Minireview

Intron mis-splicing: no alternative?

Scott William Roy* and Manuel Irimia†

Addresses: *National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20814, USA. †Departament de Genètica, Universitat de Barcelona, Barcelona 08028, Spain.

Correspondence: Scott William Roy. Email: scottwroy@gmail.com, Manuel Irimia. Email: mirimia@gmail.com

Published: 19 February 2008

*Genome **Biology*** 2008, **9:**208 (doi:10.1186/gb-2008-9-2-208)

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2008/9/2/208

© 2008 BioMed Central Ltd

Abstract

A recent report reveals widespread mis-splicing of RNA transcripts in eukaryotes, with mis-spliced RNA destroyed by nonsense-mediated mRNA decay. This striking inefficiency deepens the mystery of the proliferation and persistence of introns.

Coding sequences in eukaryotic genomes are frequently interrupted by spliceosomal introns, regions of noncoding DNA that are removed from pre-mRNA transcripts by the spliceosome, a complex of five RNAs and hundreds of proteins [1]. Why do introns exist? No general function for spliceosomal introns has been demonstrated, and both their absence from prokaryotes and the recurrent massive loss of introns in various eukaryotic lineages suggests that no such essential function may exist. One common hypothesis is that introns impose only a small (or no) burden, and so are tolerated in many lineages. However, a recent report by Jaillon *et al.* [2] reveals widespread intron mis-splicing, suggesting a significant cost associated with spliceosomal introns, and deepening the mystery of intron proliferation and persistence.

If introns are efficiently removed from transcripts before they are exported for translation, they should not respect coding meanings: whether or not an intron sequence contains a termination codon or a frameshift should be determined by chance. Jaillon *et al.* [2] report that this is not the case, however. Instead, the numerous (average 2.3 per gene) and very short (average length 25 bp) introns of the ciliate *Paramecium tetraurelia* show a pronounced preference for interrupted reading frames: 81.3% of *P. tetraurelia* introns have a frameshift (that is, they are not a multiple of three base pairs), as opposed to the expected two-thirds frequency. Moreover, those without a frameshift are twice as likely as frameshifting introns to have in-frame stop codons. Thus there is strong evolutionary selection against 'read-through' introns that could be translated into protein.

Translation of RNAs carrying premature stop codons can be prevented by the nonsense-mediated mRNA decay pathway (NMD) [3], and Jaillon *et al.* [2] experimentally tested the hypothesis that this pathway is responsible for removing mis-spliced or unspliced transcripts in *P. tetraurelia*. By knocking down a component of the NMD machinery, they revealed the intrinsic low efficiency of splicing for many introns and the essential role of NMD in preventing translation of the resulting unspliced transcripts. By a bioinformatics analysis of intron sequences in other eukaryotes, the authors conclude that such splicing inefficiency is likely to be widespread, at least for short introns.

The costs and benefits of spliceosomal introns

The findings of Jaillon *et al.* [2] have far-reaching implications for our understanding of the eukaryotic genome. First, the demonstrated inefficiency of splicing suggests that the presence of introns is even more disadvantageous to general fitness than previously appreciated. There is considerable wastage of mRNAs, and the maintenance of a complicated and effective monitoring system for the identification and removal of these transcripts - the NMD machinery - is required. In the case of *P. tetraurelia*, these burdens are borne by a species in which there is almost no functional alternative splicing [2], and so diversification of the proteome can be ruled out as a reason for intron presence in

the genome. The very short lengths of the introns in *Paramecium* also suggest a minor role, at most, in genomic stability, chromatin structure or in promoting recombination - other proposed advantages of introns. This intensifies the central mystery of eukaryotic gene structure: why did so many costly introns with no apparent function arise in the first place, and why are they retained in such a diverse array of species?

Introns in reduced genomes

Another recent report, by Pleiss *et al.* [4], shows the functional potential of unspliced transcripts. Many eukaryotes have reduced genomes with few introns. The few introns in *Saccharomyces cerevisiae* (only 0.05 per gene on average) show a previously mysterious bias towards ribosomal protein genes (0.74 introns per ribosomal gene). Pleiss *et al.* showed that introns in ribosomal protein genes, but not introns in other genes, are inefficiently spliced under amino-acid starvation, resulting in reduced production of protein-encoding spliced transcripts and thus presumably inhibiting ribosome formation and overall protein translation. This is not a general stress response, as the splicing of these introns remains unaffected under other, unrelated stress conditions (such as exposure to toxic levels of ethanol), nor does it reflect a general collapse of cellular processes due to stress, as the splicing of introns in non-ribosomal genes was not reduced [4].

The use of regulated splicing to serve some biological function is thus one potential explanation for the retention of occasional introns in reduced eukaryotic genomes. Together with the finding of widespread mis-splicing in *Paramecium* [2] and the observed low level of evolutionary conservation of alternative splice forms in metazoans (see [5] for a recent review), these results suggest an intriguingly counter-intuitive possibility. This is that a large fraction of the introns in reduced genomes may serve important functions, whereas the numerous introns and frequent alternative splicing of more intron-rich genomes may be largely non-functional [6-9].

Early eukaryotic genomes and the origins of alternative splicing

The work in *Paramecium* [2] also has important implications for the origins of alternative splicing. The inefficient splicing demonstrated suggests that eukaryotes may have been producing variable transcripts - presumably a requisite for the emergence of widespread alternative splicing [8] - early in their history. This would mean that variably spliced genes encoding multiple functional proteins need not have emerged by serendipitious, rare mutations from alleles producing single transcript forms. Instead, preexisting non-functional transcript variation is likely to have been co-opted for new functions. If the production of alternative transcripts from the same gene was already common early in eukaryote evolution [8,10,11], the question then becomes at what point(s) did functional alternative splicing emerge. On the other hand, it should be noted that most well characterized functional alternative splicing events do not involve the inclusion or exclusion of introns, but the inclusion or exclusion of exons, the evolution of which may proceed differently [10].

A recent flood of comparative genomic data indicates that early eukaryotes had complex genome structures, suggesting that compact intron-poor genomes are the most 'highly evolved' among eukaryotes. It has been shown that the last common ancestor of extant eukaryotes contained a complex spliceosome [12] and a large number of introns [13-18], and that those introns probably had degenerate sequences without strong consensus motifs [11,19]. The results of Jaillon *et al.* [2] emphasize how much of a burden those ancestral structures could have imposed - the many degenerate introns are likely to have been only inefficiently spliced, implying large numbers of unspliced and therefore useless transcripts. These results also indicate a strong selective force for the origin of intron-mediated NMD: the presence of large numbers of unspliced transcripts is likely to have driven the evolution of NMD, rather than the NMD system and introns being selected to deal with the (relatively infrequent) transcription errors. It is truly a mystery how such seemingly hapless and inefficient early eukaryotes could have succeeded in a world already well colonized by prokaryotes unburdened by such problems.

Acknowledgements

The authors thank Liran Carmel for extremely helpful comments on multiple drafts of the manuscript. MI was funded by the Spanish Ministerio of Educación y Ciencia, through the FPI grant (BFU2005-00252), and SWR by the Intramural Research Program of the National Library of Medicine at National Institutes of Health/DHHS. We thank Eugene Koonin and Jordi Garcia-Fernàndez and their groups for intellectual support and stimulation, for financial support, and for fostering environments of open intellectual exploration in their respective groups.

References

1. Nilsen TW: **The spliceosome: the most complex macromolecular machine in the cell?** *BioEssays* 2003, **25**:1147-1149.
2. Jaillon O, Bouhouche K, Gout J-F, Aury J-M, Noel B, Saudemont B, Nowacki M, Serrano V, Porcel BM, Ségurens B, Le Mouël A, Lepère G, Schächter V, Bétermier M, Cohen J, Wincker P, Sperling L, Duret L, Meyer E: **Translational control of intron splicing in eukaryotes.** *Nature* 2008, **451**:359-362.
3. Baker KE, Parker R: **Nonsense-mediated mRNA decay: terminating erroneous gene expression.** *Curr Opin Cell Biol* 2004, **16**:293-299.
4. Pleiss JA, Whitworth GB, Bergkessel M, Guthrie C: **Rapid, transcript-specific changes in splicing in response to environmental stress.** *Mol Cell* 2007, **27**:928-937.
5. Artamonova II, Gelfand MS: **Comparative genomics and evolution of alternative splicing: the pessimists' science.** *Chem Rev* 2007, **107**:3407-3430.
6. Sorek R, Shamir R, Ast G: **How prevalent is functional alternative splicing in the human genome?** *Trends Genet* 2004, **20**:68-71.
7. Rukov JL, Irimia M, Mork S, Lund VK, Vinther J, Arctander P: **High qualitative and quantitative conservation of alternative splicing in *Caenorhabditis elegans* and *Caenorhabditis briggsae*.** *Mol Biol Evol* 2007, **24**:909-917.

8. Irimia M, Rukov JL, Penny D, Roy SW: **Functional and evolutionary analysis of alternatively spliced genes is consistent with an early eukaryotic origin of alternative splicing.** *BMC Evol Biol* 2007, **7**:188.
9. Irimia M, Rukov JL, Penny D, Garcia-Fernandez J, Vinther J, Roy SW: **Widespread evolutionary conservation of alternatively spliced exons in *Caenorhabditis*.** *Mol Biol Evol* 2008, **25**:375-382.
10. Kim E, Goren A, Ast G: **Alternative splicing: current perspectives**. *BioEssays* 2008, **30**:38-47.
11. Irimia M, Penny D, Roy SW: **Coevolution of genomic intron number and splice sites.** *Trends Genet* 2007, **23**:321-325.
12. Collins L, Penny D: **Complex spliceosomal organization ancestral to extant eukaryotes.** *Mol Biol Evol* 2005, **22**:1053-1066.
13. Roy SW, Gilbert W: **Complex early genes.** *Proc Natl Acad Sci USA* 2005, **102**:1986-1991.
14. Sverdlov A, Rogozin I, Babenko V, Koonin E: **Conservation versus parallel gains in intron evolution.** *Nucleic Acids Res* 2005, **33**:1741-1748.
15. Rogozin I, Sverdlov A, Babenko V, Koonin E: **Analysis of evolution of exon-intron structure of eukaryotic genes.** *Brief Bioinform* 2005, **6**:118-134.
16. Yoshihama M, Nakao A, Nguyen HD, Kenmochi N: **Analysis of ribosomal protein gene structures: implications for intron evolution.** *PLoS Genet* 2006, **2**:e25.
17. Nguyen H, Yoshihama M, Kenmochi N: **New maximum likelihood estimators for eukaryotic intron evolution.** *PLoS Comput Biol* 2005, **1**:e79.
18. Slamovits CH, Keeling PJ: **A high density of ancient spliceosomal introns in oxymonad excavates.** *BMC Evol Biol* 2006, **6**:34.
19. Schwartz S, Silva J, Burstein D, Pupko T, Eyras E, Ast G: **Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes.** *Genome Res* 2008, **18**:88-103.