Mini Review

Can RNA Selection Pressure Distort the Measurement of Ka/Ks?

Running title: RNA selection pressure and Ka/Ks
Keywords: evolution; splicing; RNA; synonymous substitutions; nonsynonymous substitutions

Yi Xing and Christopher Lee
Molecular Biology Institute, Center for Genomics and Proteomics,
Dept. of Chemistry and Biochemistry
University of California, Los Angeles

Corresponding author: Lee, C. (leec@mbi.ucla.edu)
Address for correspondence:
611 Charles E Young Drive, Boyer Hall
University of California, Los Angeles
Los Angeles, CA, 90095, USA
TEL 310-825-7374
FAX 310-206-7286

Abbreviations: ESE, exonic splicing enhancer.

Draft 20
Dec 15, 2005

Abstract

Recently, an interesting question has emerged in the evolutionary interpretation of sequence substitution data as evidence of amino acid selection pressure. Specifically, the Ka/Ks metric was designed to measure selection pressure on amino acid substitutions, assuming that the synonymous substitution rate Ks reflects the neutral nucleotide substitution rate. However, there is increasing evidence for selection pressure at silent sites due to constraints of RNA splicing. Is Ka/Ks an appropriate metric for selection pressure on amino acid substitutions, in the presence of other selection pressures acting only at the RNA level (such as selection for exonic splicing enhancers)? Or can the resulting decreases in Ks from such selection pressures introduce bias into the Ka/Ks metric, so that it no longer gives an accurate measure of amino acid level selection pressure? In this review we present both mathematical models and empirical evidence for these divergent points of view.

1. Background: the Ka/Ks Metric of Amino Acid Selection Pressure

The Ka/Ks metric was first proposed as a means of distinguishing amino acid selection pressure ω from the background nucleotide mutation rate π (Li et al., 1985; Nei and Gojobori, 1986; Yang and Bielawski, 2000; Yang and Nielsen, 2000; Hurst, 2002). This approach took advantage of the fact that synonymous sites (nucleotides where substitutions produce no change in the encoded amino acid) are subject to the background nucleotide mutation rate π, but not to the amino acid selection pressure ω. By contrast, non-synonymous sites are subject to both processes. Thus, the rate of observed mutations at non-synonymous sites (Ka) vs. the rate at synonymous sites (Ks) can be combined to obtain an estimate of the amino acid selection pressure ω:

Non-synonymous sites: $Ka=\omega\pi$

Synonymous sites: $Ks=\pi$

$$Ka / Ks = \omega$$

Typically, Ka and Ks are normalized by the numbers of non-synonymous sites vs. synonymous sites in a gene, respectively, so that a Ka/Ks value of 1 represents the ratio of non-synonymous vs. synonymous mutations expected under a random mutation model and neutral amino acid selection pressure (i.e. ω=1). Since its introduction, the Ka/Ks

metric has been very widely used as a measure of amino acid level selection pressure (Nei, 2000; Yang and Bielawski, 2000; Hurst, 2002).

However, amino acid sequence is not the only feature of a gene that can experience selection pressure during evolution. Other selection processes can also act on a gene and affect its rate of evolution. In most genomes, synonymous codon usage shows bias towards to a set of major codons (Akashi and Eyre-Walker, 1998; Akashi, 2001; Akashi, 2003). Codon usage bias is proposed to reflect selection on optimal codons for a higher translational efficiency (often referred to as "translational selection"). Analyses of codon usage in the human genome indicate a weak positive correlation between gene expression level and frequencies of optimal codons (Lavner and Kotlar, 2005). The importance of codon usage bias and translational selection has been reviewed extensively elsewhere (Akashi and Eyre-Walker, 1998; Akashi, 2001; Duret, 2002; Akashi, 2003). It's well known that selection on codon usage reduces the synonymous divergence rate, and it can introduce bias into the Ka/Ks metric. Methods have been developed to adjust for such bias in measuring Ka/Ks (Hirsh et al., 2005).

In this review, we focus on a distinct mechanism of selection pressure in coding regions, which is associated with regulation of pre-mRNA splicing. We will refer to such selection pressure as "RNA selection pressure" throughout our manuscript. Recently, there is a new wave of evidence for widespread RNA selection pressure in the eukaryotic genomes (Cartegni et al., 2002; Buratti and Baralle, 2004; Fairbrother et al., 2004; Pagani and Baralle, 2004; Baek and Green, 2005; Carlini and Genut, 2005; Ohler et al., 2005; Pagani et al., 2005; Parmley et al., 2005; Xing and Lee, 2005). Sequence motifs that are important for transcript processing (e.g. splice sites or exonic splicing enhancers) are observed to experience strong selection pressure (Fairbrother et al., 2004; Pagani and Baralle, 2004; Parmley et al., 2005), even for mutations that do not alter the amino acid sequence and have no effect on translation. Many silent mutations are believed to disrupt splicing regulatory elements within exons and to be strongly selected against (Cartegni et al., 2002; Pagani and Baralle, 2004). Genome-wide analyses of alternatively spliced exons in mammals have also indicated strong RNA selection pressure associated with alternative splicing (Baek and Green, 2005; Xing and Lee, 2005). In a pivotal study, Pagani and colleagues systematically introduced 19 synonymous mutations to the exon

12 of *CFTR*, and found that ~30% of the mutations disrupted the splicing of *CFTR* and resulted in an inactive protein (Pagani et al., 2005). This study provides molecular evidence that synonymous mutations can affect splicing and are not neutral in evolution (Pagani et al., 2005). Other types of functional constraints on the RNA level, such as RNA secondary structure requirements (Chamary and Hurst, 2005), might affect synonymous sites as well. In fact, recent studies indicate that RNA secondary structure is under selection for the proper splicing of a number of genes (Buratti and Baralle, 2004; Kreahling and Graveley, 2005). It should be emphasized that RNA selection pressure (i.e. selection pressure on RNA functional features such as splicing regulatory motifs or secondary structure) is mechanistically distinct from translational selection, since the latter is associated with protein translation instead of pre-mRNA splicing. Does the presence of widespread RNA selection pressure distort the Ka/Ks metric, so that it no longer gives an accurate measure of ω?

Since amino acid selection pressure and RNA selection pressure arise from distinct cellular processes (e.g. the effect of a Ala → Val substitution on an enzyme's activity, vs. the effect of an A → G substitution on the binding of a specific splicing factor to this site), we must introduce an additional, independent variable ρ to represent the RNA selection pressure. In addition, we reinterpret the variable π to represent not the true, neutral rate of mutation (which we will now symbolize as μ), but instead as the product of the mutation rate (μ) and RNA selection pressure (ρ). Considering the three processes of 1) background nucleotide mutation μ (can act at all sites), 2) RNA selection pressure ρ (can act at all sites), and 3) amino acid selection pressure ω (can only act at non-synonymous sites), the extended definition of Ka/Ks becomes:

Non-synonymous sites: $Ka = \omega\rho\mu$

Synonymous sites: $Ks = \rho\mu$

$$Ka / Ks = \omega$$

On the one hand, this suggests that Ka/Ks remains a valid measure of amino acid level selection pressure even in the presence of RNA selection pressure. Specifically, ρ

is present in both the numerator and the denominator of the Ka/Ks metric, so even though it may cause decreases in Ks, this should not cause an increase in the Ka/Ks ratio. Because of the name "Ka/Ks", intuitively any decrease in Ks should cause an increase in Ka/Ks. However this is not the case: both of the factors ρ and μ that constitute Ks are also present in the numerator Ka, so they cancel. Indeed, that is the reason why Ks was included in the denominator of the Ka/Ks metric in the first place – to remove the effect of the background nucleotide substitution density $\pi=\rho\mu$ that is present in Ka. On the other hand, this also shows why it is hard to prove this definitively: in effect, we are trying to estimate three independent variables (ω, ρ, and μ) from only two observed values (Ka and Ks).

There are some empirical data on the relationship of Ks and Ka/Ks from genome-scale analyses. One systematic study of Ka/Ks has been performed recently by Zhang and Li, for a set of 1580 human-mouse orthologous genes (Zhang and Li, 2004). If low levels of Ks indeed cause systematic overestimation of Ka/Ks, one would expect to observe a trend of increasing Ka/Ks for genes with low values of Ks, especially when Ks is quite low (e.g. Ks<0.2). However the Ka/Ks vs. Ks data of Zhang and Li show no such a trend over this large set of genes (Fig. 1). For the reasons discussed above, this result should be considered more suggestive than conclusive.

2. *BRCA1*: Can decreases in Ks cause increased Ka/Ks?

Often, researchers who observe increased Ka/Ks values in a specific dataset check that it is due to an increase in Ka (observed amino acid changes) while expecting Ks to remain constant, consistent with its traditional interpretation as a measure of the background nucleotide substitution rate (Thomas et al., 2003; Zhang and Li, 2004). Thus, it was with some surprise that Hurst and colleagues reported a region of the *BRCA1* gene with greatly increased Ka/Ks, in which Ka was approximately the same as in the rest of the protein, but Ks was significantly decreased (Hurst and Pal, 2001). Subsequently, Orban and colleagues identified the likely cause of the decrease in Ks: this region corresponds to exon 11, which is observed to be alternatively spliced, and contains specific splicing enhancer sequences that are likely under selection pressure to maintain

the proper splicing of *BRCA1* (Orban and Olah, 2001) (Without these splicing enhancers, splicing of the exon might fail).

Hurst *et.al.* performed careful analysis seeking evidence for a mechanism by which the decrease in Ks could be considered to cause the observed increase in Ka/Ks. For example, codon usage bias might be able to produce such an appearance (which is specifically directed at silent sites and can cause reduced Ks, discussed in detail below). The authors examined several tests for codon usage bias, including the effective number of codons (ENC) (Wright, 1990) and the χ^2/L method (Shields et al., 1988). However, they observed no evidence for codon usage bias that could explain the simultaneous increase in Ka/Ks and decrease in Ks in this exon.

Recent studies of alternative splicing have shed light on this interesting observation in *BRCA1*. A variety of studies have shown that alternative splicing is associated with a significant relaxation of selection pressure (e.g. selection pressure against exon creation / loss (Modrek and Lee, 2003), tolerance for exonic Alu sequences (Sorek et al., 2002) or premature termination codons (Lewis et al., 2003; Xing and Lee, 2004), etc.). Thus, the fact that this exon is alternatively spliced in *BRCA1* would predict that it would indeed have increased Ka/Ks (reflecting relaxed selection pressure against amino acid mutations). At the same time, alternatively spliced exons often are observed to have increased conservation of *nucleotide* sequence, reflecting the importance of exonic splicing enhancers or other motifs that regulate splicing of the exon (Sorek and Ast, 2003; Itoh et al., 2004). This stronger RNA selection pressure (ρ) predicts a decrease in Ks. Thus alternative splicing can explain both observations in a simple way. Indeed, both patterns appear to be general, observed in alternatively spliced exons throughout eukaryotic genomes (Iida and Akashi, 2000; Xing and Lee, 2005).

However, these studies rely on Ka/Ks as a metric of amino acid selection pressure (ω). Is it possible that the decreased Ks values caused by RNA selection pressure (ρ) in such exons distort Ka/Ks so that it grossly overestimates the true value of ω?

3. How might RNA selection pressure bias Ka/Ks?

The potential source of bias is the possibility that the RNA selection pressure is different at synonymous vs. non-synonymous sites. For example, if the RNA selection

pressure at the first and second positions of each codon were different from the RNA selection pressure at the third codon, then measurements of RNA selection pressure at synonymous sites may not be applicable to non-synonymous sites.

To look into this possibility further, it would be helpful to briefly discuss other mechanisms that are known to cause such a systematic bias. One distinct mechanism that could cause such divergence is codon usage bias. By definition, codon usage bias only affects synonymous sites. If codon usage bias was strong enough to render a substantial fraction of synonymous mutations unacceptable, and furthermore the strength of this selection pressure varied greatly between two different samples, it could give rise to an artifactual appearance of a large difference in Ka/Ks between the samples. Defining the selection pressure due to codon usage bias in sample 1 as χ_1 and in sample 2 as χ_2,

Non-synonymous sites, sample 1: $Ka=\omega\rho\mu$

Synonymous sites, sample 1: $Ks= \chi_1\rho\mu$

Sample 1: $Ka / Ks = \omega/ \chi_1$

Non-synonymous sites, sample 2: $Ka=\omega\rho\mu$

Synonymous sites, sample 2: $Ks= \chi_2\rho\mu$

Sample 2: $Ka / Ks = \omega/ \chi_2$

In this case a large decrease in Ks in one of the samples ($Ks_2/Ks_1 = \chi_2/ \chi_1$) *can* indeed induce an artifactual increase in Ka/Ks [$(Ka/Ks)_2/(Ka/Ks)_1 = \chi_1/ \chi_2$], even when the actual amino acid selection pressure ω remains constant.

Thus it is essential to test a given dataset for evidence of strong changes in codon usage bias, when comparing Ka/Ks values with another dataset (Zhang and Li, 2004; Baek and Green, 2005). A number of metrics have been developed for measuring codon bias, including the codon adaptation index (CAI) (Sharp and Li, 1987), the effective number of codons (ENC) (Wright, 1990), the χ^2/L method(Shields et al., 1988), the

codon bias index (CBI) (Morton, 1994), the intrinsic codon deviation index (ICDI) (Freire-Picos et al., 1994), the frequency of optimal codons (Fop) (Ikemura, 1981), etc.

What are the general requirements such RNA selection pressure to exhibit systematic differences between adjacent synonymous and non-synonymous sites? To do so, RNA selection pressure would have to be periodic, with a periodicity of exactly 3 nt, and moreover would have to keep a fixed, constant relation to the protein reading frame in different exons and different genes, in order to produce a net bias that would not be averaged out when data from multiple exons were combined. As we have seen, codon usage bias can produce such an effect, but other sources of bias are also possible. It is conceivable that nucleotide sequence features such as binding sites for splicing factors or other proteins could be "phased" to position their most constrained nucleotides in synonymous sites, and to avoid non-synonymous sites, so that the separate requirements of the amino acid sequence and nucleotide sequence could both be fulfilled without interfering with each other.

4. Empirical data on "synonymous phasing" of splicing factor sites

Some recent studies have shed light on how RNA selection pressure might affect Ka/Ks. Krainer and colleagues have characterized the binding sites for the splicing factors SF2/ASF, SC35, SRp40 and SRp55 by RNA SELEX (Systematic Evolution of Ligands by EXponential enrichment) experiments, and have constructed a position specific scoring matrix for each motif (Liu et al., 1998; Cartegni et al., 2003) (also see: http://rulai.cshl.edu/tools/ESE/ESEmatrix.html). These data indicate the maximum possible effect that ESE motifs could have on synonymous sites, relative to non-synonymous sites. For example, for the SF2/ASF motif, positioning its strongly constrained G (position 6) at a synonymous site yields a 54% reduction of substitution probability at synonymous sites relative to nonsynonymous sites. The maximum reductions for the SC35, SRp40 and SRp55 motifs were 24%, 21%, and 21% respectively. Thus in theory SF2/ASF motifs could cause an up to 2-fold Ka/Ks artifact (i.e. an actual value of ω=0.1 might be measured as a Ka/Ks value of 0.2). However, this maximum can only be achieved if 1) every SF2/ASF motif is positioned so its G at position 6 is in a synonymous site; 2) every synonymous site is four-fold redundant,

which is not the case at all sites; 3) every nucleotide in the region being measured is covered by an SF2/ASF motif (i.e. a minimum of 21 SF2/ASF 7nt motifs to cover a 150nt exon). The effects of realistic numbers of such motifs over a typical exon or entire protein sequence would be much less.

A handful of other studies suggest that experimentally verified SF2/ASF sites do not exhibit such strong phasing. Studies by Dirksen et al. (2000), Pollard et al. (2002), and Rooke et al. (2003) have demonstrated the presence of six functional SF2/ASF sites in three genes (GH1 (Dirksen et al., 2000), GNAS (Pollard et al., 2002), and c-Src (Rooke et al., 2003)). Their data show that 3 of the 6 sites place the conserved G at the 1st codon position (non-synonymous), while the remaining three sites put it at the 3rd (synonymous) codon position (but even these sites were only two-fold redundant, halving their impact on Ks).

5. Challenges: more experimental data are needed for validated splicing factor sites

It would be useful to test whether important RNA regulatory motifs (such as exonic splicing enhancers or silencers) exhibit a strong 3nt periodicity for their strongly constrained nucleotide positions, and whether there is a significant bias for these positions to be matched to synonymous sites (and to avoid non-synonymous sites). To measure this bias in a meaningful way requires a “gold standard” set of experimentally validated nucleotide motifs; i.e. motif sites shown to be biologically active in real genes. The general weakness of such motif sequence patterns leads to very high false positive rates when searching genomic sequence, so each candidate site would have to be experimentally tested for its putative activity (e.g. via site-directed mutagenesis) (Cartegni et al., 2003). This question cannot really be resolved solely by bioinformatics, and would require a significant experimental effort. Great care would be required in the design of the motif identification analysis and experimental validation, to avoid introducing possible bias. For example, validating motifs by making site-directed mutations at synonymous sites would clearly be biased, as would any validation protocol that depended on measuring the activity of the protein product.

6. Concluding remarks

Recent studies of RNA splicing indicate widespread RNA selection pressure in eukaryotic genomes (Pagani and Baralle, 2004). This challenges a basic assumption of the Ka/Ks metric, that synonymous substitutions are selectively neutral. Is Ka/Ks still an appropriate metric for selection pressure on amino acid substitutions? If RNA selection pressure indeed introduces systematic bias into the Ka/Ks metric, what is the magnitude of the bias? Much additional work is needed on the distribution of experimentally validated RNA regulatory motifs in real genes.

Acknowledgements

The authors wish to thank Drs. Sean Eddy, Mark Gerstein, Phil Green, Laurence Hurst, Fyodor Kondrashov, and Ziheng Yang for discussions on Ka/Ks, and Dr. Peter Stoilov for discussions on exonic splicing enhancers. In particular, Phil Green and Peter Stoilov suggested the idea that RNA motifs might be "phased" to avoid non-synonymous sites.

Figure Legends

Figure 1: *Ka/Ks and Ks for 1580 human-mouse orthologs*

Ka/Ks and Ks values were obtained from a recent analysis (Zhang and Li, 2004) on 1580 housekeeping and tissue-specific genes between human and mouse. The data doesn't indicate increased Ka/Ks for genes with low Ks.

References

Akashi, H., 2001. Gene expression and molecular evolution. Curr Opin Genet Dev 11, 660-6.
Akashi, H., 2003. Translational selection and yeast proteome evolution. Genetics 164, 1291-303.
Akashi, H. and Eyre-Walker, A., 1998. Translational selection and molecular evolution. Curr Opin Genet Dev 8, 688-93.
Baek, D. and Green, P., 2005. Sequence conservation, relative isoform frequencies, and nonsense-mediated decay in evolutionarily conserved alternative splicing. Proc Natl Acad Sci U S A 102, 12813-8.
Buratti, E. and Baralle, F.E., 2004. Influence of RNA secondary structure on the pre-mRNA splicing process. Mol Cell Biol 24, 10505-14.
Carlini, D.B. and Genut, J.E., 2005. Synonymous SNPs Provide Evidence for Selective Constraint on Human Exonic Splicing Enhancers. J Mol Evol.
Cartegni, L., Chew, S.L. and Krainer, A.R., 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3, 285-98.
Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q. and Krainer, A.R., 2003. ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res 31, 3568-71.
Chamary, J.V. and Hurst, L.D., 2005. Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. Genome Biol 6, R75.
Dirksen, W.P., Li, X., Mayeda, A., Krainer, A.R. and Rottman, F.M., 2000. Mapping the SF2/ASF binding sites in the bovine growth hormone exonic splicing enhancer. J Biol Chem 275, 29170-7.
Duret, L., 2002. Evolution of synonymous codon usage in metazoans. Curr Opin Genet Dev 12, 640-9.
Fairbrother, W.G., Holste, D., Burge, C.B. and Sharp, P.A., 2004. Single Nucleotide Polymorphism-Based Validation of Exonic Splicing Enhancers. PLoS Biol 2, E268.
Freire-Picos, M.A., Gonzalez-Siso, M.I., Rodriguez-Belmonte, E., Rodriguez-Torres, A.M., Ramil, E. and Cerdan, M.E., 1994. Codon usage in Kluyveromyces lactis and in yeast cytochrome c-encoding genes. Gene 139, 43-9.
Hirsh, A.E., Fraser, H.B. and Wall, D.P., 2005. Adjusting for selection on synonymous sites in estimates of evolutionary distance. Mol Biol Evol 22, 174-7.
Hurst, L.D., 2002. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet 18, 486.
Hurst, L.D. and Pal, C., 2001. Evidence for purifying selection acting on silent sites in BRCA1. Trends Genet 17, 62-5.
Iida, K. and Akashi, H., 2000. A test of translational selection at 'silent' sites in the human genome: base composition comparisons in alternatively spliced genes. Gene 261, 93-105.
Ikemura, T., 1981. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. J Mol Biol 151, 389-409.

Itoh, H., Washio, T. and Tomita, M., 2004. Computational comparative analyses of alternative splicing regulation using full-length cDNA of various eukaryotes. Rna 10, 1005-18.

Kreahling, J.M. and Graveley, B.R., 2005. The iStem, a long-range RNA secondary structure element required for efficient exon inclusion in the Drosophila Dscam pre-mRNA. Mol Cell Biol 25, 10251-60.

Lavner, Y. and Kotlar, D., 2005. Codon bias as a factor in regulating expression via translation rate in the human genome. Gene 345, 127-38.

Lewis, B.P., Green, R.E. and Brenner, S.E., 2003. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci U S A 100, 189-92.

Li, W.H., Wu, C.I. and Luo, C.C., 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.

Liu, H.X., Zhang, M. and Krainer, A.R., 1998. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. Genes Dev 12, 1998-2012.

Modrek, B. and Lee, C., 2003. Alternative splicing in the human, mouse and rat genomes is associated with an increased rate of exon creation / loss. Nature Genet. 34, 177-180.

Morton, B.R., 1994. Codon use and the rate of divergence of land plant chloroplast genes. Mol Biol Evol 11, 231-8.

Nei, M. and Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3, 418-26.

Nei, M., Kumar,S., 2000. Molecular evolution and phylogenetics. Oxford University Press, Oxford,UK.

Ohler, U., Shomron, N. and Burge, C.B., 2005. Recognition of unknown conserved alternatively spliced exons. PLoS Comput Biol 1, 113-22.

Orban, T.I. and Olah, E., 2001. Purifying selection on silent sites -- a constraint from splicing regulation? Trends Genet 17, 252-3.

Pagani, F. and Baralle, F.E., 2004. Genomic variants in exons and introns: identifying the splicing spoilers. Nat Rev Genet 5, 389-96.

Pagani, F., Raponi, M. and Baralle, F.E., 2005. Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. Proc Natl Acad Sci U S A 102, 6368-72.

Parmley, J.L., Chamary, J.V. and Hurst, L.D., 2005. Evidence for Purifying Selection Against Synonymous Mutations in Mammalian Exonic Splicing Enhancers. Mol Biol Evol.

Pollard, A.J., Krainer, A.R., Robson, S.C. and Europe-Finner, G.N., 2002. Alternative splicing of the adenylyl cyclase stimulatory G-protein G alpha(s) is regulated by SF2/ASF and heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1) and involves the use of an unusual TG 3'-splice Site. J Biol Chem 277, 15241-51.

Rooke, N., Markovtsov, V., Cagavi, E. and Black, D.L., 2003. Roles for SR proteins and hnRNP A1 in the regulation of c-src exon N1. Mol Cell Biol 23, 1874-84.

Sharp, P.M. and Li, W.H., 1987. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15, 1281-95.

Shields, D.C., Sharp, P.M., Higgins, D.G. and Wright, F., 1988. "Silent" sites in Drosophila genes are not neutral: evidence of selection among synonymous codons. Mol Biol Evol 5, 704-16.

Sorek, R. and Ast, G., 2003. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. Genome Res 13, 1631-1637.

Sorek, R., Ast, G. and Graur, D., 2002. Alu-containing exons are alternatively spliced. Genome Res 12, 1060-7.

Thomas, M.A., Weston, B., Joseph, M., Wu, W., Nekrutenko, A. and Tonellato, P.J., 2003. Evolutionary dynamics of oncogenes and tumor suppressor genes: higher intensities of purifying selection than other genes. Mol Biol Evol 20, 964-8.

Wright, F., 1990. The 'effective number of codons' used in a gene. Gene 87, 23-9.

Xing, Y. and Lee, C., 2004. Negative selection pressure against premature protein truncation is reduced by both alternative splicing and diploidy. Trends Genet 20, 472-5.

Xing, Y. and Lee, C., 2005. Colloquium Paper: Evidence of functional selection pressure for alternative splicing events that accelerate evolution of protein subsequences. Proc Natl Acad Sci U S A 102, 13526-31.

Yang, Z. and Bielawski, J.P., 2000. Statistical methods for detecting molecular adaptation. Trends in Ecology and Evolution 15, 496-503.

Yang, Z. and Nielsen, R., 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17, 32-43.

Zhang, L. and Li, W.H., 2004. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. Mol Biol Evol 21, 236-9.

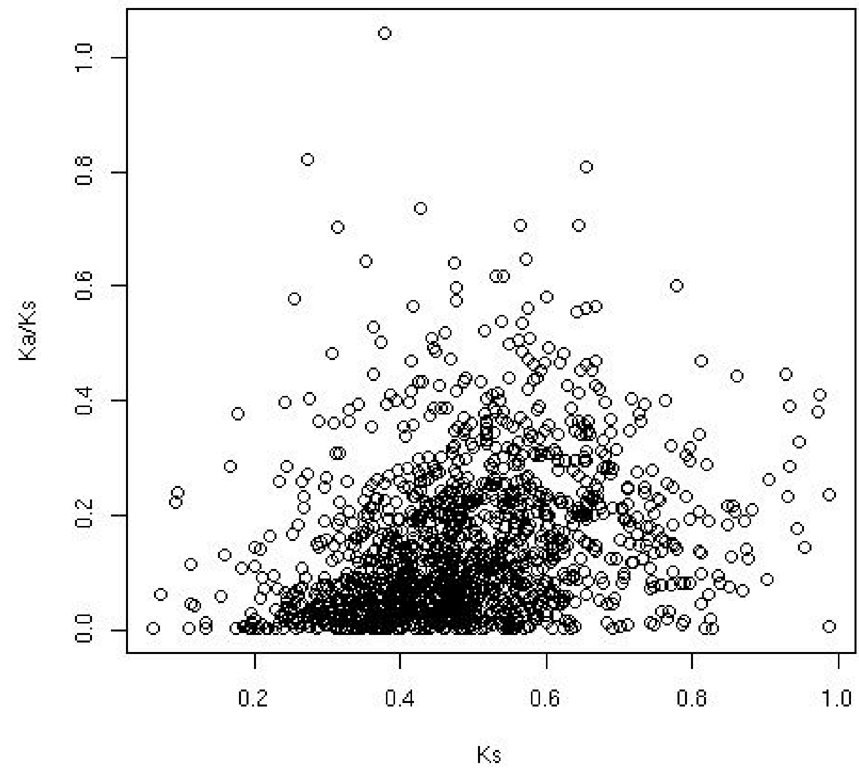

FIGURE 1