Gene 504 (2012) 102–106

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Short Communication

Revisit on the evolutionary relationship between alternative splicing and gene duplication

Zhixi Su [a], Xun Gu [a,b,*]

[a] MOE Key Laboratory of Contemporary Anthropology and Center for Evolutionary Biology, School of Life Sciences, Fudan University, Shanghai 200433, China
[b] Department of Genetics, Developmental and Cell Biology, Iowa State University, Ames, IA 50011, USA

ARTICLE INFO

Article history:
Accepted 7 May 2012
Available online 14 May 2012

Keywords:
Gene duplication
Alternative splicing
Copy number variation
Exon–intron structure

ABSTRACT

Gene duplications and alternative splicing (AS) isoforms are two widespread types of genetic variations that can facilitate diversification of protein function. A number of studies claimed that after gene duplication, two AS isoforms with differential functions can be 'fixed', respectively, in each of the duplicate copies. This simple 'functional-sharing' hypothesis was recently challenged by Roux and Robinson-Rechavi (2011). Instead, they proposed a more sophisticated hypothesis, invoking that less alternative splicing genes tend to be duplicated more frequently, and single-copy genes are younger than duplicate genes, or the 'duplicability-age' hypothesis for short. In this letter, we show that all these genome-wide analyses of AS isoforms actually did not provide clear-cut evidence to nullify the basic idea of functional-sharing hypothesis. After updating our understanding of genome-wide alternative splicing, duplicability and CNV (copy number variation), we argue that the foundation of the duplicability-age hypothesis remains to be justified carefully. Finally, we suggest that a better approach to resolving this controversy is the correspondence analysis of indels (insertions and deletions) between duplicate genes to the genomic exon–intron structure, which can be used to experimentally test the effect of functional-sharing hypothesis.

1. Introduction

Since gene duplication (GD) and alternative splicing (AS) are two widespread phenomena for increasing diversifications of protein function (Barbazuk et al., 2008; Chothia et al., 2003; Graveley, 2001; Gu et al., 2002; Koonin and Wolf, 2010; Lynch and Conery, 2000; Wang et al., 2008), the evolutionary relationship between GD and AS has been extensively studied (Birzele et al., 2008; Chothia et al., 2003; Copley, 2004; Keren et al., 2010; Kim et al., 2004; Lareau et al., 2004; Shabalina et al., 2010). Several authors (e.g., Irimia et al., 2008; Jin et al., 2008; Kopelman et al., 2005; Su et al., 2006; Talavera et al., 2007; Zhang et al., 2010) proposed that after gene duplication, two alternative splicing isoforms, representing differential functional roles in the ancient gene (prior to the duplication), can be 'fixed', respectively, in each of the duplicate copies. Hughes (1994) coined the term 'functional-sharing' to describe this type of subfunctionalization between duplicate genes.

The functional-sharing hypothesis can explain why genes belonging to large families tend to have less alternative splice (AS) forms, because gene duplication (GD) and AS mechanisms are two alternative ways to generate protein function diversifications during the course of evolution. In a recent article, Roux and Robinson-Rechavi (2011) challenged this simple explanation. Instead, they argued that genes with low AS levels tend to duplicate more frequently, plus a progressive acquisition of new splice forms, can explain the genomic data better. In the following we call it the duplicability-age hypothesis for short.

In this paper, we first address the main critic of Roux and Robinson-Rechavi (2011) on the functional-sharing hypothesis, and then discuss mechanistically the meaning of 'duplicability' in their proposed hypothesis. We conclude that the analysis of Roux and Robinson-Rechavi (2011) did not nullify the basic idea of functional-sharing hypothesis, and their 'duplicability-age hypothesis' was based on some unjustified claims. As all these genome-wide correlation analyses in both sides of debates may face the same problem of multiple, ambiguous interpretations, we suggest a more clear-cut approach that analyzes the indel (deletions and insertions)–exon structure correspondence between duplicate genes that can resolve this controversy by experimentation.

2. Results and discussion

2.1. Low proportion of AS genes in single-copy genes: age matters?

Though it may be sensitive to the experimental noise and cutoff, the proportion (f_{AS}) of genes with more than one AS isoforms (AS > 1) has

Abbreviations: AS, alternative splicing; CNV, copy number variation; GD, gene duplication; EST, expressed sequence tag; f_{AS}, the proportion of genes with more than one alternative splicing isoforms; WGD, whole-genome duplication; SD, segmental duplication; BAI, brain-specific angiogenesis inhibitor.

* Corresponding author at: MOE Key Laboratory of Contemporary Anthropology and Center for Evolutionary Biology, School of Life Sciences, Fudan University, Shanghai 200433, China. Tel./fax: +86 21 65642507.
E-mail address: xgu@iastate.edu (X. Gu).

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2012.05.012

been frequently used as a measure for the abundance of AS isoforms at the genome level (Jin et al., 2008; Kopelman et al., 2005; Roux and Robinson-Rechavi, 2011; Su et al., 2006). In these studies, f_{AS} was calculated for gene groups with the same number of paralogous genes (size-k family, $k=1, 2,\ldots$). Except for the group of single-copy genes (size-1 family) that f_{AS} is slightly lower than that of size-2 family, an inverse relationship between f_{AS} and the gene family size k was generally observed. The difference of Roux and Robinson-Rechavi (2011) from other studies was that they considered the low f_{AS} of single-copy genes in the regression analysis by using a parabola (quadratic) curve. The result became the main argument to criticize the functional-sharing hypothesis that cannot explain the low f_{AS} of single-copy genes. Instead, they explained it was the result of a progressive acquisition of new splice forms and single-copy genes are, on average, younger than duplicate genes, i.e., the gene-age matters.

If it is the case, one may expect that in those single-copy genes that are alternatively spliced, the AS level should also be relatively low. We thus did a simple calculation. For each set of genes with similar family size, the mean (m^*) number of AS forms over AS genes, i.e., after excluding genes with AS = 1, can be computed through the following equation

$$m = (1-f_{AS}) \times 1 + f_{AS} \times m*$$

where f_{AS} and m (the mean number of AS isoforms, including genes with AS = 1) were from the dataset of Roux and Robinson-Rechavi (2011). Strikingly, for single-copy genes we obtained $m^* \approx 6.54$, almost the same as $m^* \approx 6.57$ in family size-2. As shown in Fig. 1, m^* remains virtually the same between single-copy genes and genes with small family size, while it decreases remarkably when the gene family size becomes large. This suggests that the inverse relationship between the AS level (m^*) and the gene family size is logistic rather than parabola-like, though genes with no alternative splicing (AS = 1) are indeed overpresented in the group of single-copy genes. This pattern is difficult to understand from the view of gene-age hypothesis; needless to say there are a variety of technical reasons that may account for this phenomenon (Brett et al., 2002; Modrek et al., 2001; Takeda et al., 2010; Tress et al., 2007; Yeo et al., 2005).

The proportion of genes with AS = 1 is not unbiased and sensitive to the technology, because genes with low AS level are difficult to be quantitatively detected. With the development of new high-throughput transcriptome sequencing technology, many studies have showed that almost all multiple-exon genes are actually alternative spliced. Hence, the mean AS from (excluding AS = 1) would be a better unbiased measurement to represent the AS level of investigated genes, especially for the current analyses combining different data types from different technical platforms. It should be noticed that our knowledge about the genome-wide AS isoforms has been changed dramatically in recent years by advance high-throughput sequencing technologies. For instance, the initial EST dataset showed that about 30% of human genes were alternatively spliced (Brett et al., 2000; Kan et al., 2001; Maniatis and Tasic, 2002). This number has been increased up to around 60% with the increasing coverage of EST database (Gupta et al., 2004; Kim et al., 2004; Kriventseva et al., 2003). Astonishingly, deep next-generation sequencing has shown that the proportion of human AS genes could be as high as >90%, implying that almost all genes could be potentially alternatively spliced (Fujita et al., 2011; Hartmann and Valcarcel, 2009; Nilsen and Graveley, 2011). These findings of saturated genome-wide AS levels may cast some doubts on the gene-age hypothesis of Roux and Robinson-Rechavi (2011).

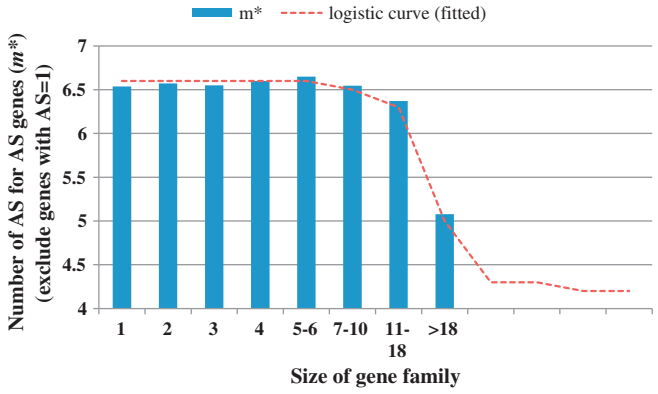

Fig. 1. The mean number of AS isoforms per AS genes (excluding AS = 1) plotted against the size of gene family.

In addition, there are at least two mechanisms that the extant single-copy genes have been generated through genome evolution. First, the genes de novo originated from previous non-genic genomic regions. Second, the current singletons actually come from gene duplication, but after a long history of evolution, we cannot find its parental genes based on sequence similarity search. Since the de novo new genes usually are single-exon genes (Wu et al., 2011) so that their AS = 1 always, we feel it is not comparable to the duplicates. To avoid this potential bias, we used the mean AS level after excluding AS = 1 in both single-copy and duplicate genes.

2.2. Deciphering the meaning of gene duplicability

Roux and Robinson-Rechavi (2011) explained the decrease of f_{AS} in large gene families by assuming that genes with low AS levels tend to duplicate more frequently, i.e., low AS means high duplicability, and *vice versa*. Yet, invoking the concept of 'duplicability' may cause some confusion, as we discuss below.

At the level of population genetics, any duplication event must occur in one or very few individuals. Most of these duplication events were lost from the population, either by deleterious dosage effects or random drifts. In some cases, it can gradually spread over the whole population due to positive selection or subfunctionalization. In this sense, the evolutionary rate (v) of any duplication event can be symbolically written as follows

$$v = d \times f$$

where d is the mutation rate leading to any duplication event, and f is the probability of a duplicate gene being preserved in the population (Force et al., 1999). There is some confusion in the literature for which one, d, f or v, should be called 'duplicability', which is relevant to the biological interpretation of Roux and Robinson-Rechavi (2011). That is, genes with low AS levels tend to duplicate more frequently, because of (*i*) high mutation rate (d-mode), (*ii*) high probability to be preserved (f-mode), or (*iii*) the joint-effect of the above (v-mode). In the following discussion we clarify this issue by assigning the d-mode for duplicability, and the f-mode for duplicate preservation.

There are two major duplication types. The first one is the well-known whole-genome duplication (WGD). Since all genes are duplicated simultaneously, 'duplicability' (d) is the same for all genes. Whether those duplicated genes can be preserved depend on a variety of factors that may influence f (Force et al., 1999). The second duplication type is the segmental duplication (SD) of a genomic region. In this case, different duplicability (d) becomes an important factor, because different chromosome regions may have different mutation rates for gene duplications.

The study of segmental duplications (SDs) has been advanced considerably by the discovery of copy number variations (CNVs) that are pervasive in virtually all eukaryotes. Simply to say, CNV and SD are the two facets of the same genomic dynamics at the population level and at the species level, respectively. Since SD is the CNV that has been fixed in the population, duplicability (d) is virtually the

mutation rate of CNV. A number of studies have shown that the mutation rate of CNV varies dramatically in different chromosomes, and identified many mutational hotspots of CNVs (Zhang et al., 2009). For instance, Fu et al. (2010) conducted a genome-wide population genetic analysis and estimated that the mutation rate of CNVs varies up to 10^3-folds. It means that, for segmental duplications, gene duplicability can have as many as 1000-fold differences.

Hence, we point out that the duplicability hypothesis of Roux and Robinson-Rechavi (2011) may have multiple meanings. In the case of genome-wide duplication, it means that duplicate genes with lower AS levels tend to be preserved more frequently. In the case of segmental duplications, it could mean that genes with lower AS levels either are located in mutational hotspots more frequently or tend to be preserved more frequently. To our best knowledge, solid empirical evidence has not been available so far to support these claims.

2.3. Reduction of genome-wide AS levels in duplicates insensitive to underlying mechanisms

It should be noticed that this debate (Roux and Robinson-Rechavi, 2011 vs. Jin et al., 2008; Kopelman et al., 2005; Su et al., 2006; Talavera et al., 2007) implied that the functional-sharing mechanism should lead to a significant reduction of the AS level in duplicate genes that can be statistically detected by the genome-wide analysis. It is true if two conditions are satisfied: (*i*) the number of AS isoforms at a gene should be far away from the saturation level; and (*ii*) the proportion of functional AS isoforms should be sufficiently large. As indicated above, it has been shown that the proportion of human AS genes could be saturated (Fujita et al., 2011; Hartmann and Valcarcel, 2009; Nilsen and Graveley, 2011), suggesting that the first condition (unsaturated AS level) may not hold. Moreover, this unexpected high AS levels have casted some serious doubts for whether these AS forms are all functional (Copley, 2004; Venables, 2002; Xing and Lee, 2004). Indeed, current wisdom is that majority of detected AS isoforms actually are nonfunctional (Fujita et al., 2011; Hartmann and Valcarcel, 2009; Nilsen and Graveley, 2011). They are simply from the splicing errors and will be removed thereafter. Since only a very few number of AS isoforms may have nontrivial contributions to the protein diversification or genome regulation, the reduced level of AS isoforms can only be statistically detected in some cases such as large gene families.

2.4. Testing the functional-sharing hypothesis of AS isoforms between duplicate genes

The question now becomes how we can test the functional-sharing hypothesis straightforwardly. Previous studies have shown that the process of partitioning of parental genes' AS forming into daughter genes may occur shortly after gene duplication (Su et al., 2006). After that, novel alternative splicing forms may be generated. Intuitively, the most straightforward method to test function-sharing model is to compare the AS form in the most young duplicate genes with its ancestral genes. Indeed, a very recent study reported that the reducing of AS degree is a general pattern in new genes of fruit fly (Zhan et al., 2010). However, it is technically difficult to detect the AS isoform change in very recently duplicated genes, due to the following reasons: 1) If many of new duplicate genes are actually pseudogenes or processed pseudogenes, the detected AS forms are nonfunctional. 2) New duplicate genes usually have very low gene expression level and high tissue specificity, resulting in low level of transcripts for AS detection. 3) The high sequence identity between recent duplicated genes may confound the mapping of the ESTs or RNA-seq short reads back to the duplicated genes.

Promoted by several interesting case studies in plants and animals (Altschmied et al., 2002; Cusack and Wolfe, 2007; Irimia et al., 2010; Lister et al., 2001; Rosti and Denyer, 2007; Yu et al., 2003; Yuan et al., 2009; Zhou et al., 2011), we propose the following procedure to resolve this problem: Consider two duplicate genes, each of which has several orthologous genes. The functional-sharing hypothesis for AS between duplicate genes predicts a correspondence between the indels (deletions and insertions) of protein sequences and the exon–intron boundaries. With the help of substantial internet protein family databases and bioinformatics tools, this approach is practically feasible. Further, these results can be considered as computational predictions of protein function diversifications between duplicate genes, based on which one can design some experimental assays for verification.

We have conducted some preliminary analyses in vertebrate gene families. For instance, the BAI (brain-specific angiogenesis inhibitor) gene family has three paralogous genes (BAI1, BAI2 and BAI3), which likely emerged during the early stage of vertebrates. We examined the transcripts of these genes in several mammals and vertebrates in the UCSC genome browser and ENSEMBL database, and identified the reference transcript as major alternative splicing events. The molecular phylogeny of BAI genes is shown in Fig. 2. Since the gene structure and AS events are highly conserved in tetrapods, to be concise we only show the BAI gene structure and AS in human and two teleost fishes, fugu and stickleback. We found that, though BAI1, BAI2 and BAI3 are highly similar in terms of the number and order of exon as well as their lengths and precise boundaries, each gene has its own specific exons or splicing forms, showing the different exon inclusion levels in three BAI genes after gene duplication (Fig. 3): (*i*) The exon orthologous to human BAI1 exon 3 is specific for BAI1 genes; (*ii*) BAI2 exons orthologous to human BAI2 exon25 are alternatively spliced and included in the major isoforms in most vertebrates, and (*iii*) the counterparts of BAI3 are not spliced in majority of the investigated vertebrates except for fugu, and the corresponding exons are lost in most BAI1 genes except in fugu BAI1.1. Note that fugu BAI genes have evolved into more complicated alternative splicing forms such as in fugu BAI2, BAI3 and one copy of late duplicated BAI1 (BAI1.1_Fugu), (Fig. 3). It would be interesting to experimentally identify the functional consequence of these events at the protein sequence level.

Because ancient AS isoforms (prior to the gene duplication) are usually unknown, the determination of exon gain or loss event along a gene family phylogeny may be technically complicated. Nevertheless, the phylogeny of the BAI genes in vertebrates can provide some clues that help in the inference of the ancestral AS state of a specific exon. For instance, based on the phylogeny, gene structure and AS mapping information, we can reasonably infer the existence of the orthologous exon of human BAI2 exon 25 in ancestral gene and it is alternatively spliced. After two rounds of gene duplications, the daughter genes showed different exon inclusion patterns. Moreover, we used the amphioxus BAI orthologous gene as an outgroup to identify the ancestral states of

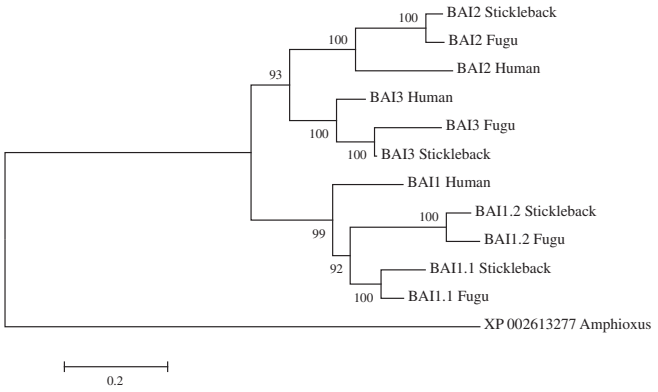

Fig. 2. Phylogeny of BAI gene family based on the longest protein isoform of each gene. Phylogenetic trees were constructed by neighbor-joining (NJ) method with the Poisson correction model using MEGA 5.0 (Tamura et al., 2011). Bootstrap values (percentage of 1000 replicates) are given at the branch nodes. Bootstrap values are indicated near the nodes.

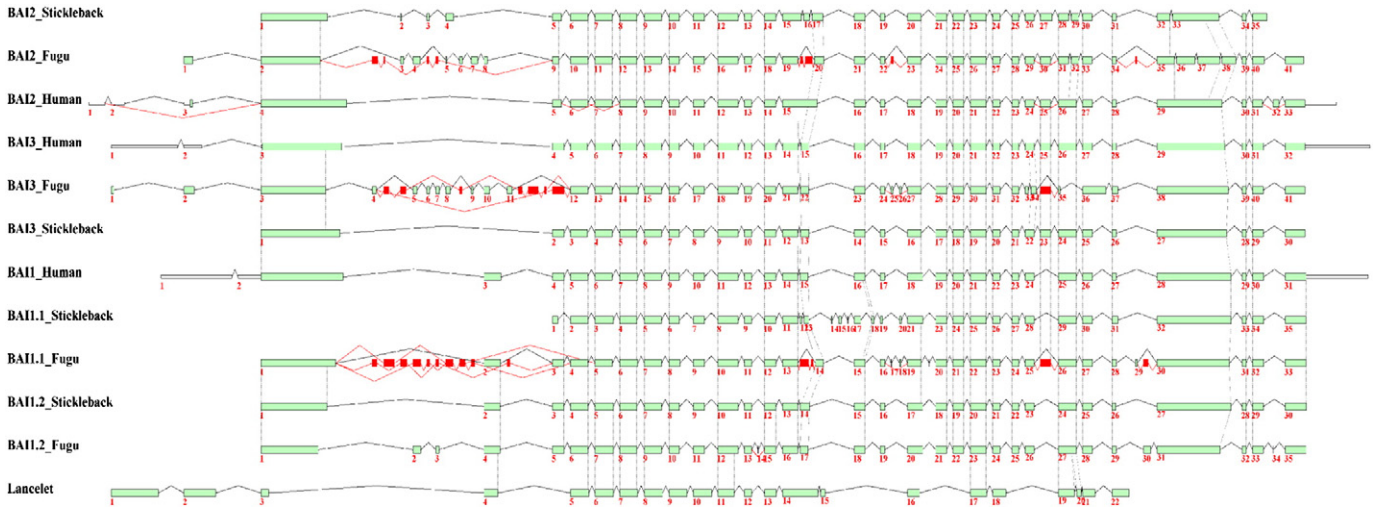

Fig. 3. Gene structure and alternative splicing of the BAI gene family members (BAI1, BAI2 and BAI3) in human, stickleback, fugu and lancelet. Exons are shown as boxes; filled boxes represent coding region; red boxes show the exons exclusively be find in minor transcripts; splices observed in transcript sequences are shown as lines. Splices observed in the major transcript form are shown as black lines; those observed in the minor form are shown as red lines. The alternative transcription start (stop) sites are not considered here. Introns are not drawn to scale. Homologous exons are indicated with dashed lines.

these exons in the vertebrate BAI gene family. Due to the low sequence and gene structure similarity between lancelet BAI ortholog and vertebrate BAI genes, we only identified a few homologous exons between them. It seems that the exon turnover after the gene duplication was prevailing. Recently, Lin et al. (2008) suggested several mechanisms for the origin of new exons during the course of evolution, including the Alu-mediated transposition, exon duplication, or intron-taken mechanism. One may speculate that these mechanisms can be also applied in the case of gene duplication.

3. Conclusive remarks

We have made three arguments in response to the critics of Roux and Robinson-Rechavi (2011): First, genome-wide pattern analyses of AS isoforms actually did not provide any clear-cut evidence to nullify the basic idea of functional-sharing hypothesis. Second, in spite that Roux and Robinson-Rechavi (2011) conducted substantial genomic analyses and claimed to favor their proposed duplicability-age hypothesis, we pointed out some fundamental flaws. Third, we show that genome-wide analyses of AS isoforms in duplicates are not capable of resolving the current debate.

We further outline a direct correspondence analysis of indels between duplicate genes to the exon–intron structure. Preliminary results indicate that this approach is promising. While loss of existed alternative exons could be explained by the hypothesis of functional-sharing, origin of new exons after the gene duplication may provide an opportunity for a protein to acquire some new functions right after gene duplication. We emphasize that the indel–exon structure correspondence analysis provides an approach to testing the functional-sharing hypothesis by experimentation, whereas in principle we are not sure how to formulate an experimental procedure for testing the duplicability-age hypothesis.

Acknowledgments

We are grateful to Jun Yu and Xiaoqiu Huang for comments in the early version of this manuscript, and Wei Huang for assistance. This work was supported by grant from the Ministry of Science and Technology China (2012CB910101 to Z.S.).

References

Altschmied, J., et al., 2002. Subfunctionalization of duplicate mitf genes associated with differential degeneration of alternative exons in fish. Genetics 161, 259–267.
Barbazuk, W.B., Fu, Y., McGinnis, K.M., 2008. Genome-wide analyses of alternative splicing in plants: opportunities and challenges. Genome Res. 18, 1381–1392.
Birzele, F., Csaba, G., Zimmer, R., 2008. Alternative splicing and protein structure evolution. Nucleic Acids Res. 36, 550–558.
Brett, D., et al., 2000. EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. FEBS Lett. 474, 83–86.
Brett, D., Pospisil, H., Valcarcel, J., Reich, J., Bork, P., 2002. Alternative splicing and genome complexity. Nat. Genet. 30, 29–30.
Chothia, C., Gough, J., Vogel, C., Teichmann, S.A., 2003. Evolution of the protein repertoire. Science 300, 1701–1703.
Copley, R.R., 2004. Evolutionary convergence of alternative splicing in ion channels. Trends Genet. 20, 171–176.
Cusack, B.P., Wolfe, K.H., 2007. When gene marriages don't work out: divorce by subfunctionalization. Trends Genet. 23, 270–272.
Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J., 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151, 1531–1545.
Fu, W., Zhang, F., Wang, Y., Gu, X., Jin, L., 2010. Identification of copy number variation hotspots in human populations. Am. J. Hum. Genet. 87, 494–504.
Fujita, P.A., et al., 2011. The UCSC Genome Browser database: update 2011. Nucleic Acids Res. 39, D876–D882.
Graveley, B.R., 2001. Alternative splicing: increasing diversity in the proteomic world. Trends Genet. 17, 100–107.
Gu, X., Wang, Y., Gu, J., 2002. Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. Nat. Genet. 31, 205–209.
Gupta, S., Zink, D., Korn, B., Vingron, M., Haas, S.A., 2004. Genome wide identification and classification of alternative splicing based on EST data. Bioinformatics 20, 2579–2585.
Hartmann, B., Valcarcel, J., 2009. Decrypting the genome's alternative messages. Curr. Opin. Cell Biol. 21, 377–386.
Hughes, A.L., 1994. The evolution of functionally novel proteins after gene duplication. Proc. Biol. Sci. 256, 119–124.
Irimia, M., Rukov, J.L., Penny, D., Garcia-Fernandez, J., Vinther, J., Roy, S.W., 2008. Widespread evolutionary conservation of alternatively spliced exons in *Caenorhabditis*. Mol. Biol. Evol. 25, 375–382.
Irimia, M., Maeso, I., Gunning, P.W., Garcia-Fernandez, J., Roy, S.W., 2010. Internal and external paralogy in the evolution of tropomyosin genes in metazoans. Mol. Biol. Evol. 27, 1504–1517.
Jin, L., et al., 2008. The evolutionary relationship between gene duplication and alternative splicing. Gene 427, 19–31.
Kan, Z., Rouchka, E.C., Gish, W.R., States, D.J., 2001. Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. Genome Res. 11, 889–900.
Keren, H., Lev-Maor, G., Ast, G., 2010. Alternative splicing and evolution: diversification, exon definition and function. Nat. Rev. Genet. 11, 345–355.
Kim, H., Klein, R., Majewski, J., Ott, J., 2004. Estimating rates of alternative splicing in mammals and invertebrates. Nat. Genet. 36, 915–916 (author reply 916–7).
Koonin, E.V., Wolf, Y.I., 2010. Constraints and plasticity in genome and molecular-phenome evolution. Nat. Rev. Genet. 11, 487–498.

Kopelman, N.M., Lancet, D., Yanai, I., 2005. Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. Nat. Genet. 37, 588–589.
Kriventseva, E.V., et al., 2003. Increase of functional diversity by alternative splicing. Trends Genet. 19, 124–128.
Lareau, L.F., Green, R.E., Bhatnagar, R.S., Brenner, S.E., 2004. The evolving roles of alternative splicing. Curr. Opin. Struct. Biol. 14, 273–282.
Lin, L., et al., 2008. Diverse splicing patterns of exonized Alu elements in human tissues. PLoS Genet. 4, e1000225.
Lister, J.A., Close, J., Raible, D.W., 2001. Duplicate mitf genes in zebrafish: complementary expression and conservation of melanogenic potential. Dev. Biol. 237, 333–344.
Lynch, M., Conery, J.S., 2000. The evolutionary fate and consequences of duplicate genes. Science 290, 1151–1155.
Maniatis, T., Tasic, B., 2002. Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature 418, 236–243.
Modrek, B., Resch, A., Grasso, C., Lee, C., 2001. Genome-wide detection of alternative splicing in expressed sequences of human genes. Nucleic Acids Res. 29, 2850–2859.
Nilsen, T.W., Graveley, B.R., 2011. Expansion of the eukaryotic proteome by alternative splicing. Nature 463, 457–463.
Rosti, S., Denyer, K., 2007. Two paralogous genes encoding small subunits of ADP-glucose pyrophosphorylase in maize, Bt2 and L2, replace the single alternatively spliced gene found in other cereal species. J. Mol. Evol. 65, 316–327.
Roux, J., Robinson-Rechavi, M., 2011. Age-dependent gain of alternative splice forms and biased duplication explain the relation between splicing and duplication. Genome Res. 21, 357–363.
Shabalina, S.A., Ogurtsov, A.Y., Spiridonov, A.N., Novichkov, P.S., Spiridonov, N.A., Koonin, E.V., 2010. Distinct patterns of expression and evolution of intronless and intron-containing mammalian genes. Mol. Biol. Evol. 27, 1745–1749.
Su, Z., Wang, J., Yu, J., Huang, X., Gu, X., 2006. Evolution of alternative splicing after gene duplication. Genome Res. 16, 182–189.
Takeda, J., et al., 2010. H-DBAS: human-transcriptome database for alternative splicing: update 2010. Nucleic Acids Res. 38, D86–D90.
Talavera, D., Vogel, C., Orozco, M., Teichmann, S.A., de la Cruz, X., 2007. The (in)dependence of alternative splicing and gene duplication. PLoS Comput. Biol. 3, e33.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Tress, M.L., et al., 2007. The implications of alternative splicing in the ENCODE protein complement. Proc. Natl. Acad. Sci. U. S. A. 104, 5495–5500.
Venables, J.P., 2002. Alternative splicing in the testes. Curr. Opin. Genet. Dev. 12, 615–619.
Wang, E.T., et al., 2008. Alternative isoform regulation in human tissue transcriptomes. Nature 456, 470–476.
Wu, D.D., Irwin, D.M., Zhang, Y.P., 2011. De novo origin of human protein-coding genes. PLoS Genet. 7, e1002379.
Xing, Y., Lee, C.J., 2004. Negative selection pressure against premature protein truncation is reduced by alternative splicing and diploidy. Trends Genet. 20, 472–475.
Yeo, G.W., Van Nostrand, E., Holste, D., Poggio, T., Burge, C.B., 2005. Identification and analysis of alternative splicing events conserved in human and mouse. Proc. Natl. Acad. Sci. U. S. A. 102, 2850–2855.
Yu, W.P., Brenner, S., Venkatesh, B., 2003. Duplication, degeneration and subfunctionalization of the nested synapsin-Timp genes in Fugu. Trends Genet. 19, 180–183.
Yuan, Y., et al., 2009. Alternative splicing and gene duplication differentially shaped the regulation of isochorismate synthase in Populus and Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 106, 22020–22025.
Zhan, Z., Ren, J., Zhang, Y., Zhao, R., Yang, S., Wang, W., 2010. Evolution of alternative splicing in newly evolved genes of *Drosophila*. Gene 470, 1–6.
Zhang, F., Gu, W., Hurles, M.E., Lupski, J.R., 2009. Copy number variation in human health, disease, and evolution. Annu. Rev. Genomics Hum. Genet. 10, 451–481.
Zhang, P.G., Huang, S.Z., Pin, A.L., Adams, K.L., 2010. Extensive divergence in alternative splicing patterns after gene and genome duplication during the evolutionary history of Arabidopsis. Mol. Biol. Evol. 27, 1686–1697.
Zhou, R., Moshgabadi, N., Adams, K.L., 2011. Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. Proc. Natl. Acad. Sci. U. S. A. 108, 16122–16127.