces determining variation in K_s. Variation in silent CpG mutation appears in our case to be an important variable and it is noteworthy that analysis of GC content alone did not detect this.

We thank Etsuko Moriyama and Josep Comeron for providing us with their programs for evolutionary analysis, Gil McVean for suggestions concerning methodology, David Haig for discussions on imprinting, Denise Barlow for advice on methylation data, and two anonymous referees for suggesting many improvements.

LITERATURE CITED

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

Alvarez-Valin, F., K. Jabbari and G. Bernardi, 1998 Synonymous and nonsynonymous substitutions in mammalian genes: intragenic correlations. J. Mol. Evol. **46:** 37–44.

Benner, S. A., M. A. Cohen and G. H. Gonnet, 1994 Amino-acid substitution during functionally constrained divergent evolution of protein sequences. Prot. Eng. **7:** 1323–1332.

Bernardi, G., D. Mouchiroud and C. Gautier, 1997 Isochores and synonymous substitutions in mammalian genes, pp. 137–168 in *DNA and Protein Sequence Analysis*, edited by M. J. Bishop and C. J. Rawlings. Oxford University Press, Oxford.

Casane, D., S. Boissinot, B. H. J. Chang, L. C. Shimmin and W. H. Li, 1997 Mutation pattern variation among regions of the primate genome. J. Mol. Evol. **45:** 216–226.

Chorney, M. J., K. Chorney, N. Seese, M. J. Owen, J. Daniels *et al.*, 1998 A quantitative trait locus associated with cognitive ability in children. Psychol. Sci. **9:** 159–166.

Comeron, J. M., 1995 A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. J. Mol. Evol. **41:** 1152–1159.

Dennis, P. A., and D. B. Rifkin, 1991 Cellular activation of latent transforming growth factor beta requires binding to the cation-independent mannose 6-phosphate/insulin-like growth factor type II receptor. Proc. Natl. Acad. Sci. USA **88:** 580–584.

Desouza, A. T., T. Yamada, J. J. Mills and R. L. Jirtle, 1997 Imprinted genes in liver carcinogenesis. FASEB J. **11:** 60–67.

Drake, J. W., 1991 A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. USA **88:** 7160–7164.

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.

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

Haig, D., 1997 Parental antagonism, relatedness asymmetries, and genomic imprinting. Proc. R. Soc. Lond. B **264:** 1657–1662.

Hughes, A. L., M. K. Hughes, C. Y. Howell and M. Nei, 1994 Natural-selection at the class-II major histocompatibility complex loci of mammals. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. **346:** 359–366.

Hurst, L. D., 1997 Evolutionary theories of genomic imprinting, pp. 211–237 in *Frontiers in Molecular Biology: Genomic Imprinting in Mammals*, edited by W. Reik and A. Surani. Oxford University Press, Oxford.

Hurst, L. D., and G. T. McVean, 1997 Growth effects of uniparental disomies and the conflict theory of genomic imprinting. Trends Genet. **13:** 436–443.

Hurst, L. D., G. T. McVean and T. Moore, 1996 Imprinted genes have few and small introns. Nat. Genet. **12:** 234–237.

Izard, J. W., S. L. Rusch and D. A. Kendall, 1996 The amino-terminal charge and core region hydrophobicity interdependently contribute to the function of signal sequences. J. Biol. Chem. **271:** 21579–21582.

Kang, J. K., Y. Li and A. Leaf, 1998 Mannose-6-phosphate/insulin-like growth factor-II receptor is a receptor for retinoic acid. Proc. Natl. Acad. Sci. USA **95:** 13671–13676.

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

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, UK.

Kornfeld, S., 1992 Structure and function of the mannose 6-phos-

phate insulin-like growth factor-II receptors. Annu. Rev. Biochem. **61:** 307–330.
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.
McVean, G. T., and L. D. Hurst, 1997a Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. Nature **386:** 388–392.
McVean, G. T., and L. D. Hurst, 1997b Molecular evolution of imprinted genes: no evidence for antagonistic coevolution. Proc. R. Soc. Lond. Ser. B Biol. Sci. **264:** 739–746.
Mills, J. J., J. G. Falls, A. T. Desouza and R. C. Jirtle, 1998 Imprinted MGP/Igf2 receptor is mutated in rat liver tumours. Oncogene **16:** 2797–2802.
Mochizuki, A., Y. Takeda and Y. Iwasa, 1996 The evolution of genomic imprinting. Genetics **144:** 1283–1295.
Moore, T., and D. Haig, 1991 Genomic imprinting in mammalian development: a parental tug-of-war. Trends Genet. **7:** 45–49.
Morgan, D. O., J. C. Edman, D. N. Standring, V. A. Fried, M. C. Smith *et al.*, 1987 Insulin-like growth factor-ii receptor as a multifunctional binding-protein. Nature **329:** 301–307.
Moriyama, E. N., and J. R. Powell, 1997 Synonymous substitution rates in *Drosophila*: Mitochondrial versus nuclear genes. J. Mol. Evol. **45:** 378–391.
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.
Ogawa, O., D. M. Becroft, I. M. Morison, M. R. Eccles, J. E. Skeen *et al.*, 1993 Human insulin-like growth-factor type-I and type-II receptors are not imprinted. Hum. Mol. Genet. **2:** 2163–2165.
Paldi, A., G. Gyapay and J. Jami, 1995 Imprinted chromosomal regions of the human genome display sex-specific meiotic recombination frequencies. Curr. Biol. **5:** 1030–1035.
Pesole, G., G. Dellisanti, G. Preparata and C. Saccone, 1995 The importance of base composition in the correct assessment of genetic-distance. J. Mol. Evol. **41:** 1124–1127.
Smith, N. G. C., and L. D. Hurst, 1998 Sensitivity of patterns of molecular evolution to alterations in methodology. J. Mol. Evol. (in press).
Spencer, H. G., M. W. Feldman and A. G. Clark, 1998 Genetic conflicts, multiple paternity and the evolution of genomic imprinting. Genetics **148:** 893–904.
Staden, R., 1996 The Staden sequence-analysis package. Mol. Biotech. **5:** 233–241.
Stoger, R., P. Kubicka, C. G. Liu, T. Kafri, A. Razin *et al.*, 1993 Maternal-specific methylation of the imprinted mouse *Igf2r* locus identifies the expressed locus as carrying the imprinting signal. Cell **73:** 61–71.
Sutton, K. A., and M. F. Wilkinson, 1997 Rapid evolution of a homeodomain: evidence for positive selection. J. Mol. Evol. **45:** 579–588.
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.
Trivers, R., and A. Burt, 1998 Kinship and genomic imprinting, in *Genomic Imprinting*, edited by R. Ohlsson. Springer, Heidelberg (in press).
Trivers, R. L., 1974 Parent-offspring conflict. Am. Zool. **14:** 249–264.
Tsaur, S. C., and C. I. Wu, 1997 Positive selection and the molecular evolution of a gene of male reproduction, Acp26Aa of *Drosophila.* Mol. Biol. Evol. **14:** 544–549.
Tucker, P. K., and B. L. Lundrigan, 1995 The nature of gene evolution on the mammalian Y-chromosome—lessons from SRY. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. **350:** 221–227.
Wallis, M., 1981 The molecular evolution of pituitary growth hormone prolactin and placental lactogen: a protein family showing variable rates of evolution. J. Mol. Evol. **17:** 10–18.
Wallis, M., 1993 Remarkably high rate of molecular evolution of ruminant placental lactogens. J. Mol. Evol. **37:** 86–88.
Wallis, M., 1994 Variable evolutionary rates in the molecular evolution of mammalian growth hormones. J. Mol. Evol. **38:** 619–627.
Wolfe, K. H., and P. M. Sharp, 1993 Mammalian gene evolution—nucleotide-sequence divergence between mouse and rat. J. Mol. Evol. **37:** 441–456.
Wolfe, K. H., P. M. Sharp and W. H. Li, 1989 Mutation-rates differ among regions of the mammalian genome. Nature **337:** 283–285.
Wright, F., 1990 The effective number of codons used in a gene. Gene **87:** 23–29.
Xie, S., J. Green, J. Bixby, B. Szafranska, J. Demartini *et al.*, 1997 The diversity and evolutionary relationships of the pregnancy-associated glycoproteins, an aspartic proteinase subfamily consisting of many trophoblast-expressed genes. Proc. R. Soc. Lond. Ser. B Biol. Sci. **94:** 12809–12816.
Xu, Y. Q., C. G. Goodyer, C. Deal and C. Polychronakos, 1993 Functional polymorphism in the parental imprinting of the human IGF2R gene. Biochem. Biophys. Res. Commun. **197:** 747–754.
Zhou, M., Z. M. Ma and W. S. Sly, 1995 Cloning and expression of the cDNA of chicken cation-independent mannose-6-phosphate receptor. Proc. Natl. Acad. Sci. USA **92:** 9762–9766.

Communicating editor: G. B. Golding