Alternative splicing and RNA selection pressure — evolutionary consequences for eukaryotic genomes

Yi Xing[*‡§] *and Christopher Lee*[*]

Abstract | Genome-wide analyses of alternative splicing have established its nearly ubiquitous role in gene regulation in many organisms. Genome sequencing and comparative genomics have made it possible to look in detail at the evolutionary history of specific alternative exons or splice sites, resulting in a flurry of publications in recent years. Here, we consider how alternative splicing has contributed to the evolution of modern genomes, and discuss constraints on evolution associated with alternative splicing that might have important medical implications.

*Molecular Biology Institute, Center for Genomics and Proteomics, Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, USA.
‡Department of Internal Medicine, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA.
§Department of Statistics, Stanford University, Sequoia Hall, 390 Serra Mall, Stanford, California 94305, USA.
Correspondence to C.L.
e-mail: leec@mbi.ucla.edu
doi:10.1038/nrg1896
Published online 13 June 2006

Soon after the discovery of exons and introns in the adenovirus *hexon* gene[1,2], Walter Gilbert proposed that different combinations of exons could produce multiple mRNA isoforms of a single gene[3]. He further suggested that such a mechanism of alternative splicing can have important implications in the evolution of gene function: "Evolution can seek new solutions without destroying the old … the genetic material does not have to duplicate to provide a second copy of an essential gene in order to mutate to a new function."[3] In the past decade, genomics has started to reveal the important role of pre-mRNA alternative splicing in shaping the evolution of mammalian genomes, something that Gilbert envisioned more than 25 years ago (see also REF. 4). High-throughput data such as ESTs and microarrays have provided a genome-wide picture of alternative splicing in multiple organisms[5–7], while whole-genome sequences and comparative genomics have made it possible to look back in detail at the evolutionary history of a specific alternative splice site or exon[8–16].

Here, we review several new directions of research that have recently emerged from studies of alternative splicing. We consider how alternative splicing has contributed to the evolution of modern genomes, and discuss constraints on evolution associated with alternative splicing that might have important medical implications.

Conservation of alternative splicing

Alternative splicing includes events such as exon skipping, alternative 5′ or 3′ splicing, mutually exclusive exons, intron retention and so on[17] (FIG. 1). In mammals, exon skipping is the most abundant type of alternative splicing[15].

In defining the conservation of alternative splicing, it is important to realize that alternative splicing represents variations in pre-mRNA splicing at the transcriptomic level. A conserved exon can have different splicing patterns in the transcriptomes of two species[11]. We define two levels of conservation for exon skipping: genomic conservation, which defines whether an alternatively spliced exon is conserved between the genomes of two species, and transcriptomic conservation, which defines whether an alternative splicing event is independently observed in the transcriptomes of two species (FIG. 2). These definitions can be extended to other types of alternative splicing event[18].

Alternative splicing and organismal complexity

In parallel with the surprising discovery of how prevalent alternative splicing is in humans (the same is true for other mammals[5–7]), the initial analysis of the human genome also revealed a strikingly small number of genes[19,20]. It has been proposed that the unexpectedly high frequency of alternative splicing might provide an attractive explanation for increasing organismal complexity in higher eukaryotes[5,21]. There is some evidence to support this hypothesis. For example, most (>95%) genes in *Saccharomyces cerevisiae* are intronless[22], and only a few alternative splicing events have been reported[23]. In addition, a genome-wide survey identified a very small number of splicing regulatory proteins in *S. cerevisiae*, compared with *Drosophila melanogaster*, worms and humans[24]. However, a trend towards more complex alternative splicing in higher eukaryotes is less clear, and even controversial[25,26]. Brett *et al.* investigated

© 2006 Nature Publishing Group

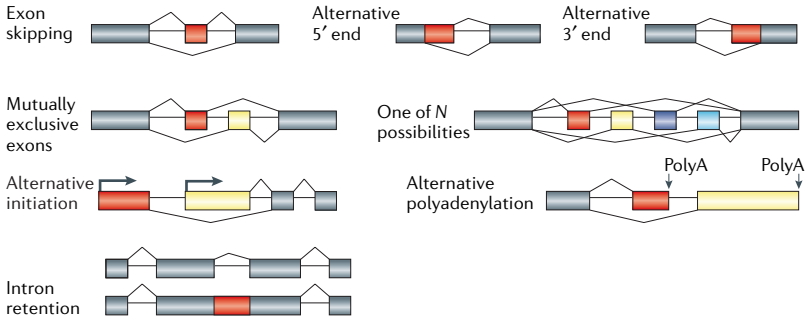


Figure 1 | **Patterns of alternative splicing.** Red boxes represent alternatively spliced exons, coloured boxes represent other possible alternative exons and grey boxes represent constitutive exons.

the frequency of alternative splicing in seven species[25]. Based on their analysis of EST data, *Arabidopsis thaliana* had a significantly lower frequency of alternative splicing than the other organisms in the study. At the same time, the frequency of alternative splicing was comparable among different animals (vertebrates and invertebrates). It is worth noting that EST data contain various sources of artefacts and biases. Microarray profiling of alternative splicing in multiple genomes is likely to provide a clearer picture of the relationship between alternative splicing and genome complexity.

Accelerated evolutionary change

Evolution can create new function in two ways: by creating new genes through processes such as duplication[27,28] or by creating new exons within existing genes. Several recent studies have suggested that alternative splicing contributes substantially to the latter, and in general can 'relax' negative selection pressure against several types of evolutionary change within a functioning gene[6,13,29–31].

As early as 1994 it was noted that transposable elements — which are now a well-documented source of new exons in the human genome[32] — such as *Alu* are sometimes found in protein-coding regions of human genes[33]. In some cases, alternative splicing of such *Alu* elements regulates gene function (in one example, by converting the membrane-bound DAF protein into a soluble isoform[33]). Recent genome-wide studies found transposable elements (mainly *Alu*s) in nearly 4% of human protein-coding regions[34]. Since many transposon-insertion events are recent (for example, *Alu* elements are restricted to the primate lineage), such widespread existence of *Alu*-containing exons suggests that transposable elements have contributed significantly to the creation of new exons during recent evolution. Further analyses confirmed that, in most cases, a transposable element was first inserted into an intron and subsequently recruited as a new exon (*Alu* elements contain several potential splice donor and acceptor sites in both sense and antisense orientations)[34].

This process was investigated in detail by Sorek and colleagues. Using cDNA and EST data, they found that all *Alu*-containing exons tested were alternatively spliced[35]. This is to be expected, as most *Alu*-containing exons introduce a frameshift or premature termination codon and have weak splice sites. Evaluating intronic *Alu* elements for the number of mutations required to convert them into an exon, Lev-Maor *et al.* estimated that nearly 75,000 intronic *Alu* elements are just one mutation away from exonization[36]. Singer *et al.* showed just how precisely such an event can be mapped in the evolutionary history of multiple genomes[37] (FIG. 3). By sequencing genomic DNA from 13 primates ranging from humans to lemurs, they studied the creation of an *Alu*-containing exon that acts as an alternative first exon in tumour-necrosis factor receptor type 2 (*p75TNFR*)[37]. Their data show that the *Alu* was first inserted 40–58 million years ago, quickly followed by an A-to-G transition that created the start codon. However, the functional exon was not created until 25 million years ago, when the splice donor site was created through a C-to-T substitution; a similar sequence of events has been observed in other cases of exonization[38].

To assess the contribution of transposable elements to the evolution of functional proteins, Gotea and Makalowski surveyed non-redundant entries in the RCSB Protein Data Bank for peptide fragments derived from transposable elements[39]. Surprisingly, although almost 4% of transcripts contain transposable elements[34], the observed proportion was much lower at the protein level (~0.1% for non-redundant Protein Data Bank entries, although this might be an underestimation[39]). All of the transposable-element fragments in the Protein Data Bank entries were derived from old transposable elements, indicating that incorporation of mobile elements into functional proteins is slow, consistent with

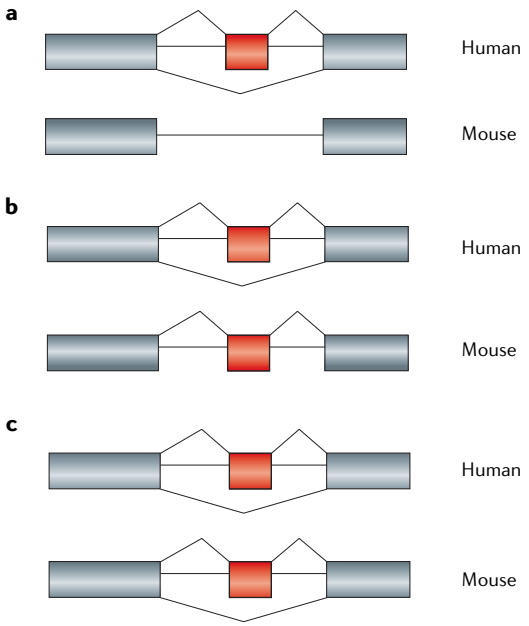


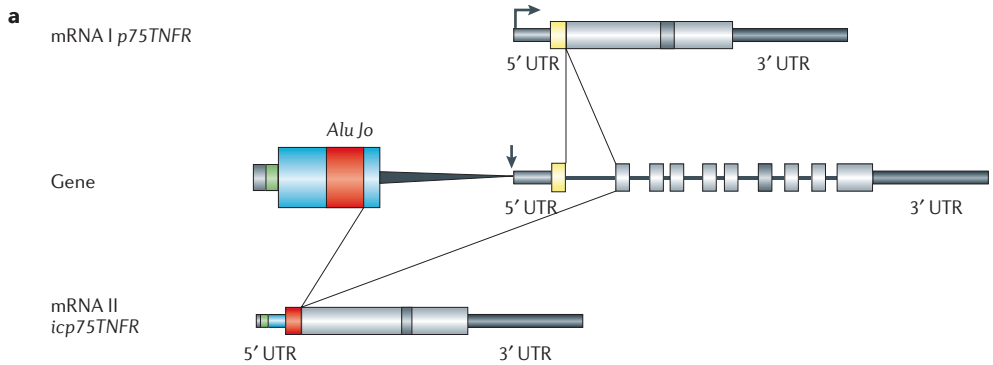
Figure 2 | **Defining the conservation of alternative splicing. a** | A non-conserved alternatively spliced exon. **b** | A conserved alternatively spliced exon, but the splicing pattern is not conserved **c** | An ancestral alternatively spliced exon. Red boxes represent alternatively spliced exons and grey boxes represent constitutive exons.

Alu
A class of retrotransposons that belongs to the primate-specific family of short interspersed elements.

© 2006 Nature Publishing Group

the observation made for *p75TNFR* (REF. 37). The predominant role of young transposable elements might be regulatory. Alternative splicing of transposable-element-derived exons might be coupled with nonsense-mediated decay (NMD) to control gene expression, a hypothesis proposed by Brenner and colleagues[40].

Another source of new exons is exon duplication. Recent work by Letunic *et al.*[41] identified a large number of human and animal genes that contain pairs of duplicated exons. In most cases, exons are mutually exclusive because inclusion of both exons of the pair would disrupt the reading frame[41]. Conversely, Kondrashov *et al.* found that exon duplication was the likely origin of about 10% of all mutually exclusive alternative splicing cases examined[42]. These studies suggest that tandem exon duplication followed by introduction of alternative splicing has had an important role in expanding the functional and regulatory diversity of the genes involved. A functional example of this regulatory mechanism was recently reported for ion-channel genes in humans, insects



b

```
Alu Jo      RGCCGGGCGC GGTGGCTCAC GCCTGTAATC CCAGCACTTT GGGAGGCC-G AGGCGGGAGG ATCGCTTGAG ---------- 69
icp75TNFR   GGCCGACTGC AGTGGCTCAC ACCTATAATC CCAGCACCTT GGGAGGCCAG AGGCGGGAAG ATCACTTGAG GGTGGGAAGA 80
Alu Jo      ---------C CCAGGAGTTC GAGACCAGCC TGGGCAAC  ATAGCGAGAC CCCGTCTCTA  CAAAAAATAC AAAAATTAGC 138
icp75TNFR   ACACGTGAGC TCAGGAGTTC GAGACCAGCC TGGACAAC  ATGGCGAAAC CCCATCTCTA  7bp delTAA AGAAATCAGC 151
Alu Jo      CGGGCGTGGT GGCGCGCGCC TGTAGTCCCA GCTACTCGGG AGGCTGAG  GCAGGAGGAT  CGCTTGAGCC CAGGAGTTCG 216
icp75TNFR   CTAGCATGGT GGCCCGAGCC TGTAGTCCCA GCTACTCGGG AGGCTGAG  GTGGGAGGAT  CGCTTGAGCG CAGGAGTTGG 229
Alu Jo      AGGCTGCAGT GAGCTATGAT CGCGCCACTG CACTCCAGCC TGGGCGACAG AGCGAGACCC TGTCTCAAAA AAAAAAAAAA 296
icp75TNFR   AGGCTGCAGT GAGCTATG-- ---------- ---------- --GGTGAAAG AGTGAGACCT TGTCTCAAAA AAAATTAAAA 285
Alu Jo      AAAAAA--                                                                              302
icp75TNFR   AATAAGAA                                                                              293
```

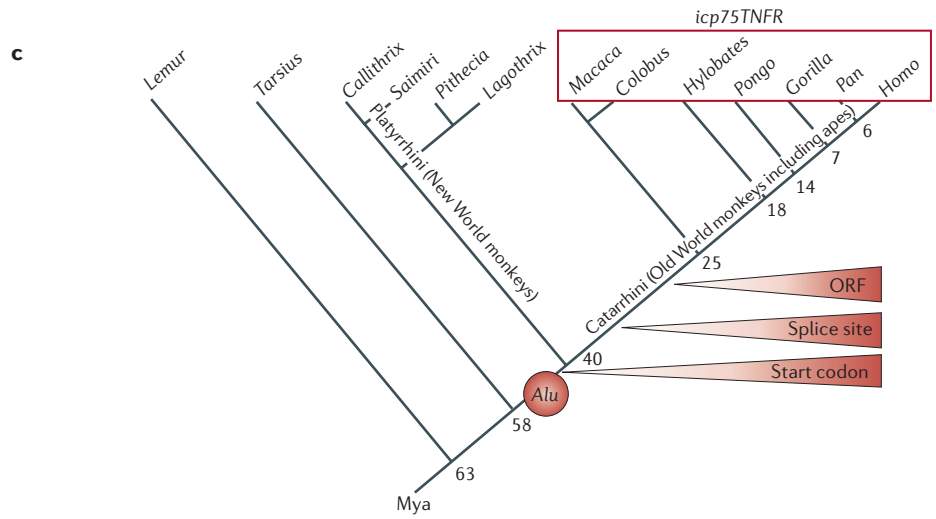


Figure 3 | **Creation of a new functional alternative exon of *p75TNFR* from an *Alu* element. a** | Gene structure and alternative splicing of *p75TNFR*. The exon 1a (red) is an alternative first exon that originates from an *Alu* element of the *Alu Jo* subfamily. **b** | Pairwise alignment of *p75TNFR* exon 1a locus to the *Alu Jo* subfamily consensus sequence. Assuming the *Alu Jo* consensus sequence is the origin for the exon 1a locus, the alignment indicates an A-to-G substitution which creates the start codon on exon 1a; a C-to-T substitution which creates a splice donor site; and a 7-bp deletion in exon 1a which creates a full-length ORF. The red box delineates the boundaries of exon 1a. **c** | Phylogenetic analyses of the *p75TNFR* exon 1a locus in primates. The *Alu* insertion occurred 40–58 million years ago (mya), followed shortly by the A-to-G substitution (which creates the start codon on exon 1a). The C-to-T substitution (which creates a splice donor site) and the 7-bp deletion (which creates a full-length ORF) occurred 25–40 million years ago. Modified with permission from REF. 37 © (2004) Elsevier Ltd.

Nonsense-mediated decay
An mRNA surveillance mechanism for removing aberrant mRNAs with premature termination codons.

© 2006 Nature Publishing Group

Major-form exons
Alternatively spliced exons with high exon-inclusion levels — they are usually included in the transcripts.

Minor-form exons
Alternatively spliced exons with low exon-inclusion levels — they are usually excluded from the transcripts.

Outgroup
In phylogenetic analysis, the taxon that is most distant from all the other taxa of interest. For example, human is an outgroup to mouse and rat.

Subfunctionalization
Two duplicated genes specialize to perform complementary functions.

and worms[43]. Indeed, although in various genes that encode ion channels the same pair of tandem exons are alternatively spliced, Copley suggested that this pattern might be so functionally important that it has evolved independently several times[43].

The association between alternative splicing and increased levels of evolutionary change seems not to be limited to specific cases such as *Alu* or tandem exons. Genome-wide analyses of alternatively spliced exons between mammalian orthologues reveal a similar pattern. For example, Modrek and Lee classified alternative exons as 'major' or 'minor', depending on the fraction of a gene's transcripts containing that exon, and showed that this exon-inclusion level (the fraction of a gene's transcripts that includes a specific alternatively spliced exon) is strongly conserved for orthologous exons between human and mouse[13]. Moreover, major-form exons are similar to constitutive exons in that they are nearly all conserved between human and mouse genomes. But minor-form exons seemed to have a much higher evolutionary turnover rate, with nearly three-quarters not conserved between human and mouse (FIG. 4). This conclusion is corroborated by a recent microarray study of alternatively spliced mouse exons[6], as well as a study of insect genomes[44]. Subsequent outgroup analyses using multiple-genome comparisons indicate that a large fraction of these non-conserved exons have been created recently[29,45]. This interpretation is consistent with recent comparisons of alternatively spliced protein products with their yeast homologues, which found 25 alternatively spliced peptide segments that were the result of insertion events (as opposed to deletion events, relative to the yeast homologue)[46]. There is also detailed evidence that alternative splicing is associated with exon creation and splice-site shifts in the *MAGE-A* family of cancer/testis antigens[47]. A remarkable example of the evolution of complex alternative splicing to rapidly produce diversified functions comes from cichlid fish, in which alternative splicing of the *hag* gene can influence pigmentation and mate choice, leading to cichlid speciation[48].

Together, these data suggest an evolutionary model in which alternative splicing can relax negative selection pressure against large-scale changes such as exon creation (FIG. 5). Ordinarily, introduction of a new exon into an existing gene is likely to cause a frameshift or to disrupt an important structural or functional element in the protein product, and therefore to incur negative selection. By contrast, if a new exon is introduced as a minor splice form, enough of the original gene product is made so that the new exon creation is likely to experience little negative selection (unless the minor form has a dominant-negative effect, in which case protective mechanisms such as NMD are likely to degrade it, again reducing negative selection pressure[49,50]). So, alternative splicing might be able to open neutral or nearly neutral evolutionary paths for large-scale evolutionary changes such as exon creation (see REF. 51 for a theoretical model). This 'neutralization' effect seems to be similar in magnitude to the effect of diploidy. The autosomes of mammals and flies have a much larger fraction of endogenous NMD targets compared with the X chromosome[52,53], indicating that diploidy is associated with a significant relaxation of negative selection against premature protein truncation. Similarly, minor isoforms are associated with a much higher frequency of premature termination codons compared with major isoforms[52,54,55]. Some of the premature transcripts might acquire new functions through subsequent mutations[37,49].

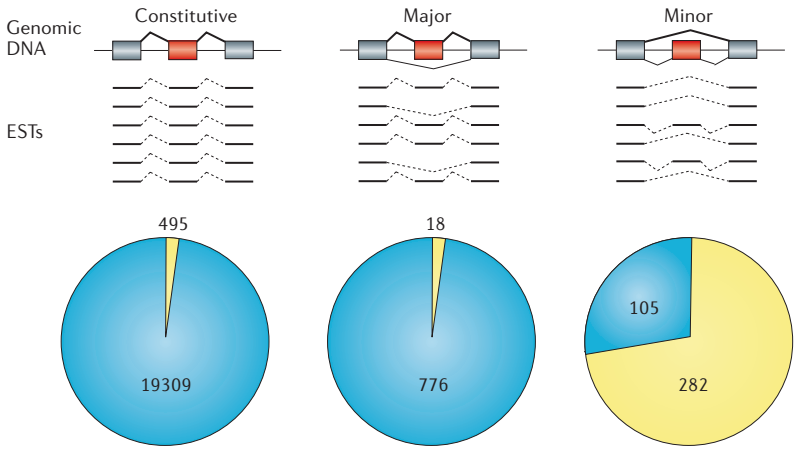


Figure 4 | **Conservation of human constitutive and alternatively spliced exons in mouse orthologues.** The majority (~98%) of human constitutive exons (left) and major-form alternative exons (middle) are conserved in the genomic sequences of their mouse orthologues. By contrast, nearly three-quarters of minor-form alternative exons are not conserved (right). Combined with subsequent outgroup analyses[29,45], these data suggest that alternative splicing (especially minor-form exons) is associated with an accelerated rate of exon creation in mammals. The blue portion represents the number of human exons that are conserved in the genomic sequences of mouse orthologues. The yellow portion represents the number of human exons with no detectable homology in the genomic sequences of mouse orthologues. Modified with permission from *Nature Genetics* REF. 13 © (2003) Macmillan Publishers Ltd.

Alternative splicing and gene duplication

How might alternative splicing and gene duplication interact during evolution? There is anecdotal evidence that gene duplication can lead to loss of alternative splicing[56,57]. For example, the common ancestor of the *Fugu* (*Takifugu*) synapsin-2 paralogues (*SYN2a* and *SYN2b*) seems to have had two splice forms, each of which is now expressed separately by the two duplicated genes[56], consistent with the subfunctionalization model of gene duplication[58]. A similar pattern has been reported for *MITF* gene duplication in teleost fish species[57].

A genome-wide comparison between the frequency of alternative splicing in multigene families and singleton genes without paralogues[59] revealed a significant inverse correlation between the frequency of alternative splicing and the size of the multigene family. Singleton genes show a significantly higher frequency of alternative splicing compared with genes in multigene families. A similar trend was also reported by Su and colleagues[60]. They estimated the approximate age of gene-duplication events using evolutionary distance of the gene duplicates and showed that alternative splicing seems to be lost rapidly after gene duplication. These findings are also consistent with the subfunctionalization of synapsin-2 and *MITF* paralogues.

© 2006 Nature Publishing Group

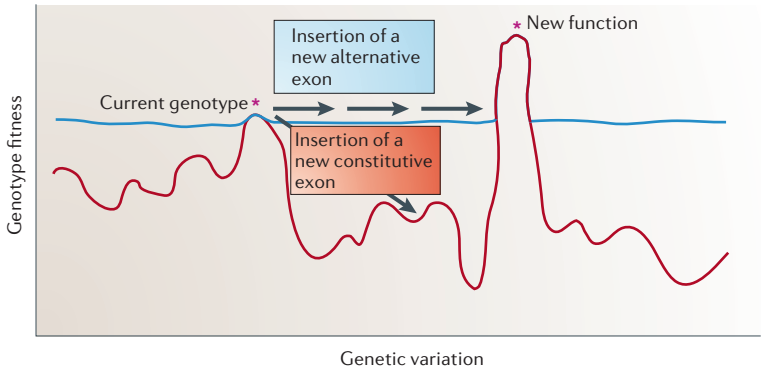


Figure 5 | **Alternative splicing opens neutral paths for an accelerated rate of new exon creation.** Because of frameshift, in-frame premature stop codons or disruption of functional and/or structural elements, the evolutionary landscape of a new constitutive exon in an existing gene is largely negative. The insertion of a new constitutive exon is likely to cause a significant reduction in fitness and is therefore likely to be prevented by strong negative selection. Insertion of a new exon as a minor-form exon through alternative splicing neutralizes regions of negative selection in the fitness landscape of the new exon. Therefore, alternative splicing opens neutral paths for an accelerated rate of new exon evolution. The new minor-form alternative exon can evolve rapidly, possibly to a new, adaptive function. The red and blues lines represent fitness landscapes for each of the two scenarios discussed.

These genome-wide studies indicate that alternative splicing and gene duplication are not evolutionarily independent. The presence of two copies of a gene as a result of duplication might reduce the need for an internal paralogue created by alternative splicing.

Amino-acid selection pressure

A related question is whether alternative splicing affects the rate of evolution in conserved protein-coding sequences. The K_a/K_s ratio (sometimes also referred to as the d_n/d_s ratio) is a widely used metric for selection pressure on amino-acid sequences[61,62]. It measures the rate of amino-acid substitutions (K_a) between a pair of homologous sequences, divided by the rate of synonymous substitutions (K_s); it is normalized so that a K_a/K_s value of 1 implies neutral evolution.

Estimates of the K_a/K_s ratio in alternatively spliced genes indicate that alternative splicing is associated with significant relaxations of protein-level selection pressure during evolution. A comparison of the rate of protein evolution in constitutively spliced genes and the constitutive regions of alternatively spliced genes revealed a correlation between a rapid change in gene structure (through alternative splicing) and a rapid divergence of amino-acid sequence (K_a/K_s ratio) in constitutive regions[29]. Alternatively spliced exons in several genes such as *CD45* and *BRCA1* have unusually high K_a/K_s values (≥1), implying nearly neutral evolution[63,64]. Genome-wide analyses of alternatively spliced exons also show a significantly higher K_a/K_s value in alternatively spliced exons, indicating that there are local 'hot spots' of elevated K_a/K_s ratio in specific segments of sequences that are alternatively spliced[65–68]. Some studies even indicate a modest increase in the absolute value of K_a, in particular for younger exons[65,66,68].

Ancestral alternatively spliced exons
Exons that are alternatively spliced in the transcripts of two species, suggesting that alternative splicing was present in the common ancestor of these species.

These findings have important implications. For example, the widely used K_a/K_s ratio test[69] to assess the protein-coding potential of genomic regions might need to be recalibrated to consider the possibility of alternative splicing[70,71]. The standard K_a/K_s model uses K_s as a proxy for the neutral nucleotide mutation rate, which affects both synonymous and non-synonymous sites[62]. In this model, the K_a/K_s ratio should be independent of K_s unless there is strong selection pressure that specifically targets synonymous sites[72] (such as codon-usage bias[73]). Ordinarily, changes in the K_a/K_s ratio (for example, between housekeeping and tissue-specific genes) are expected to be associated with a nearly constant K_s (REF. 74). However, recent data indicate that selection pressure can act at silent sites in mammalian genomes (see REF. 75 for a review).

A recent wave of evidence indicates that alternative splicing introduces strong selection pressure for RNA-sequence motifs that are involved in the regulation of alternative splicing. These data suggest many new directions for research and important implications to which we devote the remainder of this review.

RNA selection pressure

Growing interest in RNA selection pressure has been sparked by the convergence of two lines of research: the discovery of 'ultraconserved elements' in whole-genome comparisons[76] and the finding that alternatively spliced exons are more highly conserved at their silent sites and flanking intronic regions than constitutive exons[10,15,65,67,77–79]. Bejerano *et al.* identified 481 segments of absolute identity between the human, mouse and rat genomes (longer than 200 nt)[76]. Not only were these sequences almost completely conserved in chicken and dog, but many were also conserved in fish. Of the subset that mapped to known genes, about half were within introns, which ordinarily are considered to be poorly conserved, whereas the others partly overlapped known exons, although even in this case most extended far into flanking introns. The ultraconserved elements are strongly associated with genes involved in RNA binding and splicing regulation, and a very large fraction overlaps alternatively spliced regions of these genes. These data suggest that alternative splicing in these genes might be associated with strong purifying selection that both imposes nearly absolute conservation over large sequence segments (200–1,087 nt long) and is evolutionarily ancient (in many cases extending back to fish).

An independent line of research has shown that the intronic regions that flank alternatively spliced exons are much more conserved than those flanking constitutive exons. Sorek and Ast found that flanking intronic regions of 243 ancestral alternatively spliced exons were significantly more conserved than the flanking introns of constitutive exons[77]. This effect was strongest in the region immediately flanking the 5′ and 3′ splice sites (76–87% identity in the flanking 30 nt, compared with 55–63% for the 30 nt flanking constitutive exons), but extended out to at least 90 nt, compared with approximately 40 nt flanking constitutive exons. Similar results have been reported in other studies[10,15,65,67,79,80].

© 2006 Nature Publishing Group

A more detailed picture has emerged from studies of exon-inclusion levels and separate measurements of synonymous (K_s) and non-synonymous (K_a) substitution rates for alternatively spliced exons[65,67]. These data indicate that there are at least two types of conserved alternative splicing event with markedly different evolutionary patterns. Conserved major-form exons show only a slight increase in conservation: their K_s and K_a rates are similar to constitutive exons, as is their substitution density in the flanking introns[65]. By contrast, conservation is higher in conserved minor-form exons (including flanking introns up to 150 nt from the splice sites). Intriguingly, this effect is only observed at synonymous sites within an exon: in human–mouse comparisons, the K_s value for such minor-form exons was up to sixfold lower than that for neighbouring constitutive exons in the same genes, whereas the K_a value was similar for minor-form, major-form and constitutive exons[65–68]. This surprising result indicates that the increased conservation associated with alternative exons is not due to amino-acid selection pressure, but instead seems to reflect a purifying selection pressure on the RNA sequence itself. In addition, minor-form exons are shorter (87 nt on average, compared with approximately 128 nt for constitutive exons) and have markedly weaker splice sites[81]. The characteristics of ancestral alternatively spliced exons are summarized in TABLE 1. These features have been successfully used to predict alternative splicing from raw genomic sequences[71,81–83].

There are several well-studied examples of these patterns in individual genes. A classic example is *BRCA1*, in which a several-fold plunge of the K_s rate between codons 200–300 overlaps with two putative splicing-enhancer elements[63,84]. This *BRCA1* region undergoes extensive alternative splicing in several species[85–88] (see REF. 89 for a review). The plunge of K_s in this region was suggested to reflect selection on two putative elements important for alternative-splicing regulation[84]. BOX 1 provides an example of how the hypothesis of RNA selection pressure has been tested experimentally.

A recent comparative-genomics study of alternatively spliced exons in five well-studied genes found markedly higher conservation of flanking intron sequences than of constitutive exons in human and mouse[78]. The study showed that the flanking introns of minor-form alternatively spliced exons of *NF1* are much more conserved (typically 20% higher identity) than those around the constitutive exons. This region of conservation extends at least 300 nt into the flanking introns. In one example (exon 10a-2), the region is strongly conserved in mammals, birds and fish, but not *D. melanogaster*, *Caenorhabditis elegans* or fungi. Minor-form alternative exons of *CFTR*, *PER3*, *CARS* and *SYT7* also show such increased conservation of flanking introns[78]. Intriguingly, the four major-form alternative exons in *NF1* and *CFTR* show no such increase in flanking-intron conservation. These examples strongly corroborate the genome-wide analyses described above.

Many previously surprising findings are easier to understand in light of the model of RNA selection pressure. For example, the *Drosophila* ultrabithorax (*Ubx*) gene contains two alternatively spliced exons (microexons I and II). A comparison of *Ubx* sequences from *D. melanogaster*, *D. virilis*, *D. pseudoobscura* and *D. hydei* reveals extremely low K_s rates for microexons I and II (approximately fivefold lower than in the constitutive 5′ and 3′ exons), suggesting that the nucleotide sequence itself is under purifying selection[90]. Similar results have been reported for alternative exons of genes that encode fast troponin T (REF. 91), α-crystallin[92] and CD45 (REF. 64).

Evolution of functional selection

EST data provide only a crude measure of exon-inclusion level based on a pool of ESTs from the whole body. So to gain further insight into the evolution of functional regulation of alternative exons, it is useful to move to microarray data, which can give accurate measurements of tissue-specific alternative splicing (for example, see REF. 6). One microarray study revealed that the inclusion levels in different tissues shift for some exons from being a minor form in some tissues to becoming the major form in other tissues. Such tissue switching is associated with dramatic changes in evolutionary selection pressure[93]. For example, only a small fraction of minor-form exons are conserved between mouse and human[6,13]. However, if an alternative exon shows tissue switching — if a minor form in nearly all tissues becomes the major form in only a single tissue — the probability that it is conserved between human and mouse becomes the same as for exons that are major form in all tissues and for constitutive exons[93]. Similarly, major-form exons show nearly random frame preservation[93,94]. However, if an alternative exon shows tissue switching — even if it is the major form in nearly all tissues, and becomes the minor form in only a single tissue — the frame-preservation ratio becomes as high as for conserved exons that are the minor form in all tissues[93]. This tissue-switching pattern is accompanied by a large increase in conservation (and decrease in K_s), as seen for conserved minor-form exons in general. A similar trend has been seen for an independent set of tissue-specific exons[80].

These data indicate that the evolution of tissue-specific regulation of alternative splicing is a major correlate of the slow evolution of functional selection pressure on alternative exons. Such exons show a strong selection

Table 1 | **Characteristics of ancestral alternatively spliced exons**

Characteristic	Ancestral alternatively spliced exons compared with constitutive exons	References
Exon size	Shorter	67,79,81
Frame preservation	More likely to be a multiple of 3 nucleotides	81–83,94,99
Intronic conservation	Highly conserved	15,65,77,81,82
Synonymous substitution rate	Decreased	65,67,68, 106,141
K_a/K_s ratio	Increased	65,68,142
Splice-site strength	Weaker	10,67,79

© 2006 Nature Publishing Group

Box 1 | **Experimental tests of RNA selection pressure**

Synonymous mutations that disrupt splicing can be subject to strong negative selection pressure. To directly test the hypothesis that synonymous mutations in an alternatively spliced exon cause deleterious functional consequences, Pagani *et al.* performed site-directed mutagenesis of *CFTR* exon 12 and measured its exon-inclusion level in a transient transfection-splicing assay[120]. Human *CFTR* exons 9 and 12 are alternatively spliced, and skipping of these exons renders the protein product non-functional, resulting in cystic fibrosis and other less severe phenotypes.

Human and mouse *CFTR* differ in seven synonymous mutations within exon 12; Pagani *et al.* found that introducing these substitutions into human *CFTR* causes 95% skipping of exon 12. The authors also tested single nucleotide mutations at 19 synonymous sites in exon 12: of these, six caused exon skipping, resulting in an inactive protein. Based on these results, it seems that approximately 30% of synonymous sites in exon 12 are under strong purifying selection pressure to maintain the correct splicing of *CFTR* mRNA. Similar studies of *CFTR* exon 9 found that about a quarter of synonymous mutations lead to a splicing defect[139]. Similar results were also found for synonymous mutations in *NF1* exon 37 (REF. 120).

pressure to preserve the reading frame, implying that the tissue-specific splice form produces a functional protein product with a novel peptide sequence inserted in a modular fashion (without disrupting the downstream protein reading frame). It seems logical that such tissue-specific splicing would require additional splicing regulatory motifs[95], imposing stronger purifying selection pressure on K_s and the flanking introns. It is possible that a large fraction of conserved minor-form exons are actually upregulated in only a very specific tissue or cell type. High-resolution microarray studies to measure inclusion levels in many tissues and cell types and/or differentiation states will be required to assess this hypothesis.

Recent studies have contributed to a complex picture of how alternative splicing affects molecular evolution. On the one hand, alternative splicing is associated with a large reduction in selection pressure on amino-acid substitutions (increase in K_a/K_s ratio and even in the absolute value of K_a). On the other hand, it is associated with a large increase in selection pressure against nucleotide substitutions (reduction in K_s). The change of K_a in alternative exons (compared with constitutive exons) does not follow the change of K_s and even moves in the opposite direction. This observation is surprising, because a typical splicing regulatory motif spans a few codons[96], so selection on those motifs should affect both synonymous and non-synonymous sites[72]. The most plausible explanation is that two types of selection pressure — selection on proteins and selection on RNA regulatory motifs[97] — are acting very differently on alternatively spliced exons[84].

To gain further insight into this problem, it is useful to consider how the strengths of these selection pressures change over time. Comparisons of human and chimpanzee (separated by approximately 5 million years of evolution), rat and mouse (40 million years), and human and mouse (90 million years) genomes provide a broad range of timescales[98] for tracking these effects. The effect of alternative splicing on the K_a/K_s ratio is essentially constant over this time range; even at the shortest timescale (5 million years), the K_a/K_s ratio is much higher for alternative exons (and particularly minor-form exons) than for constitutive exons[65]. One possible interpretation for this observation is that a substantial fraction of minor-form exons at any given time is nearly neutral; for example, if newly created minor-form exons have no current function. Such exons would be expected to have a high K_a/K_s value. Over time, most such exons would probably be lost owing to mutation.

By contrast, the effects of alternative splicing on K_s seem to evolve more slowly. At the shortest timescale (5 million years), K_s is only 25% lower for minor-form exons relative to major-form exons. By 40 million years, K_s drops to nearly threefold lower for minor-form exons, and by 90 million years to nearly fourfold lower[65]. One possible interpretation for these data is that non-functional alternative exons are gradually lost over time, and only those that have acquired both a useful function and RNA motifs that regulate their splicing (resulting in purifying selection pressure on K_s) are retained at longer timescales. According to this interpretation, this 'functionation' process takes substantially longer than 5 million years, and is mostly complete by 40 million years. These results are reinforced by a different measure of functional selection pressure: the preservation of the protein reading frame, which follows exactly the same trend during the time course of mammalian evolution[65]. Preservation of the protein reading frame has particular relevance to alternatively spliced exons, as it determines whether exon skipping will alter the downstream reading frame during translation. Many studies have reported that ancestral alternatively spliced exons show increased preservation of the protein reading frame[15,79,82,83,94,99].

The importance of ESEs and ESSs

There is a major gap between the evidence of strong RNA selection pressure associated with alternative splicing and our current knowledge about the actual splicing regulatory sites, splicing factors and molecular mechanisms that underlie this selection pressure. Several studies have asked whether the actual splicing regulatory mechanisms can be identified on the basis of these regions of conservation. For example, Fairbrother *et al.* used statistical analyses of a large data set of human exons and introns to find motifs that were specifically enriched in exons with weak splice sites[100]. Similar methods have been used to identify exonic splicing silencers (ESSs) — exonic sequence motifs that inhibit exon recognition[101–103].

Using polymorphism and divergence data, several studies provided systematic evidence for purifying selection on exonic splicing enhancers (ESEs) — exonic sequence motifs that promote exon recognition[104–106]. Biased codon usage[107] and, more interestingly, reduced K_s near exon–intron boundaries have also been observed[75]. Nevertheless, the evidence of a link between putative ESE conservation and alternative splicing remains less clear[106]. First, the putative ESEs do not seem to explain the reduced K_s in alternatively spliced exons, and non-ESE regions in these exons show a more than twofold reduction in K_s. Second, alternatively spliced exons show no increase in predicted ESE density compared with constitutive exons. It remains possible that an increase

© 2006 Nature Publishing Group

in functional ESEs in alternative exons is obscured by a large fraction of false positives (random matches to the ESE hexamer sequences that are not functional). Some data suggest that K_s is slightly lower in predicted ESEs than in non-ESE regions of the same alternative exons[106], but this does not fully resolve the paradoxical results.

Many questions remain unanswered. The first problem is that our knowledge about functional ESEs, ESSs and intronic splicing enhancers and silencers might be incomplete. The high degree of conservation of alternative exons is puzzling: K_s is reduced by at least a factor of two for all categories of alternatively spliced exons combined[65,67,106]. Adopting the most conservative assumptions, that all synonymous sites in constitutive exons are selectively neutral and that RNA selection pressure within alternatively spliced exons is confined to the smallest possible region (in which no mutations are tolerated), a twofold reduction in K_s implies that essential splicing regulatory sites span at least half of the exon. Indeed, many of the alternatively spliced exons are part of the ultraconserved regions[76]. One possible explanation is that an alternatively spliced exon might contain several binding sites for splicing regulators. Another intriguing possibility is that formation of long RNA secondary structures is required for regulation of alternative splicing.

RNA secondary structure and splicing regulation

Complementing some early examples of RNA secondary structures that are involved in alternative splicing of troponin T (REF. 108), β-tropomyosin[109] and hnRNP A1 (REF. 110), there has been a recent flood of experimental evidence for the role of RNA secondary structures in regulating alternative splicing[111–115] (see BOX 2 for an example and REF. 116 for a review). Genome-wide analyses indicate that splicing regulation through secondary structure might be a general mechanism[117–119]. For example, Meyer and Miklos identified a putative large stem-loop fold that includes nearly all of *CFTR* exon 12 (REF. 119), indicating that the sensitivity of *CFTR* splicing to synonymous mutations in exon 12 (REF. 120) might be due to destabilization of this stem-loop structure. The requirement for secondary structures provides another possible explanation for the high level of conservation around alternative exons. The conserved stem regions are typically long (50 nt total), similar in length to the conserved regions found in alternatively spliced exons.

Box 2 | RNA secondary structure in the alternative splicing of Dscam

The *Drosophila melanogaster* axon-guidance receptor Dscam undergoes extensive alternative splicing[140], in which RNA secondary structures have an important role. Splicing of its exon 4 cluster (containing 12 mutually exclusive exons) depends on a 27-bp stem (dubbed the iStem[114]) found in the intron between exons 3 and 4.1. Mutations that disrupt the stem abolish inclusion of exon 4, whereas compensatory mutations that restore base complementarity rescue splicing of exon 4. As exon 4 skipping occurs *in vivo*, and the iStem structure is conserved over 40 million years of *D. melanogaster* evolution, the iStem seems to be an important new splicing regulatory element, unique in that it regulates the inclusion of many exons simultaneously (the 12 exons of the exon 4 cluster). By contrast, another example of RNA folding in *Dscam* seems to enforce mutual exclusivity of splicing in the exon 6 cluster, which contains 48 alternative exons. A 50-nt 'docking sequence' in the intron between exons 5 and 6.1 is complementary to 48 different 'selector sequences', each of which is located upstream of an associated exon in the exon 6 cluster[113]. The docking sequence is highly conserved from *D. melanogaster* to honeybees (250 million years), and complementary selector sequences have been identified upstream of each exon 6 variant in species spanning this evolutionary range. Graveley points out that the fact that the docking sequence can only base pair with one selector sequence naturally imposes mutual exclusivity on the exon 6 cluster[113].

Splicing mutations and human disease

A major line of evidence for RNA selection pressure comes from studies of human diseases[97,121]. Mutations that alter mRNA splicing have long been considered to constitute at least 15% of human disease mutations, based on studies of cystic fibrosis and other diseases[122]. It has been argued that the true fraction might be significantly higher because effects on splicing have often not been considered in traditional studies of disease mutations. For example, in *CFTR*, a large fraction of common coding SNPs reduce splicing of exons 9 and 12 (REF. 123).

According to a recent provocative hypothesis, splicing mutations might be the most frequent cause of hereditary disease[124]. The hypothesis is based on the following simple model. If each possible amino-acid mutation has a certain risk of causing a disease phenotype, it follows that longer genes (with more sites for possible mutations) should have a higher risk for disease mutations. Similarly, if each splice site has a certain risk of causing a disease phenotype through mutation, then genes with more introns should have a higher risk of being disease genes, and this is indeed observed. The authors report that both the length of the coding region and the number of introns correlate with disease-gene propensity[124]. Multivariate analysis indicates that both factors are required to explain the observed data. For example, the fraction of intronless genes that are classified as disease genes (4.5%) is much lower than the corresponding fractions for single-intron genes (10.6%) and two-intron genes (12.2%), even though all three groups have similar gene-coding-length distributions. According to this model, approximately 60% of disease mutations in the human genome are splicing mutations[124].

Despite the possible existence of confounding factors in this analysis, several lines of evidence suggest that this hypothesis is worth considering. First, there is convincing evidence that splicing is associated with significant negative selection pressure. For example, comparing synonymous sites between humans and chimpanzees, the K_s rate in multi-exon genes is about half that observed in single-exon genes[124]. This implies that overall splicing imposes a strong selective constraint on about half of all sites in human coding regions. Similarly, Bustamante *et al.* observed a 30% reduction in the K_s rate of functional human genes compared with pseudogenes[125]. Second, alternative splicing seems to create even stronger selection pressure. It seems inevitable that all these extra sequence constraints will create associated risks of disease mutations, perhaps even more powerfully than amino-acid mutations (as splicing mutations typically remove a large portion of the amino-acid sequence). Third, the prediction that genes with many exons, and particularly those with multiple alternative splice forms, are at risk for disease

© 2006 Nature Publishing Group

mutations that affect splicing seems to match current experience. The classic examples of splicing mutations that are known to cause disease are all large, multi-exon genes with multiple splice forms: *CFTR*, *ATM*[126], *TAU*[127] and many others. Many such disease-causing mutations have been recently reported, for example in C5 (complement 5) deficiency[128], cancer[129,130], dystrophic epidermolysis bullosa[131], familial adenomatous polyposis[132], lysinergic protein intolerance[133], mental retardation[134], Pelizaeus–Merzbacher disease[135] and tyrosinaemia type I (REF. 136). Although it is still difficult to assess the overall role of splicing defects in human diseases, the pace of discovery of 'splicing disease' mutations is likely to accelerate in the future.

Conclusions

Although the importance of alternative splicing in various biological processes such as sex determination[137] and apoptosis[138] has been known for a long time, genomics, and in particular the shotgun sequencing of ESTs, have revealed its nearly ubiquitous role in gene regulation[5]. Genome sequencing has made it possible to study the evolutionary impact and constraints of alternative splicing. By creating an 'internal paralogue' for a functional gene, alternative splicing provides an important strategy for opening nearly neutral pathways for evolution of gene structure and recruitment of novel protein-coding sequences. Moreover, mammalian genomes show widespread evidence of RNA selection pressure owing to constraints of alternative splicing regulation, emphasizing the functional importance of alternative splicing patterns.

An important future area of research will involve understanding the relative contribution of *cis*-regulatory elements (such as splice sites, enhancers and silencers) and *trans*-acting factors to the evolution of alternative splicing. An exciting opportunity to begin to dissect these contributions comes from the continued sequencing of closely related species. These comparative-genomics data will allow us to reconstruct the evolutionary history of individual exons and introns at a single-nucleotide resolution, which in turn will enhance our knowledge about the *cis*-splicing regulatory code and its impact on the evolutionary divergence of alternative splicing. Parallel evolution of splicing factors that recognize those *cis*-regulatory signals is another driving force for the divergence of alternative splicing[24]. How alternative splicing is affected by the evolutionary reconfiguration of the splicing machinery remains to be further investigated.

1. Berget, S. M. & Sharp, P. A. A spliced sequence at the 5′-terminus of adenovirus late mRNA. *Brookhaven Symp. Biol.* **29**, 332–344 (1977).
2. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12**, 1–8 (1977).
3. Gilbert, W. Why genes in pieces? *Nature* **271**, 501 (1978).
4. Mount, S. & Steitz, J. Lessons from mutant globins. *Nature* **303**, 380–381 (1983).
5. Modrek, B. & Lee, C. A genomic view of alternative splicing. *Nature Genet.* **30**, 13–19 (2002).
6. Pan, Q. *et al.* Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Mol. Cell* **16**, 929–941 (2004).
7. Johnson, J. M. *et al.* Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**, 2141–2144 (2003).
Microarray profiling of alternative splicing in humans suggests that 74% of human multi-exon genes are alternatively spliced.
8. Kan, Z., Garrett-Engele, P. W., Johnson, J. M. & Castle, J. C. Evolutionarily conserved and diverged alternative splicing events show different expression and functional profiles. *Nucleic Acids Res.* **33**, 5659–5666 (2005).
9. Kan, Z., States, D. & Gish, W. Selecting for functional alternative splices in ESTs. *Genome Res.* **12**, 1837–1845 (2002).
10. Itoh, H., Washio, T. & Tomita, M. Computational comparative analyses of alternative splicing regulation using full-length cDNA of various eukaryotes. *RNA* **10**, 1005–1018 (2004).
11 Pan, Q. *et al.* Alternative splicing of conserved exons is frequently species-specific in human and mouse. *Trends Genet.* **21**, 73–77 (2005).
12. Nurtdinov, R. N., Artamonova, I. I., Mironov, A. A. & Gelfand, M. S. Low conservation of alternative splicing patterns in the human and mouse genomes. *Hum. Mol. Genet.* **12**, 1313–1320 (2003).
13. Modrek, B. & Lee, C. Alternative splicing in the human, mouse and rat genomes is associated with an increased rate of exon creation/loss. *Nature Genet.* **34**, 177–180 (2003).
By comparing gene structure of orthologous genes in human, mouse and rat genomes, the authors found that alternative splicing was associated with a significant increase in the rate of exon creation and loss in mammals. Subsequent outgroup analyses indicate that most of the non-conserved exons are recently created (see also references 29 and 45).
14. Thanaraj, T. A., Clark, F. & Muilu, J. Conservation of human alternative splice events in mouse. *Nucleic Acids Res.* **31**, 2544–2552 (2003).
15. Sugnet, C. W., Kent, W. J., Ares, M. Jr & Haussler, D. Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac. Symp. Biocomput.* 66–77 (2004).
16 Nagasaki, H., Arita, M., Nishizawa, T., Suwa, M. & Gotoh, O. Species-specific variation of alternative splicing and transcriptional initiation in six eukaryotes. *Gene* **364**, 53–62 (2005).
17. Smith, C. W. J. & Valcarcel, J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* **25**, 381–388 (2000).
18. Hiller, M. *et al.* Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. *Nature Genet.* **36**, 1255–1257 (2004).
19. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
20. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
21 Graveley, B. R. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* **17**, 100–107 (2001).
22. Lopez, P. J. & Seraphin, B. YIDB: the Yeast Intron DataBase. *Nucleic Acids Res.* **28**, 85–86 (2000).
23. Davis, C. A., Grate, L., Spingola, M. & Ares, M. Jr. Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Res.* **28**, 1700–1706 (2000).
24. Boucher, L., Ouzounis, C. A., Enright, A. J. & Blencowe, B. J. A genome-wide survey of RS domain proteins. *RNA* **7**, 1693–1701 (2001).
25. Brett, D., Pospisil, H., Valcarcel, J., Reich, J. & Bork, P. Alternative splicing and genome complexity. *Nature Genet.* **30**, 29–30 (2002).
The authors used EST data to estimate the frequency of alternative splicing in seven species. They found no evidence for an increased level of alternative splicing in mammals compared with *C. elegans* and *D. melanogaster*.
26. Kim, H., Klein, R., Majewski, J. & Ott, J. Estimating rates of alternative splicing in mammals and invertebrates. *Nature Genet.* **36**, 915–916; author reply 916–917 (2004).
27. Long, M., Betran, E., Thornton, K. & Wang, W. The origin of new genes: glimpses from the young and old. *Nature Rev. Genet.* **4**, 865–875 (2003).
28. Taylor, J. S. & Raes, J. Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* **38**, 615–643 (2004).
29. Cusack, B. P. & Wolfe, K. H. Changes in alternative splicing of human and mouse genes are accompanied by faster evolution of constitutive exons. *Mol. Biol. Evol.* **22**, 2198–2208 (2005).
30. Lareau, L. F., Green, R. E., Bhatnagar, R. S. & Brenner, S. E. The evolving roles of alternative splicing. *Curr. Opin. Struct. Biol.* **14**, 273–282 (2004).
31. Boue, S., Letunic, I. & Bork, P. Alternative splicing and evolution. *Bioessays* **25**, 1031–1034 (2003).
32. Ast, G. How did alternative splicing evolve? *Nature Rev. Genet.* **5**, 773–782 (2004).
33. Makalowski, W., Mitchell, G. A. & Labuda, D. *Alu* sequences in the coding regions of mRNA: a source of protein variability. *Trends Genet.* **10**, 188–193 (1994).
34. Nekrutenko, A. & Li, W. H. Transposable elements are found in a large number of human protein-coding genes. *Trends Genet.* **17**, 619–621 (2001).
35. Sorek, R., Ast, G. & Graur, D. *Alu*-containing exons are alternatively spliced. *Genome Res.* **12**, 1060–1067 (2002).
36. Lev-Maor, G., Sorek, R., Shomron, N. & Ast, G. The birth of an alternatively spliced exon: 3′ splice-site selection in *Alu* exons. *Science* **300**, 1288–1291 (2003).
37. Singer, S. S., Mannel, D. N., Hehlgans, T., Brosius, J. & Schmitz, J. From "junk" to gene: curriculum vitae of a primate receptor isoform gene. *J. Mol. Biol.* **341**, 883–886 (2004).
The authors performed a phylogenetic analysis to track the birth of a *p75TNFR* alternative exon from an *Alu* element in primate genomes.
38. Krull, M., Brosius, J. & Schmitz, J. *Alu*–SINE exonization: en route to protein-coding function. *Mol. Biol. Evol.* **22**, 1702–1711 (2005).
39. Gotea, V. & Makalowski, W. Do transposable elements really contribute to proteomes? *Trends Genet.* **22**, 260–277 (2006).
40 Hillman, R. T., Green, R. E. & Brenner, S. E. An unappreciated role for RNA surveillance. *Genome Biol.* **5**, R8 (2004).
41. Letunic, I., Copley, R. R. & Bork, P. Common exon duplication in animals and its role in alternative splicing. *Hum. Mol. Genet.* **11**, 1561–1567 (2002).
42. Kondrashov, F. A. & Koonin, E. V. Origin of alternative splicing by tandem exon duplication. *Hum. Mol. Genet.* **10**, 2661–2669 (2001).
43. Copley, R. R. Evolutionary convergence of alternative splicing in ion channels. *Trends Genet.* **20**, 171–176 (2004).

© 2006 Nature Publishing Group

The author argues that the alternative splicing of tandem exons in ion-channel genes is the result of convergent evolution, that is, independent exon duplication followed by alternative splicing in multiple lineages. This study provides evidence for adaptive benefits of alternative splicing in genes encoding ion channels.

44. Malko, D. B., Makeev, V. J., Mironov, A. A. & Gelfand, M. S. Evolution of exon–intron structure and alternative splicing in fruit flies and malarial mosquito genomes. *Genome Res.* **16**, 505–509 (2006).
45. Wang, W. *et al.* Origin and evolution of new exons in rodents. *Genome Res.* **15**, 1258–1264 (2005).
46. Kondrashov, F. A. & Koonin, E. V. Evolution of alternative splicing: deletions, insertions and origin of functional parts of proteins from intron sequences. *Trends Genet.* **19**, 115–119 (2003).
47. Artamonova, I. I. & Gelfand, M. S. Evolution of the exon–intron structure and alternative splicing of the *MAGE-A* family of cancer/testis antigens. *J. Mol. Evol.* **59**, 620–631 (2004).
48. Terai, Y., Morikawa, N., Kawakami, K. & Okada, N. The complexity of alternative splicing of hagoromo mRNAs is increased in an explosively speciated lineage in East African cichlids. *Proc. Natl Acad. Sci. USA* **100**, 12798–12803 (2003).
49. Lynch, M. & Kewalramani, A. Messenger RNA surveillance and the evolutionary proliferation of introns. *Mol. Biol. Evol.* **20**, 563–571 (2003).
50. Maquat, L. E. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nature Rev. Mol. Cell Biol.* **5**, 89–99 (2004).
51. Masel, J. Cryptic genetic variation is enriched for potential adaptations. *Genetics* **172**, 1985–1991 (2006).
52. Xing, Y. & Lee, C. Negative selection pressure against premature protein truncation is reduced by alternative splicing and diploidy. *Trends Genet.* **20**, 472–475 (2004).
53. Rehwinkel, J., Letunic, I., Raes, J., Bork, P. & Izaurralde, E. Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. *RNA* **11**, 1530–1544 (2005).
54. Lewis, B. P., Green, R. E. & Brenner, S. E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl Acad. Sci. USA* **100**, 189–192 (2003).
55 Pan, Q. *et al.* Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. *Genes Dev.* **20**, 153–158 (2006).
56. Yu, W. P., Brenner, S. & Venkatesh, B. Duplication, degeneration and subfunctionalization of the nested synapsin-Timp genes in *Fugu*. *Trends Genet.* **19**, 180–183 (2003).
57. Altschmied, J. *et al.* Subfunctionalization of duplicate *mitf* genes associated with differential degeneration of alternative exons in fish. *Genetics* **161**, 259–267 (2002).
58. Force, A. *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545 (1999).
59. Kopelman, N. M., Lancet, D. & Yanai, I. Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. *Nature Genet.* **37**, 588–589 (2005).
60. Su, Z., Wang, J., Yu, J., Huang, X. & Gu, X. Evolution of alternative splicing after gene duplication. *Genome Res.* **16**, 182–189 (2006).
References 59 and 60 show a significant inverse correlation between the frequency of alternative splicing and the size of a multigene family, suggesting that gene duplication was frequently followed by the loss of alternative splicing.
61. Hurst, L. D. The K_a/K_s ratio: diagnosing the form of sequence evolution. *Trends Genet.* **18**, 486 (2002).
62. Yang, Z. & Nielsen, R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**, 32–43 (2000).
63. Hurst, L. D. & Pal, C. Evidence for purifying selection acting on silent sites in *BRCA1*. *Trends Genet.* **17**, 62–65 (2001).
A several-fold plunge in K_s rate is correlated with sites of alternative splicing regulation for *BRCA1* (see also reference 84).
64. Filip, L. C. & Mundy, N. I. Rapid evolution by positive Darwinian selection in the extracellular domain of the abundant lymphocyte protein CD45 in primates. *Mol. Biol. Evol.* **21**, 1504–1511 (2004).
65 Xing, Y. & Lee, C. Evidence of functional selection pressure for alternative splicing events that accelerate evolution of protein subsequences. *Proc. Natl Acad. Sci. USA* **102**, 13526–13531 (2005).
66. Iida, K. & Akashi, H. A test of translational selection at 'silent' sites in the human genome: base composition comparisons in alternatively spliced genes. *Gene* **261**, 93–105 (2000).
67. Baek, D. & Green, P. Sequence conservation, relative isoform frequencies, and nonsense-mediated decay in evolutionarily conserved alternative splicing. *Proc. Natl Acad. Sci. USA* **102**, 12813–1288 (2005).
68. Chen, F. C., Wang, S. S., Chen, C. J., Li, W. H. & Chuang, T. J. Alternatively and constitutively spliced exons are subject to different evolutionary forces. *Mol. Biol. Evol.* **23**, 675–682 (2006).
References 65–68 show that evolutionarily conserved alternative exons on average have a significantly reduced K_s but a nearly constant, or even increased, K_a. These data suggest that two types of selection pressure — selection on proteins and selection on RNA splicing motifs — act very differently on alternatively spliced exons.
69. Nekrutenko, A., Makova, K. D. & Li, W. H. The K_a/K_s ratio test for assessing the protein-coding potential of genomic regions: an empirical and simulation study. *Genome Res.* **12**, 198–202 (2002).
70 Xing, Y. & Lee, C. Assessing the application of K_a/K_s ratio test to alternatively spliced exons. *Bioinformatics* **21**, 3701–3703 (2005).
71. Ohler, U., Shomron, N. & Burge, C. B. Recognition of unknown conserved alternatively spliced exons. *PLoS Comput. Biol.* **1**, 113–122 (2005).
72. Xing, Y. & Lee, C. Can RNA selection pressure distort the measurement of K_a/K_s? *Gene* **370**, 1–5 (2006).
73. Hirsh, A. E., Fraser, H. B. & Wall, D. P. Adjusting for selection on synonymous sites in estimates of evolutionary distance. *Mol. Biol. Evol.* **22**, 174–177 (2005).
74. Zhang, L. & Li, W. H. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. *Mol. Biol. Evol.* **21**, 236–239 (2004).
75 Chamary, J. V., Parmley, J. L. & Hurst, L. D. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nature Rev. Genet.* **7**, 98–108 (2006).
An excellent review on selection at silent sites in mammals.
76. Bejerano, G. *et al.* Ultraconserved elements in the human genome. *Science* **304**, 1321–1325 (2004).
77 Sorek, R. & Ast, G. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res.* **13**, 1631–1637 (2003).
Evolutionarily conserved alternative exons have highly conserved flanking intronic regions, suggesting purifying selection on intronic sites that are important for alternative splicing regulation (see also reference 80).
78. Kaufmann, D., Kenner, O., Nurnberg, P., Vogel, W. & Bartelt, B. In *NF1*, *CFTR*, *PER3*, *CARS* and *SYT7*, alternatively included exons show higher conservation of surrounding intron sequences than constitutive exons. *Eur. J. Hum. Genet.* **12**, 139–149 (2004).
79. Zheng, C. L., Fu, X. D. & Gribskov, M. Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse. *RNA* **11**, 1777–1787 (2005).
80. Sugnet, C. W. *et al.* Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Comput. Biol.* **2**, e4 (2006).
81. Sorek, R. *et al.* A non-EST-based method for exon-skipping prediction. *Genome Res.* **14**, 1617–1623 (2004).
82. Philipps, D. L., Park, J. W. & Graveley, B. R. A computational and experimental approach toward *a priori* identification of alternatively spliced exons. *RNA* **10**, 1838–1844 (2004).
83. Yeo, G. W., Van Nostrand, E., Holste, D., Poggio, T. & Burge, C. B. Identification and analysis of alternative splicing events conserved in human and mouse. *Proc. Natl Acad. Sci. USA* **102**, 2850–2855 (2005).
84 Orban, T. I. & Olah, E. Purifying selection on silent sites — a constraint from splicing regulation? *Trends Genet.* **17**, 252–253 (2001).
85. Orban, T. I. & Olah, E. Expression profiles of *BRCA1* splice variants in asynchronous and in G1/S synchronized tumor cell lines. *Biochem. Biophys. Res. Commun.* **280**, 32–38 (2001).
86. Hakem, R. *et al.* The tumor suppressor gene *Brca1* is required for embryonic cellular proliferation in the mouse. *Cell* **85**, 1009–1023 (1996).
87. Xu, X. *et al.* Centrosome amplification and a defective G2–M cell cycle checkpoint induce genetic instability in *BRCA1* exon 11 isoform-deficient cells. *Mol. Cell* **3**, 389–395 (1999).
88. Mixon, M., Kittrell, F. & Medina, D. Expression of *Brca1* and splice variant *Brca1 δ11* RNA levels in mouse mammary gland during normal development and tumorigenesis. *Oncogene* **19**, 5237–5243 (2000).
89. Orban, T. I. & Olah, E. Emerging roles of *BRCA1* alternative splicing. *Mol. Pathol.* **56**, 191–197 (2003).
90. Bomze, H. M. & Lopez, A. J. Evolutionary conservation of the structure and expression of alternatively spliced Ultrabithorax isoforms from *Drosophila*. *Genetics* **136**, 965–977 (1994).
91. Bucher, E. A., Dhoot, G. K., Emerson, M. M., Ober, M. & Emerson, C. P. Jr. Structure and evolution of the alternatively spliced fast troponin T isoform gene. *J. Biol. Chem.* **274**, 17661–17670 (1999).
92. van Dijk, M. A., Sweers, M. A. & de Jong, W. W. The evolution of an alternatively spliced exon in the αA-crystallin gene. *J. Mol. Evol.* **52**, 510–515 (2001).
93. Xing, Y. & Lee, C. Protein modularity of alternatively spliced exons is associated with tissue-specific regulation of alternative splicing. *PLoS Genet.* **1**, e34 (2005).
94. Resch, A., Xing, Y., Alekseyenko, A., Modrek, B. & Lee, C. Evidence for a subpopulation of conserved alternative splicing events under selection pressure for protein reading frame preservation. *Nucleic Acids Res.* **32**, 1261–1269 (2004).
95. Han, K., Yeo, G., An, P., Burge, C. B. & Grabowski, P. J. A combinatorial code for splicing silencing: UAGG and GGGG motifs. *PLoS Biol.* **3**, e158 (2005).
96. Zheng, Z. M. Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression. *J. Biomed. Sci.* **11**, 278–294 (2004).
97. Pagani, F. & Baralle, F. E. Genomic variants in exons and introns: identifying the splicing spoilers. *Nature Rev. Genet.* **5**, 389–396 (2004).
The authors review two types of selection pressure on exons: selection on proper RNA splicing and selection on protein function.
98. Hedges, S. B. The origin and evolution of model organisms. *Nature Rev. Genet.* **3**, 838–849 (2002).
99. Sorek, R., Shamir, R. & Ast, G. How prevalent is functional alternative splicing in the human genome? *Trends Genet.* **20**, 68–71 (2004).
100. Fairbrother, W. G., Yeh, R. F., Sharp, P. A. & Burge, C. B. Predictive identification of exonic splicing enhancers in human genes. *Science* **297**, 1007–1013 (2002).
101. Wang, Z. *et al.* Systematic identification and analysis of exonic splicing silencers. *Cell* **119**, 831–845 (2004).
102. Zhang, X. H., Leslie, C. S. & Chasin, L. A. Computational searches for splicing signals. *Methods* **37**, 292–305 (2005).
103. Zhang, X. H. & Chasin, L. A. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev.* **18**, 1241–1250 (2004).
104. Fairbrother, W. G., Holste, D., Burge, C. B. & Sharp, P. A. Single nucleotide polymorphism-based validation of exonic splicing enhancers. *PLoS Biol.* **2**, e268 (2004).
105. Carlini, D. B. & Genut, J. E. Synonymous SNPs provide evidence for selective constraint on human exonic splicing enhancers. *J. Mol. Evol.* **62**, 89–98 (2005).
106. Parmley, J. L., Chamary, J. V. & Hurst, L. D. Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. *Mol. Biol. Evol.* **23**, 301–309 (2006).
107. Chamary, J. V. & Hurst, L. D. Biased codon usage near intron–exon junctions: selection on splicing enhancers, splice-site recognition or something else? *Trends Genet.* **21**, 256–259 (2005).
108. Cooper, T. A., Cardone, M. H. & Ordahl, C. P. *Cis* requirements for alternative splicing of the cardiac troponin T pre-mRNA. *Nucleic Acids Re.s* **16**, 8443–8465 (1988).
109. Sirand-Pugnet, P., Durosay, P., Clouet d'Orval, B. C., Brody, E. & Marie, J. β-Tropomyosin pre-mRNA folding around a muscle-specific exon interferes with several steps of spliceosome assembly. *J. Mol. Biol.* **251**, 591–602 (1995).
110. Blanchette, M. & Chabot, B. A highly stable duplex structure sequesters the 5′ splice site region of hnRNP A1 alternative exon 7B. *RNA* **3**, 405–419 (1997).
111. Baraniak, A. P., Lasda, E. L., Wagner, E. J. & Garcia-Blanco, M. A. A stem structure in fibroblast growth factor receptor 2 transcripts mediates cell-type-specific splicing by approximating intronic control elements. *Mol. Cell. Biol.* **23**, 9327–9337 (2003).

© 2006 Nature Publishing Group

112. Hefferon, T. W., Groman, J. D., Yurk, C. E. & Cutting, G. R. A variable dinucleotide repeat in the *CFTR* gene contributes to phenotype diversity by forming RNA secondary structures that alter splicing. *Proc. Natl Acad. Sci. USA* **101**, 3504–3509 (2004).
113. Graveley, B. R. Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* **123**, 65–73 (2005).
114. Kreahling, J. M. & Graveley, B. R. The iStem, a long-range RNA secondary structure element required for efficient exon inclusion in the *Drosophila* Dscam pre-mRNA. *Mol. Cell. Biol.* **25**, 10251–10260 (2005).
References 113 and 114 show the importance of RNA secondary structure in alternative splicing of the *Drosophila* axon-guidance receptor Dscam.
115. McAlinden, A., Havlioglu, N., Liang, L., Davies, S. R. & Sandell, L. J. Alternative splicing of type II procollagen exon 2 is regulated by the combination of a weak 5′ splice site and an adjacent intronic stem-loop *cis* element. *J. Biol. Chem.* **280**, 32700–32711 (2005).
116. Buratti, E. & Baralle, F. E. Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol. Cell. Biol.* **24**, 10505–10514 (2004).
117. Chamary, J. V. & Hurst, L. D. Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biol.* **6**, R75 (2005).
118. Lian, Y. & Garner, H. R. Evidence for the regulation of alternative splicing via complementary DNA sequence repeats. *Bioinformatics* **21**, 1358–1364 (2005).
119. Meyer, I. M. & Miklos, I. Statistical evidence for conserved, local secondary structure in the coding regions of eukaryotic mRNAs and pre-mRNAs. *Nucleic Acids Res.* **33**, 6338–6348 (2005).
120. Pagani, F., Raponi, M. & Baralle, F. E. Synonymous mutations in *CFTR* exon 12 affect splicing and are not neutral in evolution. *Proc. Natl Acad. Sci. USA* **102**, 6368–6372 (2005).
Experimental tests of RNA selection pressure. The authors systematically introduced synonymous mutations to exon 12 of *CFTR*. Approximately 30% of those mutations induced exon skipping and resulted in an inactive protein.
121. Cartegni, L., Chew, S. L. & Krainer, A. R. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Rev. Genet.* **3**, 285–298 (2002).
122. Krawczak, M., Reiss, J. & Cooper, D. N. The mutational spectrum of single basepair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum. Genet.* **90**, 41–54 (1992).
123. Steiner, B., Truninger, K., Sanz, J., Schaller, A. & Gallati, S. The role of common single-nucleotide polymorphisms on exon 9 and exon 12 skipping in nonmutated *CFTR* alleles. *Hum. Mutat.* **24**, 120–129 (2004).
124. Lopez-Bigas, N., Audit, B., Ouzounis, C., Parra, G. & Guigo, R. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett.* **579**, 1900–1903 (2005).
Based on the disease-gene propensity of human genes in terms of their coding-region length and intron number, the authors estimate that ~60% of human disease mutations are splicing mutations.
125. Bustamante, C. D., Nielsen, R. & Hartl, D. L. A maximum likelihood method for analyzing pseudogene evolution: implications for silent site evolution in humans and rodents. *Mol. Biol. Evol.* **19**, 110–117 (2002).
126. Eng, L. *et al.* Nonclassical splicing mutations in the coding and noncoding regions of the *ATM* gene: maximum entropy estimates of splice junction strengths. *Hum. Mutat.* **23**, 67–76 (2004).
127. Buee, L., Bussiere, T., Buee-Scherrer, V., Delacourte, A. & Hof, P. R. τ Protein isoforms, phosphorylation and role in neurodegenerative disorders. *Brain Res. Rev.* **33**, 95–130 (2000).
128. Pfarr, N. *et al.* Linking C5 deficiency to an exonic splicing enhancer mutation. *J. Immunol.* **174**, 4172–4177 (2005).
129. Chen, L. L. *et al.* A mutation-created novel intra-exonic pre-mRNA splice site causes constitutive activation of KIT in human gastrointestinal stromal tumors. *Oncogene* **24**, 4271–4280 (2005).
130. Harland, M. *et al.* A mutation hotspot at the p14ARF splice site. *Oncogene* **24**, 4604–4608 (2005).
131. Masse, M. *et al.* Identification of novel type VII collagen gene mutations resulting in severe recessive dystrophic epidermolysis bullosa. *Clin. Exp. Dermatol.* **30**, 289–293 (2005).
132. Garcia-Lozano, J. R. *et al. APC* germ-line mutations in southern Spanish patients with familial adenomatous polyposis: genotype–phenotype correlations and identification of eight novel mutations. *Genet. Test.* **9**, 37–40 (2005).
133. Sperandeo, M. P. *et al.* Lysinuric protein intolerance: identification and functional analysis of mutations of the *SLC7A7* gene. *Hum. Mutat.* **25**, 410 (2005).
134. Gabut, M. *et al.* The SR protein SC35 is responsible for aberrant splicing of the E1α pyruvate dehydrogenase mRNA in a case of mental retardation with lactic acidosis. *Mol. Cell. Biol.* **25**, 3286–3294 (2005).
135. Hubner, C. A. *et al.* Mild Pelizaeus–Merzbacher disease caused by a point mutation affecting correct splicing of *PLP1* mRNA. *Neuroscience* **132**, 697–701 (2005).
136. Bliksrud, Y. T., Brodtkorb, E., Andresen, P. A., van den Berg, I. E. & Kvittingen, E. A. Tyrosinaemia type I — *de novo* mutation in liver tissue suppressing an inborn splicing defect. *J. Mol. Med.* **83**, 406–410 (2005).
137. Lopez, A. J. Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.* **32**, 279–305 (1998).
138. Boise, L. H. *et al. bcl-x*, a *bcl-2*-related gene that functions as a dominant regulator of apoptotic cell death. *Cell* **74**, 597–608 (1993).
139. Pagani, F., Buratti, E., Stuani, C. & Baralle, F. E. Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9. *J. Biol. Chem.* **278**, 26580–26588 (2003).
140. Schmucker, D. *et al. Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**, 671–684 (2000).
141. Schattner, P. & Diekhans, M. Regions of extreme synonymous codon selection in mammalian genes. *Nucleic Acids Res.* **34**, 1700–1710 (2006).
142. Ermakova, E. O., Nurtdinov, R. N. & Gelfand, M. S. Fast rate of evolution in alternatively spliced coding regions of mammalian genes. *BMC Genomics* **7**, 84 (2006).

Acknowledgements
We thank D. Black and B. Blencowe for comments on our manuscript, and S. Mount for discussions. This work was supported by grants from the US National Institutes of Health and the US Department of Energy, and a Ph.D. dissertation fellowship from the University of California, Los Angeles, to Y.X.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Lee laboratory homepage:
http://www.bioinformatics.ucla.edu/leelab
RCSB Protein Data Bank: http://www.rcsb.org/pdb
Access to this links box is available online.

© 2006 Nature Publishing Group