2006

Evolution of Alternative Splicing after Gene Duplication

Zhixi Su
Zhejiang University

Jianmin Wang
Zhejiang University

Xiaoqiu Huang
Iowa State University, xqhuang@iastate.edu

Xun Gu
Zhejiang University

Follow this and additional works at: http://lib.dr.iastate.edu/cs_pubs
Part of the Bioinformatics Commons, Computer Sciences Commons, and the Genomics Commons

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/cs_pubs/7. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.
Evolution of alternative splicing after gene duplication

Zhixi Su,1,2,6 Jianmin Wang,3,6 Jun Yu,1,5 Xiaoqiu Huang,3,4 and Xun Gu1,2,4,7

1James D. Watson Institute of Genome Sciences, Zhejiang University, Hangzhou 310008, China; 2Department of Genetics, Development, and Cell Biology, 3Department of Computer Science, and 4Center for Bioinformatics and Biological Statistics, Iowa State University, Ames, Iowa 50011, USA; 5Beijing Genomics Institute, Chinese Academy of Sciences, Beijing 101300, China

Alternative splicing and gene duplication are two major sources of proteomic function diversity. Here, we study the evolutionary trend of alternative splicing after gene duplication by analyzing the alternative splicing differences between duplicate genes. We observed that duplicate genes have fewer alternative splice (AS) forms than single-copy genes, and that a negative correlation exists between the mean number of AS forms and the gene family size. Interestingly, we found that the loss of alternative splicing in duplicate genes may occur shortly after the gene duplication. These results support the subfunctionization model of alternative splicing in the early stage after gene duplication. Further analysis of the alternative splicing distribution in human duplicate pairs showed the asymmetric evolution of alternative splicing after gene duplications; i.e., the AS forms between duplicates may differ dramatically. We therefore conclude that alternative splicing and gene duplication may not evolve independently. In the early stage after gene duplication, young duplicates may take over a certain amount of protein function diversity that previously was carried out by the alternative splicing mechanism. In the late stage, the gain and loss of alternative splicing seem to be independent between duplicates.

[Supplemental material is available online at www.genome.org.]

Results

Detecting human alternative splice forms

Our interest here is to identify all potential AS forms for well annotated human genes. To this end, we developed a computational pipeline to predict the number of AS isoforms for each human gene, which was used as the proxy of the alternative splicing functional divergence. We then conducted a comprehensive analysis to investigate the evolutionary pattern of alternative splicing after gene duplications.
Yu et al. (2003) conducted a few case studies to support the function-sharing model. To further test whether it is the general case, we used a BLAST search to classify all well annotated human genes under study into 8819 single-copy genes and 6603 duplicate genes (see Methods). The percentage of alternative splicing of single-copy genes is 7090/8819 = 80.4%, while that of duplication genes is given by 4924/6603 = 74.6%. The difference between single-copy genes and duplicates is statistically highly significant ($\chi^2 = 74.3$, $P < 10^{-7}$). Similarly, the mean number of alternative splicing duplicate genes is 3.52 ± 0.05, which is significantly lower than that of single-copy genes (4.11 ± 0.05 [t-test, $P < 0.001$]). Besides, we found a higher proportion of single-copy genes with many AS forms (8.3% for >10 forms) than duplicates (6.6%). In short, our analysis suggests that, at the genome level, functional divergence among duplicate genes may reduce the number of AS forms, which is more consistent with the prediction of the function-sharing model than the independent model. Roughly speaking, one duplicate gene may take over a certain amount of proteomic diversity that previously was carried out by, on average, 0.59 (≈4.11 – 3.52) AS forms. In other words, in the human genome, there are ~6603 × 0.59 = 3896 losses of AS forms due to gene duplications.

**Large gene families tend to have fewer alternative splice forms**

Another prediction from the function-sharing model is that large gene families tend to have fewer AS forms, because multiple rounds of gene duplication may result in more loss of AS forms. Except for 1486 genes that cannot be mapped to Ensembl gene ID or were not assigned an Ensembl gene family ID, we further classified the remaining 13,936 human genes into 8211 gene families (see Methods). For each gene family size, we estimated the mean of AS forms and the percentage of genes that are not alternatively spliced (no-AS). Figure 3A shows the mean of AS forms plotted against the gene family size; a single-copy gene is regarded as a gene family with size 1. It appears that the mean of AS forms remains roughly constant for gene families with small to moderate sizes, e.g., size 1 to 4, but decreases for larger gene families ($R = -0.85$, $P < 0.0037$). Similarly, Figure 3B shows that the proportion of no-AS forms increases with increasing gene family size ($R = 0.93$, $P < 0.0004$). In short, the negative correlation between the number of AS forms and the gene family size is likely to be caused by the effect of function-sharing after gene duplications.

**Loss of alternative splicing may occur only shortly after gene duplication**

To further explore the evolutionary pattern of alternative splicing after gene duplications, we compiled independent 2875 hu-
man duplicate pairs and calculated the protein sequence distance (d) between duplicates (see Methods). As shown by a histogram (Supplemental Fig. 1), there is a peak of duplicate pairs around d = 0.7, roughly corresponding to 500–600 million years ago when the duplication time was estimated by the method of molecular clock (Gu et al. 2002b).

Using the protein sequence distance as a proxy of the age of the duplicate, we first grouped duplicate gene pairs with similar protein sequence distance (with a bin of 0.1 distance unit, ∼80 million years ago, or the time of mammalian radiation), and calculated the percentage of no-AS. Interestingly, the percentage of no-AS in the most recent duplicate group (d < 0.1) is 42%, which is almost two times higher than that in the other more ancient groups (22% on average, Fig. 4A); a $\chi^2$ test showed the difference is statistically significant ($P < 0.001$). Similarly, the mean number of AS forms (2.7) in the most recent duplicate genes is smaller than that of more ancient duplicate genes (3.5; Fig. 4B). Hence, both measures, the percentage of no-AS and the mean number of AS forms, indicated a rapid reduction of AS forms in young duplicates.

**Figure 3.** Fewer AS forms in larger gene families. Error bar, standard error. (A) Mean number of AS forms plotted against human gene family size (size = 1 means single-copy gene). (B) Proportion of genes that are not alternatively spliced plotted against the human gene family size.

**Figure 4.** The proportion of no alternatively spliced (no-AS) genes (A) and the mean of AS forms (B) plotted against the protein distance between duplicates (with a bin of 0.1 distance unit). Error bar, standard error.

Duplication versus alternative splicing in other model organisms

The above analysis was based on human AS isoforms. It would be interesting to test whether duplicates in other organisms also tend to have a lower number of AS isoforms. We developed a simple approach to compare the evolutionary patterns of alternative splicing in several model animals, and found that it might be the case (Fig. 5). First, we used the amino acid identity (I) and the BLAST search E-value as the criteria to define duplicate genes. For example, the identity $I = 30$ and E-value $1E-10$ mean genes have identity $\geq 30\%$ and E-value is $\leq 1E-10$. Second, we changed the criterion $I$ from 30% to 50%, 70%, and 90%, respectively, and calculated the corresponding proportion of no-AS genes. Since the amino acid identity (I) is a proxy for young and ancient duplicate genes, the proportion of no-AS genes is expected to increase with the increasing of I, if the loss of AS isoforms in duplicates may occur shortly after the gene duplication. For comparison, we also included the human genome. As shown in Figure 5, we found that in all organisms we examined (human, mouse, *Drosophila*, and *Caenorhabditis elegans*), the proportion of no-AS genes tends to increase with the sequence identity criterion. Our analysis indicated that the loss of AS forms in young duplicates may be a general pattern in animals.

**Alternative splicing of duplicates after the human-mouse split**

To further test the hypothesis that loss of alternative splicing may occur only shortly after gene duplication, we classified all human genes into two groups using mouse orthologs as the reference. The A group (H1) includes 9640 one-to-one and 282 one-to-many human-mouse orthologous genes, while the B group (Hx) includes 419 human duplicate genes that were duplicated after the human–mouse split. We analyzed the alternative splicing evolution for the B group, i.e., human recent duplicates after the human–mouse split. We found that the mean AS forms of the Hx group (2.83 ± 0.2) is significantly lower than that of the H1 group (3.66 ± 0.04), as well as that of whole human duplicate genes (3.52 ± 0.05). Similarly, the percentage of no-AS forms in Hx (0.369 ± 0.024) is significantly higher than in the H1 group (0.221 ± 0.0042). These results indicate that the reduction of AS forms in young duplicates is a general pattern in animals.
Evolution of alternative splicing

forms in duplicate genes mainly happened in the early stage after the gene duplication.

Put together, we propose a scenario for the evolution of alternative splicing after the gene duplication (Fig. 2). For simplicity, consider a gene that has two AS forms, \( L \) (long) and \( S \) (short), for distinct physiological roles. After the gene/genome duplication, two duplicate copies \( (A \) and \( B) \) inherit both AS forms, but may start to have differential expression profiles. Because of the functional redundancy, the transcription of the \( L \) form in gene \( A \) becomes more dominant, while that of the \( S \) form in gene \( B \) becomes dominant. This evolutionary transition from alternative splicing diversity to functional divergence of duplicates may occur in the early stage after gene duplication. In the late stage, novel alternative splicing may be added, increasing the overall number of AS forms.

Testing asymmetric evolution of alternative splicing after gene duplication

Although there are 571 (19.8%) and 575 (20%) duplicate pairs with no or only one difference, respectively (Fig. 6), we indeed observed a significant portion of duplicate pairs showing dramatic differences in their number of AS forms. For instance, 206 duplicate pairs (7.2%) were found to have >10 differences of AS forms, indicating the possibility of an asymmetric pattern of alternative splicing evolution after gene duplication. It implies that not all duplicate genes have a similar number of AS forms. To address this issue rigorously, we used the binomial test to obtain the \( P \)-value (type-one error) for each duplicate pair under the null hypothesis of no difference in the number of AS forms. As a result, 418 pairs (14.5%) have significantly different AS forms \( (P < 0.05) \), and 181 pairs (6.3%) have \( P \)-values <0.01 (Fig. 7). Table 1 shows the 36 duplicate pairs that have the largest AS form difference \((P < 0.0001)\). For all 181 pairs, see Supplementary Table.) However, since it involves 2875 simultaneous statistical tests, the multiple-test problem should not be neglected. For instance, at the significance level of 0.05, there are \(-2875 \times 0.05 = 144\) significant cases by pure chance. In other words, at this 0.05 significance level, we observed 418 significant cases; only \( 418 - 144 = 274 \) cases are likely to be truly significant, whereas others are false-positive. Statistically, it may be evaluated by the false-positive discovery rate (FDR). At the 0.05 significance level, we calculated FDR = 34%. For 181 significant cases at the level of 0.01, the FDR is \(-16\%). At any rate, this analysis indicates that a significant portion of human duplicate genes evolve asymmetrically in the AS forms.

The effects of EST coverage and expression level

Clearly, AS form detecting is affected by EST coverage, because the more ESTs found for a given gene, the more likely AS forms(s) can be detected (Hide et al. 2001; Kan et al. 2002). When the size of the library is sufficiently large, the EST coverage of genes is mainly determined by the gene expression levels. Indeed, we observed a positive correlation between the detected AS form number and the EST hit number in our data set \((P < 0.0001, \text{Supplemental Fig. 2A})\). In addition, the number of genes decays exponentially with the number of AS forms (Fig. 1), as well as the number of EST hits \((R = 0.93, \text{Supplemental Fig. 2B})\).

We have run several tests to examine whether our duplicate AS analysis is affected by the EST coverage, measured by the number of ESTs aligned to a given gene (EST hits; see Methods). We found a very similar EST hit distribution between single-copy and duplicate genes; the mean number of EST hits in single-copy genes is 196 ± 3.7, with no significant difference from that in duplicate genes (206.9 ± 6.0). Moreover, we grouped duplicate genes and single-copy genes with similar EST coverage (with a bin of 40 EST hits) and found duplicate genes always have a larger percentage of no-AS genes and a smaller mean of AS forms in any EST coverage span (Supplemental Fig. 3).

In the AS–gene family size and AS–duplicate distance analyses, we classified all human genes into two categories: the \( H \) category for 7713 highly expressed genes with \( >112 \) EST hits, and the \( L \) category for 7709 other weakly expressed genes \((<112 \) EST hits). We found a positive correlation between the proportion of no-AS genes and gene family size in both \( H \) and \( L \) categories (Supplemental Fig. 4). For the \( L \) category, the proportion of no-AS decreases with increasing duplicate protein distance. However, in the \( H \) category, the proportion of no-AS in single-copy genes and duplicate genes decreases with increasing of the duplicate protein distance when \( d < 0.3 \), but seems to increase in more ancient duplicate pairs (Supplemental Fig. 4). Though this pattern needs to be elaborated further, it may imply that highly expressed duplicate genes may lose some AS forms in the late stage after gene duplication.

Finally, we compared the duplicate pairs that have significant asymmetric evolution of AS forms (Table 2). For 181 dupli-
We conducted a large-scale alternative splicing analysis of the human genome, using human AS forms predicted from the EST databases. Many methods or programs for identification of alternative splicing have been developed recently, such as mapping ESTs onto miRNA sequences (Brett et al. 2002); aligning ESTs with genomic sequences (Mironov et al. 1999); multiply aligning ESTs, mRNAs, and genomic sequences (Modrek et al. 2001; Modrek and Lee 2003); and aligning transcribed consensus sequences to genomic sequences (Gupta et al. 2004). A number of studies have relied on EST self-clustering to assemble alternatively spliced transcripts (Mironov et al. 1999; Quackenbush et al. 2001). Some studies have combined genome-based EST clustering and transcripts’ assembly approaches to reconstruct the alternatively spliced isoforms (Kan et al. 2001; Xing et al. 2004; Kim et al. 2005b). All these efforts attempt to address the “garbage EST” issue. The analysis pipeline we developed has adopted several techniques to reduce the potential garbage EST effects (also see Methods). First, a simple method was implemented to verify AS forms based on the pairwise alignments between EST/mRNA pairs with highly significant asymmetric AS forms (Su et al. 2005). Second, any EST that results in an extremely short exon or intron was removed because extremely short exons and introns are error-prone. Third, splice-site motifs (GT-AG/GC-AG) were used to validate our detected AS forms. And finally, to avoid the high false-positive rate of the transcript assembly procedure (Bouck et al. 2005a), we conducted a large-scale alternative splicing analysis of the human genome, using human AS forms predicted from the EST databases. Many methods or programs for identification of alternative splicing have been developed recently, such as mapping ESTs onto miRNA sequences (Brett et al. 2002); aligning ESTs with genomic sequences (Mironov et al. 1999); multiply aligning ESTs, mRNAs, and genomic sequences (Modrek et al. 2001; Modrek and Lee 2003); and aligning transcribed consensus sequences to genomic sequences (Gupta et al. 2004). A number of studies have relied on EST self-clustering to assemble alternatively spliced transcripts (Mironov et al. 1999; Quackenbush et al. 2001). Some studies have combined genome-based EST clustering and transcripts’ assembly approaches to reconstruct the alternatively spliced isoforms (Kan et al. 2001; Xing et al. 2004; Kim et al. 2005b). All these efforts attempt to address the “garbage EST” issue. The analysis pipeline we developed has adopted several techniques to reduce the potential garbage EST effects (also see Methods). First, a simple method was implemented to verify AS forms based on the pairwise alignments between EST/mRNA pairs with highly significant asymmetric AS forms (Su et al. 2005). Second, any EST that results in an extremely short exon or intron was removed because extremely short exons and introns are error-prone. Third, splice-site motifs (GT-AG/GC-AG) were used to validate our detected AS forms. And finally, to avoid the high false-positive rate of the transcript assembly procedure (Bouck et al. 2005a),
al. 1999; Xing et al. 2004), we did not assemble the ESTs to full-length transcripts. Instead, we excluded these garbage targets of nonsense-mediated decay (NMD) (Lewis et al. 2003; Lareau et al. 2004), such as aberrant splicing in some tumors, genome contamination, unspliced mRNA, splicing errors that arose from various reasons, or background splicing without any effect on the cellular function. Some of these artifacts can be detected by identifying the premature termination codon (PTC) in alternative splicing isoforms to find the putative targets of nonsense-mediated decay (NMD) (Lewis et al. 2003; Xing and Lee 2002; Lareau et al. 2004). We therefore conclude that AS evolution after gene duplication, which was calculated by the binomial test under the null hypothesis of no difference in the number of AS forms (see Methods). (A) represents all duplicate pairs; (B) represents duplicate pairs having P-value < 0.1.

Conclusions

In this study, we discovered that the level of alternative splicing in duplicate genes is usually lower than that of single-copy genes. Further analyses indicated that the number of AS forms is negatively associated with the gene family size, and the loss of AS forms may happen shortly after the gene duplication. Moreover, we demonstrated asymmetric AS evolution; i.e., the AS numbers between some duplicates may differ dramatically. Some of these results have been confirmed in other organisms, including mouse, Drosophila, and C. elegans. We therefore conclude that AS and gene duplication, two mechanisms for proteomic function diversity, may not evolve independently, supporting the model of function-sharing. That is, in the early stage after gene duplication, the young duplicates may take over a certain amount of protein function diversity that previously was carried out by the alternative splicing mechanism. After that, the evolution of AS forms may be independent between duplicates.

Several studies have shown that the evolutionary rate of coding sequences may accelerate shortly after gene duplication (Lynch and Force 2000; Kondrashov et al. 2002; Conant and Wagner 2003; Zhang et al. 2003). Gu et al. (2005) found that the initial rate for either expression or regulatory network evolution after yeast gene duplications is much higher than that of the late stage. The current study for the loss of AS forms between duplicates presents another example to support the viewpoint of rapid evolution in the early stage after gene duplication. Moreover, the pattern of rapid AS form loss in young duplicates is consistent with the model of subfunctionalization (Lister et al. 2001; Altschmied et al. 2002; Yu et al. 2003). We suggested that AS subfunctionalization of duplicate genes may be a general phenomenon that happened in the early stage after gene duplication. On the other hand, alternative splicing may contribute to the neofunctionalization by acquiring new functional AS variants, resulting in the increase of AS isoforms for ancient duplicate genes (Fig. 2).

Table 2. Summary of expression and alternative splicing divergence of duplicates

<table>
<thead>
<tr>
<th>Duplicate pair categories</th>
<th>Number of pairs</th>
<th>Proportion</th>
<th>Number of pairs</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P &lt; 0.01$ (181)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HH*</td>
<td>89</td>
<td>0.49</td>
<td>798</td>
<td>0.3</td>
</tr>
<tr>
<td>LL*</td>
<td>9</td>
<td>0.05</td>
<td>1018</td>
<td>0.37</td>
</tr>
<tr>
<td>LH*</td>
<td>83</td>
<td>0.46</td>
<td>878</td>
<td>0.33</td>
</tr>
</tbody>
</table>

| $P > 0.01$ (2694)         |                |            |                |            |

Shown here is the classification of the two groups of duplicate pairs that have different alternative splice form difference levels based on EST coverage.

*Both pairs have >112 EST hits.

Both pairs have ≤112 EST hits.

One gene of the duplicate pair has >112 EST hits and another one has ≤112 EST hits.
Methods

Sequence data

The reference sequences (RefSeq) of human, mouse, Drosophila, and C. elegans were downloaded from NCBI (ftp://ftp.ncbi.nih.gov/refseq/). Only the mRNA sequences of IDs starting with the prefix “NM_” were extracted, and the corresponding protein sequences were also extracted. There were 21,267, 16,863, 18,648, and 20,785 mRNA reference sequences for human, mouse, Drosophila, and C. elegans, respectively. The alignments between mRNA reference sequences and genome sequence were downloaded from the UCSC genome center (http://genome.ucsc.edu/), and they were generated by the BLAT program (Kent et al. 2002). The alignment information was used to remove redundant sequences. In the case of multiple mRNA sequences being mapped to one genomic region, only the sequence resulting in the longest protein product was accepted.

Identification of duplicate genes

We used the method of Gu et al. (2002a) to identify duplicate genes. Briefly, every protein was used as the query to search against all other proteins by using BLASTP ($E = 10$). Two proteins are scored as forming a link if (1) the BLASTP alignable region between them is $>80\%$ of the longer protein, and (2) the identity ($I$) between them is $\geq 30\%$ if the alignable region is longer than 150 amino acids; $I \geq 0.01n + 4.87 - 0.325\exp(\frac{-1}{1000}L - H)$. Duplicate pairs were seeded for all other protein pairs, in which $n$ is 6 and $L$ is the alignable length between the two proteins. Duplicate pairs were seeded with a two-way best pairwise match. The protein sequence distance ($d$) between duplicates was calculated by $d = -\ln(I/100)$, the Poisson correction.

Identifying gene family size and genes duplicated after the human–mouse split

All human gene family IDs were extracted using the EnsMart tool (Kasprzyk et al. 2004). The size of each gene family was obtained from counting the frequency of this gene’s family ID in whole genome. Orthologs and in-paralogs between human and mouse were extracted from the Inparanoid database (version 4.0) (O’Brien et al. 2005). Human genes having one-to-one and one-to-many relationships with mouse were classified into the H category. Similarly, human genes having many-to-one and many-to-many relationships with mouse were classified into the Hx category. RefSeq/Ensembl gene mapping was extracted using EnsMart. Only reciprocally unique pairs were further analyzed.

Identification of alternative splice forms

Alternative splice forms were identified by using alignments of human EST/RefSeq and genomic sequences as follows: The EST/RefSeq sequences that were highly similar to only one genomic region were selected. The remaining EST/RefSeq sequences were discarded. The selected EST/RefSeq sequences were grouped together by genomic regions, one group per genomic region. If there were multiple RefSeq sequences in a group, then the RefSeq sequence with the longest protein sequence was the leader sequence for the group. For each EST/RefSeq sequence, splice sites in the genomic region were identified by the GeneSplicer program (Pertea et al. 2001), and exons and introns of sufficient lengths in the genomic region were identified based on the splice sites and the alignment of the EST/RefSeq sequence and the genomic region. Any EST/RefSeq sequence without any exons and introns was discarded. The pattern of splicing for each EST/RefSeq sequence was indicated by a list of all exons in order of increasing coordinate. For each group of EST/RefSeq sequences, alternative splice forms were identified by comparing the exon list of each EST/RefSeq sequence with the exon list of the leader sequence and finding differences in exon coordinates. Redundant alternative splice forms were removed.

The method given above finds and removes garbage ESTs from aberrant transcripts or abnormal cell lines by three measures: First, any EST that is not highly similar to a genomic region over a majority of its length is removed. Second, any EST that results in an extremely short exon or intron is removed because extremely short exons and introns are error-prone. Third, any EST that results in a weak splice site in a genomic region is removed.

Additional methods for finding garbage ESTs have been suggested by several previous studies (Modrek and Lee 2002; Lareau et al. 2004; Xing and Lee 2004). One of them is to use only AS forms that are confirmed by two independent libraries. We are currently experimenting with this method and other methods to see if they affect the main observations in the paper.

Acknowledgments

This work was supported by NIH grants (X.G.).

References


188 Genome Research www.genome.org
Evolution of alternative splicing


Received May 27, 2005; accepted in revised form October 16, 2005.
Complex evolution of 7E olfactory receptor genes in segmental duplications
Tera Newman and Barbara J. Trask

Using a newer human genome assembly (May 2004, hg17), a better alignment and more stringent criteria, the authors find far fewer putative instances of ectopic exchanges involving the 7E subfamily of odorant receptor pseudogenes than originally reported (Newman and Trask 2003). Each of the three legs of support for this phenomenon is weakened. First, the shift in best-matching partners in Figure 5, and associated percent identities, is an artifact of the August 2001 genome assembly (hg 8). Second, the original alignment did not optimally position some frame-shifting mutations in these genes. An improved alignment of genes from the newest assembly (coordinates in Supplemental Table A; alignment in Supplemental Data) reveals that all but six of the 35 taxa in the B clade share a 4-nucleotide frameshift mutation in TM3 and a stop mutation (TTA→TGA) in TM6 (Supplemental Fig. A). The TM6 mutation is also shared by 24 of 46 taxa in clade A. Although the presence of the TM6 mutation in some but not all members of both clades is consistent with a past gene-conversion event, this pattern also could have resulted from two independent, identical mutations. Finally, use of GeneConv (Sawyer 1989; http://www.math.wustl.edu/~sawyer/geneconv/) on various subsets of 7E sequences using the updated and improved alignment implicates only a few genes in putative conversion events that are statistically significant after Bonferroni correction for multiple tests. Thus, while 7E-containing segmental duplications might be subject to ectopic exchange events (e.g., Giglio et al. 2001), statistically supported signatures of these events are not prevalent in the 7E pseudogene sequences. The authors regret any confusion these artifacts in their analysis might have caused, and thank Eleanor Williams for bringing them to their attention.

References


Evolution of alternative splicing after gene duplication
Zhixi Su, Jianmin Wang, Jun Yu, Xiaoqiu Huang, and Xun Gu

The authors inadvertantly failed to give precedence to the recent paper of N.M. Kopelman, D. Lancet, and I. Yanai (2005), “Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms,” Nat. Genet. 37: 588–589, which reports equivalent findings. Namely, that an inverse correlation exists between the size of a gene’s family and its use of alternatively spliced isoforms. The authors sincerely apologize to those concerned for this oversight.

References

Evolution of alternative splicing after gene duplication


Genome Res. 2006 16: 182-189
Access the most recent version at doi:10.1101/gr.4197006

Supplemental Material
http://genome.cshlp.org/content/suppl/2005/12/20/gr.4197006.DC1.html

Related Content
Correction for Volume 16, p. 182
Genome Res. April, 2006 16: 557 Age-dependent gain of alternative splice forms and biased duplication explain the relation between splicing and duplication
Julien Roux and Marc Robinson-Rechavi
Genome Res. March, 2011 21: 357-363

References
This article cites 44 articles, 24 of which can be accessed free at:
http://genome.cshlp.org/content/16/2/182.full.html#ref-list-1

Articles cited in:
http://genome.cshlp.org/content/16/2/182.full.html#related-urls

Creative Commons License
This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at http://creativecommons.org/licenses/by-nc/3.0/.

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

To subscribe to Genome Research go to:
http://genome.cshlp.org/subscriptions