Changes in Alternative Splicing of Human and Mouse Genes Are Accompanied by Faster Evolution of Constitutive Exons

Brian P. Cusack and Kenneth H. Wolfe

Department of Genetics, Smurfit Institute, University of Dublin, Trinity College, Dublin, Ireland

Alternative splicing is known to be an important source of protein sequence variation, but its evolutionary impact has not been explored in detail. Studying alternative splicing requires extensive sampling of the transcriptome, but new data sets based on expressed sequence tags aligned to chromosomes make it possible to study alternative splicing on a genome-wide scale. Although genes showing alternative splicing by exon skipping are conserved as compared to the genome as a whole, we find that genes where structural differences between human and mouse result in genome-specific alternatively spliced exons in one species show almost 60% greater nonsynonymous divergence in constitutive exons than genes where exon skipping is conserved. This effect is also seen for genes showing species-specific patterns of alternative splicing where gene structure is conserved. Our observations are not attributable to an inherent difference in rate of evolution between these two sets of proteins or to differences with respect to predictors of evolutionary rate such as expression level, tissue specificity, or genetic redundancy. Where genome-specific alternatively spliced exons are seen in mammals, the vast majority of skipped exons appear to be recent additions to gene structures. Furthermore, among genes with genome-specific alternatively spliced exons, the degree of nonsynonymous divergence in constitutive sequence is a function of the frequency of incorporation of these alternative exons into transcripts. These results suggest that alterations in alternative splicing pattern can have knock-on effects in terms of accelerated sequence evolution in constant regions of the protein.

Introduction

Proteome diversity is expanded by the twin evolutionary engines of gene duplication and alternative splicing. Completion of whole-genome sequences for a range of eukaryotes has revealed the pervasiveness of gene duplication in evolution. However, an appreciation of the prevalence of alternative splicing has had to await deeper sampling of the transcriptome. Alternative pre-mRNA splicing enables a single gene to encode many different mature mRNA transcripts and potentially several different protein products. Estimates of the fraction of alternatively spliced human genes have increased as expressed sequence (expressed sequence tag [EST] and cDNA) databases have grown (Kan, States, and Gish 2002; Boue, Letunic, and Bork 2003) and with the development of new technologies such as exon junction microarrays (Johnson et al. 2003). Current estimates are that at least 70% of human multiexon genes are alternatively spliced (Johnson et al. 2003). At the same time both the fraction of genes that are alternatively spliced and the number of isoforms generated per gene, appear to be roughly constant over a broad phylogenetic range of metazoa (Brett et al. 2002; Harrington et al. 2004).

Alternative splicing may result in exon truncation or extension, intron retention, or the inclusion/exclusion of entire exons by exon skipping. Different protein isoforms encoded by a single gene are likely to be variants of a constant protein backbone with the addition or deletion of entire alternative domains (Kriventseva et al. 2003). This enables some alternative isoforms to encode distinct functions, as has been demonstrated for transmembrane domains and protein-protein interactions (Xing, Xu, and Lee 2003; Resch et al. 2004b).

A definitive catalog of the types of alternative splicing occurring in a given organism would require both extensive transcriptome sampling and a finished genome sequence (Modrek and Lee 2002). These requirements are closest to being fulfilled in humans and in mouse. Mapping ESTs onto genomic sequence (Modrek and Lee 2002) reduces the contaminating effect of mixing paralogous sequences and other EST artifacts and allows alternatively spliced variants to be assigned to specific gene structures. This genomic-confirmation approach was used to create the Alternative Splicing Annotation Project (ASAP) database which provides a high-quality platform for the annotation of alternative splicing in humans and mouse (Lee et al. 2003).

Despite our growing appreciation of the incidence of alternative splicing in the generation of protein diversity, little is known about its evolutionary impact (Kopelman, Lancet, and Yanai 2005). This contrasts with the depth of research into the evolutionary impact of the other major mechanism of proteome expansion, gene duplication (Lynch and Conery 2000; Kondrashov et al. 2002; Nembaware et al. 2002). In a key study, Modrek and Lee described an association between alternative splicing and changes in the exon-intron structure of orthologous mouse and human genes resulting from lineage-specific gain or loss of exons (Modrek and Lee 2003). This work suggested that alternative splicing may be used as a mechanism for evolution to try incorporating novel exons into a minority of a gene's transcripts (so-called "minor-form" transcripts). Because the gene's ancestral function is maintained by the "major-form" transcripts, this may free the minor-form transcript from functional constraint, thus reducing purifying selection. This situation can be likened to the relaxation of constraints on recent gene duplicates, and for this reason minor transcripts generated by alternative splicing have been termed "internal paralogs" (Modrek and Lee 2003). Evidence for relaxed selection on alternatively spliced sequence regions includes the observations that Alu-containing exons are always alternatively spliced (Sorek, Ast, and Graur 2002; Xing and Lee 2004) and that a larger proportion of minor-form transcripts contain premature termination codons (PTCs) (Xing and Lee 2004). Furthermore, it has recently been shown that

alternatively spliced exons themselves show relaxation on sequence constraint with respect to amino acid substitutions (Xing and Lee 2005).

The model of Modrek and Lee (2003) predicts that relaxation of selective constraint on the minor-transcript isoform will result in faster evolution of the alternatively spliced exon alone, but it makes no predictions about constraints on constitutively translated parts of the gene. Here we investigated whether the generation of an internal paralog through alternative splicing has an impact on selection operating on the entire gene. We considered only alternative splice events in humans and mouse that result from exon skipping and distinguished between conserved alternative splicing and alternative splicing that is specific to either humans or mouse. We show that these "genome-specific" alternatively spliced exons appear to be the result of exon gains following the human-mouse split. We find that although genes showing alternative splicing by exon skipping tend to be slowly evolving, the possible impact of change in alternative splice pattern is acceleration of sequence evolution in the entire gene. Notably this acceleration is detected in constitutive exon sequence and may be a consequence of amino acid substitutions correlated with the accommodation of an alternatively spliced exon.

Methods
Human-Mouse Exon-Skip Conservation

We downloaded the ASAP data set (Lee et al. 2003) from http://www.bioinformatics.ucla.edu/ASAP/ in December 2003. This data set includes conservation information for alternatively spliced exons (i.e., exon-skip events) in human and mouse genes assigned as orthologous using Homologene data (Wheeler et al. 2004). Conservation of an exon-skip event is recorded first with respect to sequence conservation of the alternatively spliced exon in the genomic DNA of the ortholog and second by determining whether expressed sequence information supports both the inclusion and exclusion of the homologous exon from transcripts in the second species (transcriptomic evidence of alternative splicing of the exon). We defined conserved alternatively spliced exons as those having transcriptomic evidence of alternative splicing in both species. We defined an alternatively spliced exon as "genome specific" when there is transcriptomic evidence for its alternative splicing in one species but no genomic evidence for its presence in the other species.

Orthology Mapping

ASAP lists the UniGene identifiers of human and mouse genes. We extracted the Human Genome Organisation (HUGO) gene name for each human UniGene identifier (ID) in ASAP and mapped this to unique human and mouse LocusLink IDs using Homologene (two versions: December 2003, January 2004). LocusLink IDs were then used as queries for the Ensmart tool (http://www.ensembl.org/Multi/martview) to obtain the associated human (NCBI build 34) and mouse (NCBIM build 32) Ensembl gene names and predicted protein and transcript sequences. Linking to Ensembl using direct Homologene information in this way yielded the sequences of 224 pairs of orthologs.

For UniGene IDs that we could not map to recent versions of Homologene, we used Ensmart to map the human gene name to a human Ensembl gene identifier and used reciprocal best BlastP (Altschul et al. 1997) to assign a mouse ortholog. Sequences for a further 56 pairs of orthologs were derived in this step. For those genes that could not be linked to Ensembl via either Homologene or human gene name, we used high-stringency BlastP to map a translation product inferred by the Alternatively Spliced Proteins Database (ASP) (Xing, Resch, and Lee 2004) for each gene to a human Ensembl-predicted protein, followed by a reciprocal best BlastP to assign a mouse Ensembl ortholog. This step found an additional 93 pairs of orthologs. Finally, we used BlastN to verify all assignments of genes to Ensembl IDs by ensuring that the Ensembl-predicted transcript for a given gene matched its sequence derived from ASAP.

Identification of "Representative Orthologs" in Fish

For each gene showing either conserved alternative splicing or genome-specific alternative splicing we used the human protein as query to detect a reciprocal best hit in fugu and in zebrafish with $E < 1 \times 10^{-10}$ and requiring a coverage of at least 50% of the longer sequence. Each pair of fugu and zebrafish orthologs identified is a representative ortholog pair (Davis and Petrov 2004) belonging to one of two categories, one representing the evolution of genes for which alternative splicing is conserved between humans and mouse and the other representing the evolution of genes with genome-specific alternative splicing where the patterns of alternative splicing differ between humans and mouse.

Assessing Levels of Selective Constraint

Human and mouse protein sequences were aligned using ClustalW (Thompson, Higgins, and Gibson 1994) and back-translated to generate a codon-based alignment of transcripts. For each gene we used ASAP annotations to extract the sequences of exons that undergo alternative splicing by exon skipping. Parts of the transcript alignment corresponding to these exons were masked. We calculated K_a and K_s for the unmasked (constitutive) sequence using the yn00 program in the PAML package (Yang 1997). For fugu-zebrafish representative orthologs, K_a and K_s were calculated based on the entire transcript alignments because no information was available on alternative splicing of exons in these organisms.

Determining Alternatively Spliced Exon Presence/ Absence in the Human-Mouse Ancestor

We used chicken as an out-group to determine whether a given alternatively spliced exon was present in the human-mouse ancestor. Three strategies were employed to detect homologs of human alternatively spliced exons in either the chicken genome or transcriptome. First, translations of the alternatively spliced exon sequence plus 90 nt from each flanking exon were used as TBlastN queries against the chicken genome. These were required to hit a stretch of chicken chromosome having an "anchoring" TBlastN

match ($E \leq 1 \times 10^{-5}$) to the Ensembl-predicted translation of the human gene. The flanking sequence was used to further "anchor" hits to the chromosome, and only hits in which at least two-thirds of both flanks were aligned with $\geq 50\%$ amino acid similarity were scored. The alternatively spliced exon was scored as "detected" if at least two-thirds of its length was aligned with $\geq 50\%$ similarity or as "not detected" otherwise. Second, the alternatively spliced exon sequence alone was used as a BlastN query against chicken ESTs. Alternatively spliced exons with $\geq 80\%$ of their length aligned, $\geq 70\%$ identity, and $E < 0.001$ were scored as detected, or not detected otherwise. The third approach involved a "low-stringency" search strategy that did not require the detection of conservation of the alternatively spliced exon sequence itself. The alternatively spliced exon sequence plus 90 nt from each flanking exon was used as a BlastN query against chicken ESTs. Only hits in which $\geq 50\%$ of both flanks were aligned were scored. The alternatively spliced exon was scored as detected if the intervening EST sequence between the aligned flanks was ≥ 10 nt or not detected otherwise. Finally, to identify "high-confidence" cases where we expect to see a chicken homolog for an alternatively spliced human exon if it exists we applied a binomial test as outlined in Kan, States, and Gish (2002). We only scored those alternatively spliced exons having sufficiently high splicing frequency in humans and for which chicken ortholog EST coverage is deep enough that we expect to detect chicken homologs of these exons.

Influence of Frequency of Incorporation of Alternatively Spliced Sequence

For each human genome–specific alternatively spliced exon classified as translated and incorporated into productive transcripts according to ASP annotation, we counted the number of ESTs in ASAP supporting each of the two alternative splices: inclusion and exclusion of the exon. We performed the binomial test employing two threshold cutoffs (we chose $t = 0.03$ and $t = 0.12$ to produce roughly equally sized subdivisions of the data) to categorize each exon skip as belonging to one of the three frequency classes as used in figure 2. For example, an alternatively spliced exon whose inclusion frequency satisfied the binomial test at the 95% confidence level (ie., $P < 0.05$) for the lower threshold frequency ($t = 0.03$) but not for the upper threshold frequency ($t = 0.12$) was classified as incorporated at medium frequency. We compared each frequency category of human genome–specific alternatively spliced exons with respect to the nonsynonymous divergence (K_a) calculated for constitutive sequence in the cognate gene.

Level and Breadth of Constitutive Exon Expression

The expression level of the constitutive exons of each gene was approximated using a simple count of all EST/cDNA sequences mapped to that gene by ASAP. Breadth of expression was determined by assigning each EST/cDNA to one of 34 tissue classes using TissueInfo (Skrabanek and Campagne 2001).

Estimating Adequacy of Mouse EST Sampling in Genes with Putatively Human-Specific Alternative Splicing

Genes with putatively human-specific alternative splicing of conserved exons were identified as those where gene structure is conserved in humans and mouse but where alternative splicing is only observed in humans. For each gene in this group we determined the number of mouse ESTs as a fraction of the number of human ESTs sampled. This scaling gives a measure of how adequate mouse EST coverage should be in recovering any conserved alternative splicing events under the assumption that alternative splicing occurs with equal frequency in humans and mouse. We considered three different measures of depth of EST coverage by counting: (1) all human and mouse ESTs assigned to a gene; (2) human and mouse ESTs from a set of named tissues only (this excludes ESTs from cancerous sources); and (3) human and mouse ESTs from the tissue(s) in which the putatively human-specific splice event (exon inclusion or skipping) is observed.

Results
Genes Showing Exon Skipping Are More Conserved than the Genome Average

We downloaded a set of 14,596 human-mouse orthologs with assigned gene names from Ensembl and classified them as either exhibiting exon skipping (6,580 genes) or as having no evidence of exon skipping ("control set" of 8,016 genes) based on ASAP annotation (Lee et al. 2003). We compared sequence constraint in the alternatively spliced genes to that of genes in the control set. A commonly used measure of the degree of evolutionary constraint on a sequence is the ratio of nonsynonymous substitutions per nonsynonymous site (K_a) to synonymous substitutions per synonymous site (K_s). For values of K_a/K_s ≤ 1, this ratio is generally highest for genes whose sequences are weakly constrained by purifying selection. However, in the case of alternatively spliced sequence the K_a/K_s ratio has to be interpreted with greater caution because the inherent assumption that silent sites in codons are selectively neutral is more likely to be incorrect. The presence of exonic splicing enhancer (ESE) motifs in alternatively spliced exons means that nucleotide changes that disrupt these motifs are likely to be detrimental to function and are therefore subject to purifying selection (Iida and Akashi 2000; Orban and Olah 2001). It is not known if the constraint imposed by ESE motifs is of equal strength at synonymous and nonsynonymous sites or whether these motifs have evolved to have minimal impact on the encoded amino acid sequence. For this reason the usefulness of K_a/K_s as a measure of selective constraint on alternatively spliced exons is uncertain (but see Xing and Lee 2005).

We found that genes undergoing alternative splicing by exon skipping were more constrained than the control set (table 1). We could also compare K_a for these human-mouse orthologs because they all share a common divergence time. The slower evolution of alternatively spliced genes relative to the genome average is equally striking when we consider K_a alone (table 1). The observed differences are made more conservative by the fact that

Table 1
Medians and Standard Deviations of K_a/K_s and K_a from Human/Mouse Orthologs Showing Alternative Splicing (AS) by Exon Skipping and from a Control Set of Human/Mouse Orthologs for Which No Exon Skipping Has Been Described

	N	K_a/K_s	K_a
AS genes	6,580	0.089 (0.123)	0.053 (0.085)
Control genes	8,016	0.117 (0.136)	0.071 (0.096)
		$P < 1 \times 10^{-15}$	$P < 1 \times 10^{-15}$

NOTE.—Significance was tested using a two-tailed Wilcoxon rank-sum test.

Table 2
Medians and Standard Deviations of K_a and K_a/K_s from Orthologous Comparisons for Genes with Alternative Splicing (AS) Conserved Between Humans and Mouse Compared to Genes with Genome-Specific Alternative Splicing in Humans or Mouse

	K_a	K_a/K_s	N
All exons			
Conserved AS	0.046 (0.065)	0.086 (0.099)	68
Genome-specific AS	0.061 (0.089)	0.108 (0.140)	286
	$P < 0.050$	$P < 0.050$	
Constitutive exons			
Conserved AS	0.046 (0.068)	0.075 (0.108)	66
Genome-specific AS	0.060 (0.091)	0.111 (0.148)	285
	$P < 0.050$	$P = 0.055$	
Productive AS, constitutive exons			
Conserved AS	0.049 (0.063)	0.075 (0.091)	51
Genome-specific AS	0.078 (0.102)	0.130 (0.182)	93
	$P < 0.005$	$P < 0.01$	
Fish representative orthologs[a]			
Conserved AS	0.134 (0.108)	0.058 (0.055)	30
Genome-specific AS	0.160 (0.161)	0.061 (0.070)	124
	$P > 0.1$	$P > 0.1$	

NOTE.—Significance was tested using a two-tailed Wilcoxon rank-sum test.
[a] K_a and K_a/K_s for fugu versus zebrafish orthologs of mammalian genes that show conserved AS or genome-specific AS in humans or mouse.

estimates of both K_a and K_a/K_s for alternatively spliced exons are higher than for constitutive exons (Iida and Akashi 2000; Xing and Lee 2005).

Genome-Specific Alternative Splicing Is Associated with Faster Protein Evolution and Weaker Selective Constraint in Constitutive Regions

The level of constraint on a protein sequence is likely to differ according to the protein's function. If genes in different functional categories employ alternative splicing to different extents, this could explain why alternatively spliced genes are conserved compared to the genome as a whole. To test the influence of functional bias we focused only on genes that undergo alternative splicing through exon skipping and classified them as showing either (1) exon skipping conserved between humans and mouse or (2) genome-specific exon skipping where the alternatively spliced exon is found in genomic DNA of one species only. The latter group was defined based on the failure of a BlastN search to detect the alternatively spliced exon in genomic DNA of the ortholog. This fact coupled with lack of evidence for a homologous exon among ESTs in the second species indicates unambiguously that the alternatively spliced exon, and therefore the alternative splicing event, is genome specific. Throughout this paper, we use the terms "human genome–specific alternative splicing" and "mouse genome–specific alternative splicing" to denote alternative splicing events that are specific to one genome and where the other genome has no ortholog of the alternatively spliced exon. We confirmed using GOstat (Beissbarth et al. 2004) that although genes with exon skips are biased toward certain functional terms, there was no difference in the functions performed by genes with conserved alternatively spliced exons and genes with genome-specific alternatively spliced exons (data not shown). The set of genes with conserved alternatively spliced exons therefore serves as a function-matched control for comparison to genes with genome-specific alternatively spliced exons. Although the classification of genes in these two categories is unambiguous, it should be noted that the groups differ in their degree of gene-structure conservation. This potential source of bias was assessed in a later test comparing genes with putatively species-specific alternative splicing patterns but conserved gene structures (see the final section of *Results*).

Using human/mouse orthologs for the two groups we find that K_a in genes with genome-specific alternatively spliced exons is 33% greater than in genes with conserved alternative splicing ($K_a = 0.061$ vs. 0.046; table 2). There is

also a comparable difference in the K_a/K_s ratio. However, a strict comparison between these groups requires us to account for the possibility of differing selective pressures on alternatively spliced exons compared to constitutive exons (Xing and Lee 2005). For genes with conserved exon skipping the conserved alternatively spliced exon is included in the human-mouse alignment and thus contributes to the calculation of K_a and K_s but this is not the case for genome-specific alternatively spliced exons. Omitting the sequence of all alternatively spliced exons from all genes had a negligible effect on the calculated values of K_a, but for genes with conserved alternative splicing it reduced the estimate of K_a/K_s as expected (table 2). Thus, when we consider constitutive sequence only we see that K_a/K_s is 48% greater in genes with genome-specific alternative splicing than in genes where alternative splicing is conserved. This result suggests that there is an association between changes in a gene's alternative splicing pattern and an increase in the rate of sequence evolution in the constant part of the protein.

Productive Alternative Splicing

The inclusion of an alternatively spliced exon may induce a frameshift and introduce a PTC into the transcript resulting in transcript degradation by nonsense-mediated decay (NMD) (Nagy and Maquat 1998). Although alternative splicing–coupled NMD can have a regulatory role, these alternative splicing events do not increase the gene's protein-coding potential, and we therefore consider them "unproductive." Notably, a greater proportion of nonconserved alternatively spliced exons induces frameshifts than conserved alternatively spliced exons, and many of these are likely to initiate NMD (Sorek, Shamir, and Ast 2004).

We used the ASP database (Xing, Resch, and Lee 2004) of predicted alternatively spliced transcript sets

inferred from EST and cDNA data for a given human gene and repeated our analysis, excluding all cases generating transcripts with a PTC. We further focused on those cases in which the alternatively spliced exon overlaps the reading frame of the transcript. This restricts the analysis to genes undergoing productive alternative splicing that is likely to generate a distinct protein product. Because the ASP database currently contains only inferred transcripts from humans we compared genes with conserved alternative splicing in humans and mouse (51 genes) to those with human genome–specific alternative splicing (93 genes), considering only productive alternative splicing in both cases. The distributions of K_a and K_a/K_s for constitutive exons from human/mouse orthologous comparisons for both sets of genes are shown in figure 1. Genes with human genome–specific alternative splicing showed a 59% increase in median K_a ($P < 0.005$) and a 73% increase in median K_a/K_s ($P < 0.01$) in their constitutive sequence when compared to genes with conserved alternative splicing (table 2). Therefore, the observed increase in evolutionary rate in genes undergoing genome-specific alternative splicing holds for productive alternative splicing events. It is important to note that this increase in K_a was observed in constitutive exons and is distinct from the acceleration reported in alternatively spliced exons (Xing and Lee 2005).

Differences in Strength of Selective Constraint in Mammals Are Not a Reflection of Inherent Constraint Differences

The difference in substitution rates associated with conserved versus genome-specific alternative splicing may lie in an inherent difference between these two classes of genes. Genes under relaxed selective constraint may be more liable both to change their gene structure by gaining or losing an alternatively spliced exon and to have faster rates of sequence evolution.

We addressed this issue by examining the substitution rates in genes independently of the effects of changes in alternative splicing that have emerged during the course of mammalian evolution, by using the "representative orthologs" method of Davis and Petrov (2004). For each pair of human/mouse orthologs we searched for fugu and zebrafish orthologs. We calculated the divergence between the two fish species for two groups of genes, according to whether their mammalian orthologs showed conserved alternative splicing or genome-specific alternative splicing. In contrast to the differences seen for the mammalian genes, we found no significant difference in K_a or K_a/K_s between the two groups of fish orthologs (table 2). This is partly a consequence of the smaller size of the representative ortholog sample because we did detect an increase in K_a in fish orthologs for genes that show genome-specific alternative splicing in mammals, but this was less dramatic than the increase seen in the study orthologs (table 2). However, because no difference was seen in K_a/K_s in the fish comparison we conclude that there is no inherent difference in selective constraint between the two classes of alternatively spliced gene. In addition, this result suggests that a simple sampling bias does not underlie the difference we observe between these two classes of genes in mammals.
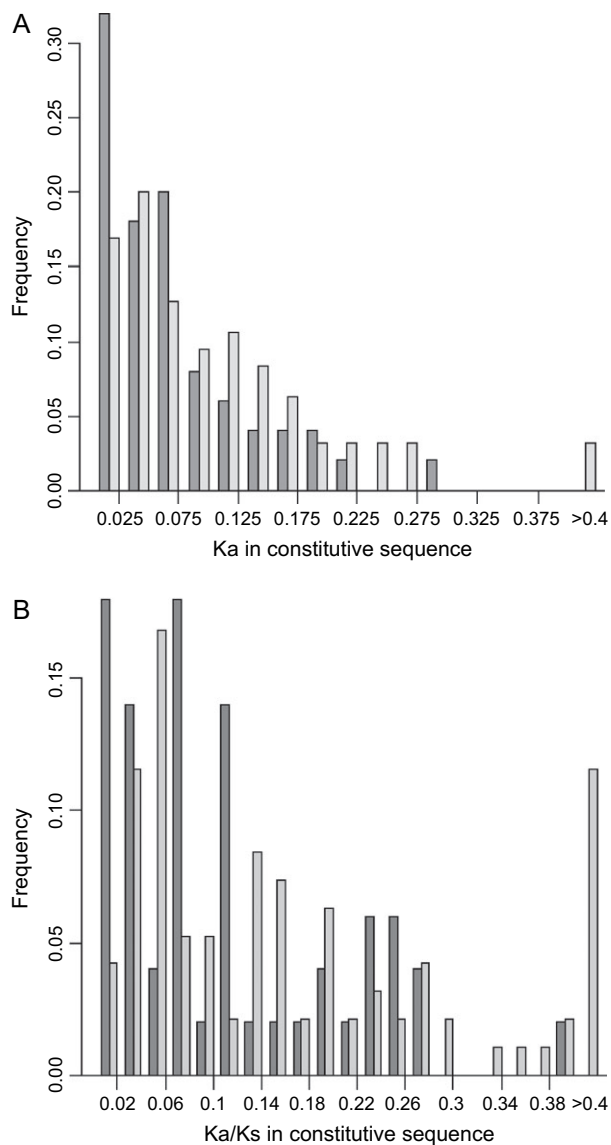

Fig. 1.—Distributions of (A) K_a and (B) K_a/K_s for constitutive exons from human/mouse orthologous comparisons of 51 genes with conserved alternative splicing (dark gray) and 93 genes with human genome–specific alternative splicing (light gray). All alternative splicing events overlap the open reading frame and generate productive transcripts without PTCs.

Genes That Have Changed in Alternative Splicing Pattern Have Also Undergone Changes in K_a/K_s Ratio

We next looked for indications that changes in alternative splicing pattern have resulted in changes in selective constraint on a gene during the course of mammalian evolution. For a given gene, comparing the K_a/K_s ratio for human/mouse to that for fugu/zebrafish gives an indication of any change in the strength of selective constraint operating on that gene. We considered only those genes for which we were able to calculate K_a/K_s for both the mammal and the fish species pairs. We found that of 124 genes showing genome-specific alternative splicing (i.e., either alternative splicing of an exon in humans but not in mouse or vice versa), 77 had higher K_a/K_s in mammals than in fish

($P = 0.003$, binomial test). In contrast, of 29 genes with alternative splicing conserved between humans and mouse, only 15 had higher K_a/K_s in mammals than in fish ($P = 0.355$, binomial test). This test is conservative because it ignores the magnitude of the difference in K_a/K_s. We therefore performed a second comparison considering the distributions of K_a/K_s from human/mouse and fugu/zebrafish orthologs. For the 29 genes in which alternative splicing is conserved between humans and mouse, K_a/K_s did not differ significantly in the cross-taxon comparison between mammals (median $K_a/K_s = 0.066$) and fish (median $K_a/K_s = 0.055$) (Wilcoxon rank-sum test $P = 0.97$) (table 2). On the other hand, the 124 genes showing alternative splicing in humans but not in mouse (or vice versa) are significantly less constrained in mammals (median $K_a/K_s = 0.086$) than in fish (median $K_a/K_s = 0.061$) (Wilcoxon rank-sum test $P = 0.003$).

Difference in K_a/K_s Ratio Is Not Due to Bias with Respect to Known Predictors of Evolutionary Rate

We looked for alternative explanations of our results using three important predictors of rate of sequence evolution of a gene, namely expression level, breadth of expression, and genetic redundancy. Highly expressed genes are more conserved than genes expressed at low levels (Krylov et al. 2003), and broadly expressed genes are more conserved than genes expressed only in a subset of tissues (Duret and Mouchiroud 2000; Huminiecki and Wolfe 2004; Zhang and Li 2004). It is not known whether there are differences in expression level or breadth between genes with conserved alternative splicing and genes with genome-specific alternative splicing (Resch et al. 2004a), and there is no a priori reason to suspect any. However, if these variables cannot be eliminated as possible explanations for our result we do not need to invoke any other, less trivial, explanations. The difference we see in evolutionary rate relates to constitutive exons. So we set out to determine the level and breadth of expression of constitutive exons in each gene by pooling EST information from all its alternative transcript isoforms.

Using the number of assigned ESTs mapped to the genome for each gene as a simple measure of its expression we found no difference in expression levels of genes in the two categories of alternative splicing conservation. In humans the median number of ESTs for genes with conserved alternative splicing and genome-specific alternative splicing were 72 and 69, respectively. The corresponding numbers for mouse are 37 and 39, respectively. Similarly, there is no difference in breadth of expression for genes in the two alternative splicing conservation categories. The median number of human tissues showing evidence of expression was nine both for genes with conserved alternative splicing and for genes with genome-specific alternative splicing.

Selective constraint can also be affected by the presence of a close paralog. Genes that have undergone recent duplication experience relaxation of purifying selection corresponding to a period of functional redundancy (Kondrashov et al. 2002), and this is detected as an increase in K_a/K_s between the paralogs (Lynch and Conery 2000;

Jordan, Wolf, and Koonin 2004). We tested whether our two categories of genes (conserved alternative splicing and genome-specific alternative splicing) differed with respect to possession of a close paralog. The median value of K_s to the nearest paralog did not differ between categories (data not shown); therefore, the difference in orthologous K_a/K_s between categories cannot be explained as resulting from different propensities to undergo gene duplication.

Genome-Specific Alternatively Spliced Exons Are Likely to Be Exon Gains

If the association between having a genome-specific alternatively spliced exon and faster protein evolution reflects causation, our observations suggest one of two possibilities. First, genome-specific alternatively spliced exons could be recent gains in one lineage that have had a knock-on effect of speeding up protein sequence evolution. Alternatively, genome-specific alternatively spliced exons could be due to recent exon losses in the sister lineage, which would imply that loss of alternatively spliced sequence accelerates the substitution rate.

We attempted to distinguish between these two possibilities by using chicken (Hillier et al. 2004) as an out-group species to determine the direction of change. We used human alternatively spliced exons absent from mouse to search both the chicken genome and transcriptome and compared their detection rate to that of alternatively spliced exons conserved in humans and mouse. We did not use mouse-specific alternatively spliced exons for this analysis because the number of cases and the EST coverage of mouse are lower.

Direct sequence matches between alternatively spliced exons and chicken chromosomes recovered putative chicken homologs for conserved human-mouse alternatively spliced exons much more frequently than for alternatively spliced exons that are present in humans but not in mouse. The detection rate for the latter category was close to zero (table 3). These results point toward exon gain as the source of exons that are alternatively spliced in humans but are absent from the mouse genome, lending support to the first possibility above.

The validity of this assertion depends on the assumption that Blast has the same power to detect chicken homologs of exons in the two classes (conserved alternative splicing and genome-specific alternative splicing) between humans and mouse. This may not be the case if exons in the latter category are faster evolving, in which case failure to detect a Blast hit in the chicken genome for a given exon cannot be taken as evidence of its absence from chicken. However, we think it is unlikely that a difference in evolutionary rates alone could produce the sort of qualitatively different results for the two classes seen in table 3.

One way to partly account for possible rate differences among exons is to use a low-stringency search of the chicken transcriptome for putatively homologous chicken exons without requiring a direct sequence match to the alternatively spliced exon. We did this by searching chicken ESTs with a human query consisting of the alternatively spliced exon plus additional sequence from its flanking exons. Chicken ESTs aligning to the sequence of both flanking

Table 3
Results of Searching for Homologs of Human Alternatively Spliced (AS) Exons in the Chicken Genome

	Detected	Not Detected	Not Scored[a]
TBlastN, exon + 90 bp flanks to chicken genome			
Conserved AS	15 (75%)	5 (25%)	41
Genome-specific AS	1 (1%)	71 (99%)	155
BlastN, exon to chicken ESTs			
Conserved AS	21 (34%)	40 (66%)	not applicable
Genome-specific AS	2 (1%)	225 (99%)	not applicable
BlastN, exon + 90 bp flanks to chicken ESTs			
Conserved AS	12 (80%)	3 (20%)	46
Genome-specific AS	1 (4%)	25 (96%)	201
BlastN, exon + 90 bp flanks to genes adequately sampled with chicken ESTs			
Conserved AS	10 (83%)	2 (16%)	49
Genome-specific AS	1 (50%)	1 (50%)	225

Note.—''Conserved'' alternatively spliced exons are conserved with respect to humans and mouse. ''Genome-specific'' alternatively spliced exons are found in a human gene but are absent from genomic sequence of the mouse ortholog. Chicken genes with sufficient EST coverage to be confident of recovering homologs for a given human alternatively spliced exon were identified on the basis of a binomial test (Kan, States, and Gish 2002). See *Methods* for details.
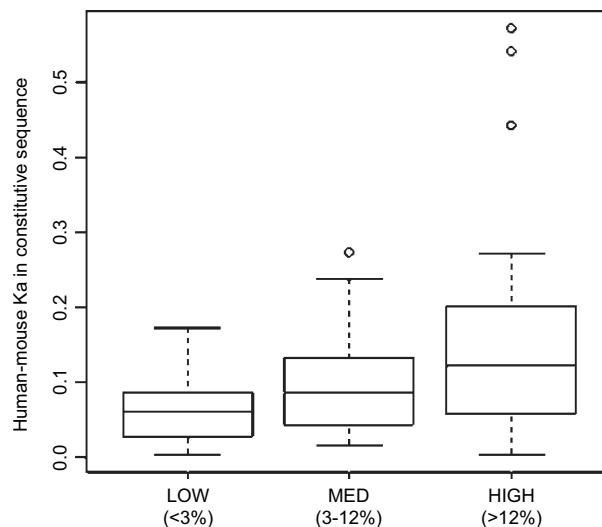
[a] Exons without anchoring matches to genomic or EST sequence were not scored. In addition, in the bottom panel, exons for which chicken EST coverage was inadequate were not scored.

Fig. 2.—Association between frequency of incorporation of human genome–specific alternatively spliced exons that are translated into productive alternatively spliced variants and K_a of constitutive gene sequence between humans and mouse. The lower and upper bounds of each box depict the first and third quartiles, respectively, while the horizontal line within each box corresponds to the median. The lower and upper whiskers extend to the most extreme data point within 1.5 times the interquartile range of the first and third quartiles, respectively.

exons and which contain a stretch of intervening EST sequence were scored as containing a chicken homolog of the human alternatively spliced exon even in the absence of any detectable sequence similarity to the exon itself. However, this approach is itself based on the assumption that all the alternatively spliced exons in question are spliced at equivalent frequencies, but the two sets of alternatively spliced exons under study here show a significant difference in their frequency of incorporation. Nonconserved alternatively spliced exons are spliced at low frequencies into minor-form transcripts, whereas alternatively spliced exons conserved between humans and mouse are generally represented among major-form transcripts (Modrek and Lee 2003). This means that a given number of chicken ESTs may be sufficient to detect a homolog of a human alternatively spliced exon if it is found in the major-form transcript but not if it is exclusive to the minor form.

We attempted to allow for splicing frequency differences by considering only high-confidence cases, i.e., alternatively spliced exons whose splicing frequency in humans and EST coverage in chicken is such that we expect to detect homologs in chicken if they do exist (Kan, States, and Gish 2002). Because only a small number of such high-confidence cases exists among alternatively spliced exons found in humans but not in mouse, we had insufficient evidence from this low-stringency strategy to determine the ancestry of many exons. Thus, we conclude that it is likely, but not certain, that genome-specific alternative exons are gains.

Influence of Frequency of Incorporation of Alternatively Spliced Exons

If the gain of an alternatively spliced exon is responsible for increasing the rate of amino acid change in con-

stitutive regions of the gene, then we might expect the strength of this effect to be proportional to the frequency at which the alternatively spliced exon is spliced into mRNA. Considering only genome-specific alternatively spliced human exons that are translated and productive, we classified each alternatively spliced exon by its frequency of incorporation and binned the alternatively spliced exons into three frequency categories on this basis. A strong correlation was detected between the binned splicing frequency and K_a for constitutive exons (Spearman's rank correlation $\rho = 0.353$, $P < 0.001$, $n = 107$). The median values of K_a for genes with genome-specific alternatively spliced exons incorporated at low ($n = 36$), medium ($n = 37$), and high frequency ($n = 35$) were 0.053, 0.086, and 0.122, respectively (fig. 2). A different method of classifying exons by splicing frequency is based on the counts of ESTs that either include or exclude the exon and uses inclusion thresholds of 33% and 66% (Resch et al. 2004a) to produce low-, medium-, and high-frequency bins. Using this approach gave us a very similar result (not shown). However, the classification of alternative splicing frequency using either approach introduces a bias because low-frequency alternative splicing events are more easily detectable in highly expressed genes, and gene expression level is a known correlate of evolutionary rate (Krylov et al. 2003).

To establish whether the slower evolution of genes with lower alternative exon inclusion frequency is explained by their higher expression level, we calculated the partial correlation between splicing frequency and K_a controlling for EST coverage (Spearman's partial correlation $\rho = 0.288$, $P < 0.01$, $n = 107$). This confirms that there

is a positive correlation between frequency of human genome–specific alternative exon inclusion and evolutionary rate in the constitutive parts of the gene, independent of EST coverage level. In contrast no such relationship was found between alternative splicing frequency and K_a among a control set of alternatively spliced exons conserved between humans and mouse (not shown).

Species-Specific Alternative Splicing in Genes with Conserved Exon-Intron Structure

The results described above indicate a higher rate of protein evolution among genes having genome-specific alternatively spliced exons compared to genes where there is conservation of both the alternatively spliced exon itself and of each alternative splicing event (exon inclusion and exclusion) as detected in human and mouse ESTs. The advantage of contrasting these two groups is that genes in the former group can be unambiguously classified as undergoing species-specific alternative splicing because there is no genomic evidence of the alternatively spliced exon in the orthologs of these genes. However, these two classes of genes differ not only in the degree of conservation of their alternative splicing patterns but also in the degree of conservation of gene structure. Thus, alternative splicing alone may not underlie the described disparity in rates of protein evolution because this may simply reflect faster sequence evolution of genes for which evolution of gene structure is also fast.

In order to account for any effect of changes in gene structure during evolution, we considered genes whose gene structure is conserved between humans and mouse but whose alternative splicing pattern appears to have changed. These genes show evidence of both inclusion and exclusion of an alternatively spliced exon in the first species (e.g., humans) but no evidence for alternative splicing of this exon in the second species (e.g., mouse). Classification of these genes is problematic because for some genes conserved alternative splicing may not be detected in one species due to undersampling of ESTs in the second species. Alternatively, some genes in this group may be undergoing truly species-specific alternative splicing. Because this group is likely to be a mixture of both types of genes we consider these genes as having genomically conserved exons whose alternative splicing conservation is "unclassified."

We retrieved two sets of such unclassified genes from the ASAP database. Both sets consist of genes showing evidence of alternative splicing in humans. In the first set ("mouse-skip" set) the mouse ortholog shows no EST evidence of inclusion of the exon, but there is sufficient sequence conservation in the mouse genome to suggest that a cryptic, possibly functional, exon exists. In the second set ("mouse-inclusion" set) there is no EST evidence for skipping of the relevant exon in the mouse ortholog.

We consider genes in these two sets to show putatively human-specific alternative splicing. It is important to note that we cannot distinguish truly human-specific alternative splicing in these genes from alternative splicing that is simply more frequent in humans than in mouse and has not yet been detected in mouse. In determining the effect on

Table 4
Mouse EST Coverage (Expressed As the Median Percentage of Human EST Coverage for Each Gene) for Genes Having Alternatively Spliced (AS) Exons in Humans But for Which the Homologous Mouse Exon Is Either Consistently Skipped (Mouse Skip) or Consistently Included (Mouse Inclusion)

	Conserved Exon Skip	Mouse Skips	Mouse Inclusion
All ESTs	53%	47% ($P > 0.1$)	48% ($P > 0.1$)
ESTs from named tissues	41%	58% (NS)	53% (NS)
ESTs from tissues in which human AS exon is:			
Included	29%	36% (NS)	—
Skipped	36%	—	50% (NS)

Note.—The median values shown are compared to those of the control group with conserved exon skipping, where mouse EST coverage is adequate by definition. Coverage in the mouse-skip and mouse-inclusion groups is not significantly lower than in the control group for any EST set ($P > 0.1$, one-tailed Wilcoxon rank-sum test; NS [not significant]: $P > 0.8$).

evolutionary rate, these sets of genes are only informative if they can be considered to be enriched for genes having human-specific alternative splicing. We therefore asked whether EST sampling of the mouse genes in these sets is sufficiently deep that we would expect to observe both exon inclusion and exon skipping in the mouse ESTs if alternative splicing were conserved. If a given mouse gene has been adequately sampled with ESTs and we still fail to observe a mouse counterpart for an alternative splicing event seen in humans, then we can be more confident that the apparent human-specific alternative splicing in this gene is real. Using a range of approaches we found that mouse EST coverage in the mouse-skip set and in the mouse-inclusion set is comparable with or even better than the control set (table 4). It is notable that, even in control genes where mouse EST coverage is adequate and detects conservation of alternative splicing, mouse EST coverage is only half that of humans.

On this basis both the mouse-skip and mouse-inclusion categories can be considered to be enriched for human-specific alternative splicing of conserved exons. This assertion is supported by a recent study which exploited the fact that most putatively human-specific alternatively spliced exons (corresponding to our mouse-inclusion group) have sequence features that can be used to discriminate them from conserved alternatively spliced exons. This led to an estimate that for 89% of such exons alternative splicing is likely to be human specific (Yeo et al. 2005).

It is therefore meaningful to compare K_a for constitutive sequence between these genes and the control group of genes with conserved alternative splicing. We saw a significant increase in K_a for genes in the mouse-skip category ($n = 163$, median $K_a = 0.068$, $P < 0.05$) and in the mouse-inclusion category ($n = 364$, median $K_a = 0.062$, $P < 0.01$) compared to the control set ($n = 66$, median $K_a = 0.046$): an increase of 48% and 35%, respectively. This suggests that genes with species-specific alternative splicing but conserved gene structure also show accelerated protein evolution in constitutive regions.

Discussion

Our results indicate that gaining an alternatively spliced exon is associated with an increased rate of evolution in the constitutive exons of a gene. We have not attempted to determine the origin of these gained exons. We note that other studies have reported that tandem exon duplication is one source of alternatively spliced exons (Kondrashov and Koonin 2001; Letunic, Copley, and Bork 2002), but none of the probable recent exon gains that we identified showed evidence of this. If genome-specific exons are created by tandem duplication, then the lack of detectable sequence homology in the orthologous gene must be due to rapid sequence change following duplication. By restricting our comparison to genes undergoing alternative splicing by exon skipping and subdividing these into those cases where alternative splicing occurred in the ancestor of humans and mouse and those where alternative splicing emerged in the human or mouse branch only, we have been able to focus on the impact of alternative splicing on recent mammalian sequence evolution. This approach was designed to eliminate the influence of functional differences between genes, unlike the comparison of sequence constraint in alternatively spliced genes to genes in the genome as a whole. Thus, although genes showing alternative splicing by exon skipping are a slow-evolving subset of the human genome, there is an increased rate of sequence evolution in the immediate aftermath of the appearance of alternative splicing. This result is reminiscent of observations about the evolution of duplicated genes. A number of studies have reported relaxation of sequence constraint in duplicated genes compared to singletons (Lynch and Conery 2000; Van de Peer et al. 2001; Nembaware et al. 2002; Seoighe, Johnston, and Shields 2003), but it has recently been shown that genes that tend to remain duplicated are generally more slowly evolving than genes that are found in single copy (Davis and Petrov 2004; Jordan, Wolf, and Koonin 2004). It therefore appears that conserved genes are more likely than faster evolving genes to undergo diversification by either gene duplication or alternative splicing and that both processes result in an increased rate of sequence change.

Several sources of error are linked to observations of alternative splicing at the genomic level. The primary question is the following: how reliable is any given observation of alternative splicing? Many EST sequences are derived from cancerous tissue sources, and these may exhibit a high rate of aberrant splice events that are not relevant to normal function (Sorek, Shamir, and Ast 2004). This is likely to have a disproportionate effect on observations seen in only one species because alternative splicing events conserved across species are more likely to be functional. This may reduce confidence in our observation of a difference in evolutionary rate between the two categories of alternatively spliced genes. However, two sources of evidence reinforce our result. First, if a given alternative splicing event occurs at high frequency we can be more confident that the event is functional (Kan, States, and Gish 2002). Our results show that genes with genome-specific alternative splicing occurring at high frequency ($>12\%$) show the greatest elevation of evolutionary rate. Second, restricting our analysis to include only those alternative splicing events that do not initiate NMD and which encode a distinct translation product shows that the observed rate difference is robust.

The limitations of the analogy between the evolution of gene duplicates and genes undergoing alternative splicing become apparent when we consider that alternatively spliced isoforms are not as free to evolve as paralogs. Nevertheless, we note that a recent study implicitly suggests an evolutionary equivalence between gene duplicates and alternative isoforms (Kopelman, Lancet, and Y. I. 2005). In the case of paralogs the increase in evolutionary rate observed following gene duplication is often explained as resulting from functional redundancy between duplicates because the fates of the two paralogs are uncoupled, thus, leading to relaxed selection on one of them (Van de Peer et al. 2001; Nembaware et al. 2002; Seoighe, Johnston, and Shields 2003). In contrast, accelerated sequence evolution in the constitutive parts of alternatively spliced genes cannot be attributed to simple sequence redundancy. When a gene becomes alternatively spliced, the evolutionary fates of the two transcripts are tightly coupled because some exons remain common to both transcripts. In this case only the alternatively spliced sequence itself would be expected to provide raw material for evolutionary change. This is implied by the original model of Modrek and Lee, where alternative splicing generates an internal paralog that is shielded from the constraints imposed by purifying selection and has been supported by more recent results (Modrek and Lee 2003; Xing and Lee 2005). However, our results show that the constitutive exons shared between transcripts are themselves subject to alteration of sequence constraint following the acquisition of alternative splicing.

The slower evolution we observe in genes undergoing alternative splicing by exon skipping compared to the average genome-wide rate of evolution is consistent with the classical model of evolutionary constraint accompanying pleiotropy (Fisher 1930). Thus, the fact that an alternatively spliced gene may have multiple roles associated with its multiple isoforms (Xing, Xu, and Lee 2003; Resch et al. 2004*b*) means that an individual mutation is more likely to be deleterious. On the other hand, we can imagine the constitutive exons in an alternatively spliced gene as being subjected to two distinct selective regimes corresponding to the different functions of its isoforms. This can be likened to a state of adaptive conflict (Piatigorsky and Wistow 1991), where changes beneficial to one function may be deleterious to the other. Selective constraint will be imposed by the need to maintain ancestral gene function (encoded by the major-form transcript), which will tend to slow sequence change in the constitutive region of the gene. However, the potential functional innovation associated with an internal paralog (encoded by the minor-form transcript) may demand correlated sequence changes in constitutive regions, thus increasing the rate of sequence evolution in the gene as a whole. These amino acid changes may be fixed if they have an adaptive benefit in the context of the function of the minor isoform while being selectively neutral, or even slightly deleterious, to the function of the major isoform. Piatigorsky and Wistow (1991) proposed that gene duplication can resolve the stalemate between these opposing selective forces. Our results demonstrate that the constitutive

exons of alternatively spliced genes possess sufficient plasticity to accommodate the competing functional demands of their isoforms. This is underlined by our observation of a correlation between frequency of alternative exon incorporation and evolutionary rate in constitutive regions. These observations mirror results from a recent directed evolution study which demonstrated that negative trade-offs between different enzyme functions are much weaker than expected (Aharoni et al. 2004).

We should, however, be cautious before interpreting the strong correlation between the apparent gain of genome-specific alternative splicing and the increased rate of protein evolution as reflecting an actual causation. Both variables may be under the influence of some untested variable whereby, following the human-mouse split, a change in selective pressure operating on a gene may manifest itself both as a change in gene structure and in an increased rate of nonsynonymous evolution.

Acknowledgments

Literature Cited

Aharoni, A., L. Gaidukov, O. Khersonsky, S. McQ Gould, C. Roodveldt, and D. S. Tawfik. 2004. The 'evolvability' of promiscuous protein functions. Nat. Genet. **37**:73–76.

Altschul, S. F., T. L. Madden, A. A. Schaeffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.

Beissbarth, T., and T. P. Speed. 2004. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. Bioinformatics **20**:1464–1465.

Boue, S., I. Letunic, and P. Bork. 2003. Alternative splicing and evolution. Bioessays **25**:1031–1034.

Brett, D., H. Pospisil, J. Valcarcel, J. Reich, and P. Bork. 2002. Alternative splicing and genome complexity. Nat. Genet. **30**:29–30.

Davis, J. C., and D. A. Petrov. 2004. Preferential duplication of conserved proteins in eukaryotic genomes. PLoS Biol. **2**:E55.

Duret, L., and D. Mouchiroud. 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. Mol. Biol. Evol. **17**:68–74.

Fisher, R. 1930. The genetical theory of natural selection. Dover, New York.

Harrington, E. D., S. Boue, J. Valcarcel, J. G. Reich, and P. Bork. 2004. Estimating rates of alternative splicing in mammals and invertebrates. Nat. Genet. **36**:916–917.

Hillier, L. W., W. Miller, E. Birney et al. (175 co-authors). 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature **432**:695–716.

Huminiecki, L., and K. H. Wolfe. 2004. Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse. Genome Res. **14**:1870–1879.

Iida, K., and H. Akashi. 2000. A test of translational selection at 'silent' sites in the human genome: base composition comparisons in alternatively spliced genes. Gene **261**:93–105.

Johnson, J. M., J. Castle, P. Garrett-Engele, Z. Kan, P. M. Loerch, C. D. Armour, R. Santos, E. E. Schadt, R. Stoughton, and D. D.

Shoemaker. 2003. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science **302**:2141–2144.

Jordan, I. K., Y. I. Wolf, and E. V. Koonin. 2004. Duplicated genes evolve slower than singletons despite the initial rate increase. BMC Evol. Biol. **4**:22.

Kan, Z., D. States, and W. Gish. 2002. Selecting for functional alternative splices in ESTs. Genome Res. **12**:1837–1845.

Kondrashov, F., and E. V. Koonin. 2001. Origin of alternative splicing by tandem exon duplication. Hum. Mol. Genet. **10**:2661–2669.

Kondrashov, F. A., I. B. Rogozin, Y. I. Wolf, and E. V. Koonin. 2002. Selection in the evolution of gene duplications. Genome Biol. **3**:RESEARCH0008.

Kopelman, N., D. Lancet, and I. Yanai. 2005. Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. Nat. Genet. **37**:588–589.

Kriventseva, E. V., I. Koch, R. Apweiler, M. Vingron, P. Bork, M. S. Gelfand, and S. Sunyaev. 2003. Increase of functional diversity by alternative splicing. Trends Genet. **19**:124–128.

Krylov, D. M., Y. I. Wolf, I. B. Rogozin, and E. V. Koonin. 2003. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. Genome Res. **13**:2229–2235.

Lee, C., L. Atanelov, B. Modrek, and Y. Xing. 2003. ASAP: the Alternative Splicing Annotation Project. Nucleic Acids Res. **31**:101–105.

Letunic, I., R. R. Copley, and P. Bork. 2002. Common exon duplication in animals and its role in alternative splicing. Hum. Mol. Genet. **11**:1561–1567.

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

Modrek, B., and C. Lee. 2002. A genomic view of alternative splicing. Nat. Genet. **30**:13–19.

———. 2003. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. Nat. Genet. **34**:177–180.

Nagy, E., and L. E. Maquat. 1998. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem. Sci. **23**:198–199.

Nembaware, V., K. Crum, J. Kelso, and C. Seoighe. 2002. Impact of the presence of paralogs on sequence divergence in a set of mouse-human orthologs. Genome Res. **12**:1370–1376.

Orban, T. I., and E. Olah. 2001. Purifying selection on silent sites—a constraint from splicing regulation? Trends Genet. **17**:252–253.

Piatigorsky, J., and G. Wistow. 1991. The recruitment of crystallins: new functions precede gene duplication. Science **252**:1078–1079.

Resch, A., Y. Xing, A. Alekseyenko, B. Modrek, and C. Lee. 2004a. Evidence for a subpopulation of conserved alternative splicing events under selection pressure for protein reading frame preservation. Nucleic Acids Res. **32**:1261–1269.

Resch, A., Y. Xing, B. Modrek, M. Gorlick, R. Riley, and C. Lee. 2004b. Assessing the impact of alternative splicing on domain interactions in the human proteome. J. Proteome Res. **3**:76–83.

Seoighe, C., C. R. Johnston, and D. C. Shields. 2003. Significantly different patterns of amino acid replacement after gene duplication as compared to after speciation. Mol. Biol. Evol. **20**:484–490.

Skrabanek, L., and F. Campagne. 2001. TissueInfo: high-throughput identification of tissue expression profiles and specificity. Nucleic Acids Res. **29**:102.

Sorek, R., G. Ast, and D. Graur. 2002. Alu-containing exons are alternatively spliced. Genome Res. **12**:1060–1067.

Sorek, R., R. Shamir, and G. Ast. 2004. How prevalent is functional alternative splicing in the human genome? Trends Genet. **20**:68–71.

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

Van de Peer, Y., J. S. Taylor, I. Braasch, and A. Meyer. 2001. The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes. J. Mol. Evol. **53**:436–446.

Wheeler, D. L., D. M. Church, R. Edgar et al. (13 co-authors). 2004. Database resources of the National Center for Biotechnology Information: update. Nucleic Acids Res. **32**:D35–D40.

Xing, Y., and C. Lee. 2005. Evidence of functional selection pressure for alternative splicing events that accelerate evolution of protein subsequences. Proc. Natl. Acad. Sci USA (in press).

———. 2004. Negative selection pressure against premature protein truncation is reduced by alternative splicing and diploidy. Trends Genet. **20**:472–475.

Xing, Y., A. Resch, and C. Lee. 2004. The multiassembly problem: reconstructing multiple transcript isoforms from EST fragment mixtures. Genome Res. **14**:426–441.

Xing, Y., Q. Xu, and C. Lee. 2003. Widespread production of novel soluble protein isoforms by alternative splicing removal of transmembrane anchoring domains. FEBS Lett. **555**: 572–578.

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

Yeo, G. W., E. Van Nostrand, D. Holste, T. Poggio, and C. B. Burge. 2005. Identification and analysis of alternative splicing events conserved in human and mouse. Proc. Natl. Acad. Sci. USA **102**:2850–2855.

Zhang, L., and W. H. Li. 2004. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. Mol. Biol. Evol. **21**:236–239.