Genome Analysis

Rapidly regulated genes are intron poor

Daniel C. Jeffares[1*], Christopher J. Penkett[1,2*] and Jürg Bähler[1,3]

[1] Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK
[2] Current address: European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, UK
[3] Current address: UCL Cancer Institute and Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK

We show that genes with rapidly changing expression levels in response to stress contain significantly lower intron densities in yeasts, thale cress and mice. Therefore, we propose that introns can delay regulatory responses and are selected against in genes whose transcripts require rapid adjustment for survival of environmental challenges. These findings could provide an explanation for the apparent extensive intron loss during the evolution of some eukaryotic lineages.

Intron loss during evolution

Spliceosomal introns, the noncoding portions of genes between the protein-coding exons, are ancient and ubiquitous features of eukaryotic genomes but are absent from bacterial and archaeal genomes. Some eukaryotes maintain numerous large introns that can contain functional elements [1] and contribute to proteome diversity by facilitating alternative splicing [2]. Other eukaryotic lineages seem to have undergone extensive intron loss, resulting in virtually intronless genomes in some cases [3,4]. Intronloss events should be favoured by selection if introns were detrimental to their host [5]. However, any disadvantages that introns might confer are subject to much debate [6–9], and understanding the uneven phylogenetic distribution of introns is a major challenge for evolutionary genomics.

Various hypotheses have been proposed to explain the detrimental nature of introns with respect to gene expression. The observation that intron length is inversely correlated with transcript levels in humans and worms [10] supported the hypothesis of a selection for short introns in highly expressed genes. Intron length, however, is positively correlated with expression level in plants [11]. Human antisense transcripts have shorter introns than their sense partner, suggesting that long introns are deleterious because they prolong transcription [8]. Another hypothesis is that tissue-specific genes can contain shorter intronic and intergenic regions because they require more complex regulation and more compact chromatin to minimise ectopic expression [12]. Here, we examine the relationships between introns and gene expression in budding yeast (*Saccharomyces cerevisiae*), fission yeast (*Schizosaccharomyces pombe*), thale cress (*Arabidopsis thaliana*) and mice (*Mus musculus*).

Fewer introns in genes that are rapidly regulated during stress

We found that transcripts that were strongly modulated during stress in fission yeast, either increasing or decreasing in levels under five conditions [13], were significantly underenriched for introns ($P_{Hypergeometric} \sim 1.5 \times 10^{-22}$). This bias against introns among stress-response genes was strongest for the most highly regulated genes (Figure 1).

To further examine the relationship between gene regulation and introns, we analysed microarray time course data that follow global gene expression in response to various stress conditions in budding yeast [14], fission yeast [13], thale cress [15] and mice [16,17]. For each gene, we determined the maximal rate of change in transcript levels in each experiment (R_{stress}); we took the highest R_{stress} value as an estimate of the fastest rate of transcript change (R_{max}; see Supplementary Material). R_{max} estimates were inversely correlated with intron number in yeasts and thale cress (Figure 2a,b; Supplementary Figure S1a–c), whereas R_{max} estimates were positively correlated with intron number in mouse (Spearman rank correlation coefficients; budding yeast = −0.08, fission yeast = −0.22, thale cress = −0.26, mouse = 0.12, all $P < 4 \times 10^{-9}$). These observations remained robust when using only genes with expression levels greater than the median, indicating that R_{max} estimates were not biased by noise in array data (see Supplementary Material).

Possible costs of introns: kinetics of transcription and splicing

To understand the relative costs of introns, we must consider the kinetics of transcription and splicing, which occur concurrently in the nucleus [18]. Because transcription proceeds at 1200–1500 nucleotides per minute [19], yeasts and thale cress genes (with mean unspliced transcript lengths of ~1500 and 2000 nucleotides, respectively) are transcribed in ~1 min. Half-lives for splicing reactions are <1 min for the first intron, but 2–8 min for second and third introns [20]. Hence, splicing of two or more introns requires more time than transcription and becomes rate-limiting. The kinetics for splicing of additional introns, as is common in mice, is not well understood [20]. Unspliced mouse transcripts are much longer than those of yeast or thale cress (mean length, ~30 000 nucleotides) and have nine introns on average. Transcription of such genes requires ~23 min and can become rate-limiting.

Introns therefore could have at least three possible deleterious effects on gene expression: a delay in transcript

Corresponding author: Jeffares, D.C. (dcj@sanger.ac.uk).
[*] These authors contributed equally to this work.

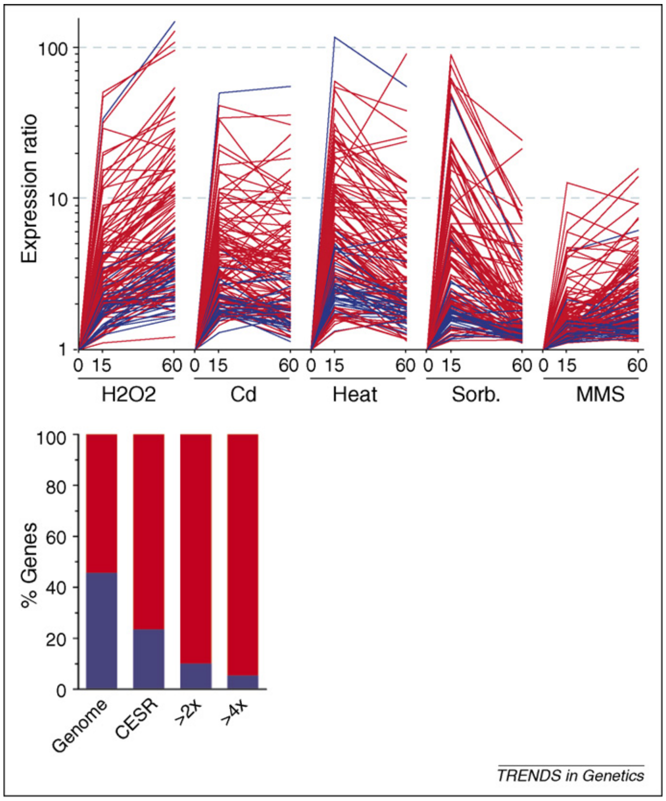

Figure 1. Fission yeast genes rapidly regulated during stress tend to have no introns. **(a)** Expression profiles of 96 intronless (red) and 33 intron-containing (blue) genes in five stress time course experiments: oxidative stress (H_2O_2), cadmium (Cd), heat shock, osmotic stress (sorbitol) and a DNA-damaging agent (MMS) [13]. Transcript profiles relative to unstressed cells are shown at 0, 15 and 60 min after stress exposure for the 129 genes that showed >1.1-fold induction in all stress conditions. **(b)** Histogram showing the percentages of genes with (red) and without (blue) introns in the following lists. Genome: all known ~5000 genes; CESR: 311 induced core environmental stress response genes [13]; >2×: 168 genes that were >2-fold induced within 15 min in at least three experiments; >4×: 55 genes that were >4-fold induced within 15 min in at least three experiments.

production caused by splicing, a delay in transcript production caused by the added length of the nascent transcript and an added energetic cost from increased transcript length. Any of these might be deleterious. The cost of splicing is independent of intron length but depends on intron number, whereas the kinetic and energetic costs of transcription are proportional to intron length. In yeasts and thale cress, where introns are relatively short, the costs of transcription are negligible compared with the cost of splicing. In mice, however, the considerably longer introns add much energetic cost to transcription, outweighing the cost of splicing. Consistent with this hypothesis, strongly transcribed human genes contain shorter introns than weakly transcribed genes [10]; we observed a similar inverse relationship between transcript levels and total intron length for mouse, but positive correlations between total intron length and expression level in both yeasts and thale cress (Supplementary Table 1), in accordance with published findings [10,11].

Consistent inverse relationships between intron density and rapid regulation

To normalize for transcript length, we used an 'intron density' measure (intron number/unspliced transcript length), which reflects only the cost of splicing by correcting for bias caused by the cost of transcription. Strikingly, R_{max} values showed significant inverse relationships with intron density in all organisms examined including mouse (Figure 2c; Supplementary Figure S1a–d; Spearman rank correlation coefficients; budding yeast = −0.08, fission yeast = −0.21, thale cress = −0.31, mouse = −0.07, all $P < 3 \times 10^{-7}$). Again, statistics were robust to variance from low expression genes (Supplementary Material).

Given the different correlations between gene expression and the length, number or density of introns, we used principal components analysis to determine the most consistent relationships for the four species tested. Each species showed different relationships between intron and gene-expression properties, and only the inverse correlation between R_{max} and intron density was found consistently in all species (Supplementary Figure S2; Supplementary Table S1). We conclude that rapidly regulated genes show significantly lower intron densities across a broad evolutionary spectrum.

Rapidly regulated cell cycle genes are also intron-poor

The inverse correlation between rapid gene expression changes and intron density should not be limited to stress-response genes but also apply to genes undergoing rapid regulation during cell proliferation, differentiation or development. Organisms with short generation times will require overall more rapid changes in gene expression, which might be reflected by the intron content in the regulated genes. Indeed, generation times are in general inversely correlated with genome-wide intron density [7]. To examine the relationship between introns and genes regulated during cell proliferation, we estimated R_{max} from fission yeast cell cycle data [21]. Transcripts whose levels changed most rapidly as a function of the cell cycle contained significantly fewer introns (Supplementary Figure S1e). This finding is consistent with introns being deleterious for the rapidly regulated intrinsic gene expression program that drives the short lifecycle of fission yeast.

Does the need for rapid gene regulation drive intron loss?

The inverse relationships between intron densities and gene expression changes in four diverse eukaryotes, covering fungi, plants and animals, raise the possibility that selection against introns in rapidly regulated genes has been a pervasive force in genome evolution. It is speculative to infer selection based on correlations, but this is certainly a plausible model given the current data. The inverse correlation between intron density and rapid gene regulation was the most consistent relationship across the four species tested. Moreover, in fission yeast, where we did additional analyses, this inverse correlation was evident for independent gene groups with different properties such as expression levels: rapidly regulated cell cycle genes, stress-induced genes and stress-repressed genes (that are rapidly induced on stress recovery). Together, these findings suggest that the relationship between intron density and rapid gene regulation is direct. Because R_{max} measures changes of total transcripts from whole cell extracts, it represents overall changes in transcript levels and not changes in spliced transcripts; the trivial explanation that intron-poor genes show higher R_{max} because they are rapidly spliced can therefore be discarded.

Intron loss events can occur via the recombination of reverse-transcribed cDNAs [4], providing the genetic variation for selection to act on. Although this process is biased in some ways [22,23], changes in the rate of transcription are not expected to affect intron loss. The model of selection against introns in rapidly regulated genes is also consistent with our understanding of molecular biology: because splicing is relatively slow compared with transcription [19,20], an intronless allele will produce its protein product more rapidly than a corresponding intron-containing allele. This decrease in response time could provide the selective advantage to fix intron loss events in the population or to prevent accumulation of intron gain alleles.

We propose a balance between selection against introns in genes that require rapid expression changes and against long introns in highly expressed genes. These distinct selection pressures seem to differ between organisms: the genomes of yeasts and thale cress reflect selection against introns to expedite gene regulation, whereas the mouse genome primarily reflects selection against long introns in highly expressed genes, with only weak selection for low intron density. Thus, mouse genes seem to be less optimized for rapid regulation, possibly reflecting that animals are less exposed to environmental challenges than plants and microorganisms. The proposed model can explain some of the large differences in intron numbers between eukaryotes, and it provides testable predictions about intron gain and loss.

Acknowledgements

The authors thank Anthony Poole, Samuel Marguerat, Manolis Dermitzakis and Paul Gardner for comments on the manuscript. This research was supported by Cancer Research UK and Wellcome Trust.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2008.05.006.

References

1 Coulombe-Huntington, J. and Majewski, J. (2007) Characterization of intron loss events in mammals. *Genome Res.* 17, 23–32

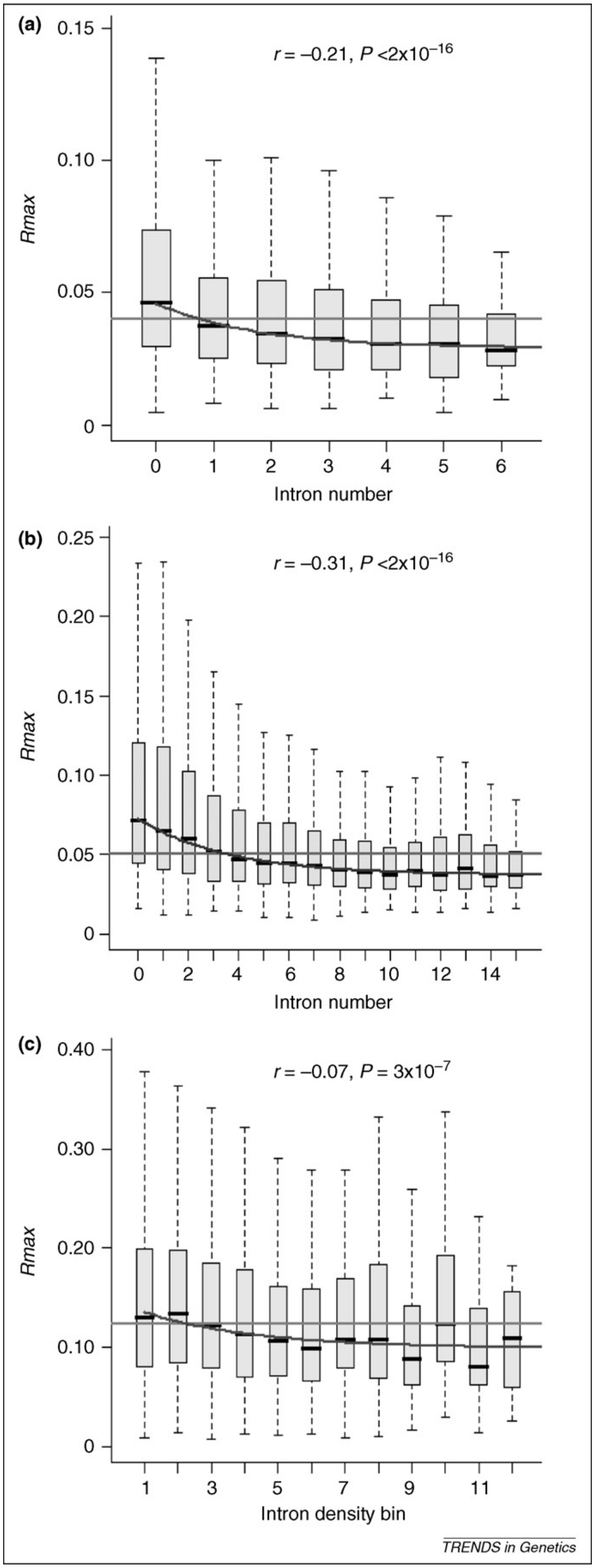

Figure 2. Rapidly regulated genes tend have lower intron content. Box plots show the relationship between introns and the maximum rate of change in transcript levels (R_{max}) during stress. To avoid bias, plots were generated from equal populations using a subsampling technique (see Supplementary Materials). Boxes indicate the range of R_{max} values for gene groups with different intron number or density; the bold central lines represent the median, whereas upper and lower boundaries represent 75th and 25th percentiles and whiskers extend to the furthest datapoint up to 1.5× the interquartile range. The horizontal grey lines show the median of all R_{max} values, and the bold curved lines show the best fitting exponential model for the median values of the entire datasets. The corresponding Spearman rank correlations are provided in the top right corners. Plots showing all data without subsampling are presented in Supplementary Figure S1. **(a)** Fission yeast R_{max} estimates from five stress experiments (as in Figure 1) as a function of intron number. Intronless genes have significantly higher R_{max} than intron-containing genes (median R_{max} of intron-less genes = 0.046, intron-containing genes = 0.035, Mann-Whitney, $P < 2.2 \times 10^{-16}$). **(b)** Thale cress R_{max} estimates from eight abiotic stresses (drought, UV-B, cold, heat, genotoxic, salt, wounding, osmotic [15]) as a function of intron number. Intronless genes have significantly higher R_{max} than intron-containing genes (median R_{max} of intron-less genes = 0.071, intron-containing genes = 0.047, Mann-Whitney, $P < 2.2 \times 10^{-16}$). **(c)** Mouse R_{max} estimates derived from two stress experiments (oxidative stress, bovine serum factor [16,17]) as a function of 20 equally sized bins of intron density. Genes in the lowest two density bins have significantly higher R_{max} than genes in the other bins (median R_{max} of genes in the lowest two density bins = 0.13, all other genes = 0.11, Mann-Whitney, $P = 2.6 \times 10^{-11}$).

2 Stetefeld, J. and Ruegg, M.A. (2005) Structural and functional diversity generated by alternative mRNA splicing. *Trends Biochem. Sci.* 30, 515–521
3 Roy, S.W. and Gilbert, W. (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat. Rev. Genet.* 7, 211–221
4 Rodriguez-Trelles, F. *et al.* (2006) Origins and evolution of spliceosomal introns. *Annu. Rev. Genet.* 40, 47–76
5 Lynch, M. (2002) Intron evolution as a population-genetic process. *Proc. Natl. Acad. Sci. U. S. A.* 99, 6118–6123
6 Lynch, M. and Conery, J.S. (2003) The origins of genome complexity. *Science* 302, 1401–1404
7 Jeffares, D.C. *et al.* (2006) The biology of intron gain and loss. *Trends Genet.* 22, 16–22
8 Chen, J. *et al.* (2005) Human antisense genes have unusually short introns: evidence for selection for rapid transcription. *Trends Genet.* 21, 203–207
9 Carmel, L. *et al.* (2007) Three distinct modes of intron dynamics in the evolution of eukaryotes. *Genome Res.* 17, 1034–1044
10 Castillo-Davis, C.I. *et al.* (2002) Selection for short introns in highly expressed genes. *Nat. Genet.* 31, 415–418
11 Ren, X.Y. *et al.* (2006) In plants, highly expressed genes are the least compact. *Trends Genet.* 22, 528–532
12 Vinogradov, A.E. (2004) Compactness of human housekeeping genes: selection for economy or genomic design? *Trends Genet.* 20, 248–253
13 Chen, D. *et al.* (2003) Global transcriptional responses of fission yeast to environmental stress. *Mol. Biol. Cell* 14, 214–229
14 Causton, H.C. *et al.* (2001) Remodeling of yeast genome expression in response to environmental changes. *Mol. Biol. Cell* 12, 323–337
15 Craigon, D.J. *et al.* (2004) NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. *Nucleic Acids Res.* 32, D575–D577
16 Philippar, U. *et al.* (2004) The SRF target gene Fhl2 antagonizes RhoA/MAL-dependent activation of SRF. *Mol. Cell* 16, 867–880
17 Madsen, M.A. *et al.* (2004) Altered oxidative stress response of the long-lived Snell dwarf mouse. *Biochem. Biophys. Res. Commun.* 318, 998–1005
18 Neugebauer, K.M. (2002) On the importance of being co-transcriptional. *J. Cell Sci.* 115, 3865–3871
19 Izban, M.G. and Luse, D.S. (1992) Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. *J. Biol. Chem.* 267, 13647–13655
20 Audibert, A. *et al.* (2002) *In vivo* kinetics of mRNA splicing and transport in mammalian cells. *Mol. Cell. Biol.* 22, 6706–6718
21 Rustici, G. *et al.* (2004) Periodic gene expression program of the fission yeast cell cycle. *Nat. Genet.* 36, 809–817
22 Mourier, T. and Jeffares, D.C. (2003) Eukaryotic intron loss. *Science* 300, 1393
23 Zhang, Z. *et al.* (2003) Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res.* 13, 2541–2558

0168-9525/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tig.2008.05.006 Available online 27 June 2008

Genome Analysis

Origin of introns by 'intronization' of exonic sequences

Manuel Irimia[1], Jakob Lewin Rukov[2], David Penny[3], Jeppe Vinther[2], Jordi Garcia-Fernandez[1] and Scott William Roy[4]

[1] Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Barcelona 08028, Spain
[2] Molecular Evolution Group, Department of Biology, University of Copenhagen, Copenhagen DK-2200, Denmark
[3] Allan Wilson Centre for Molecular Evolution and Ecology, Massey University, Palmerston North, New Zealand
[4] National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20892, USA

The mechanisms of spliceosomal intron creation have proved elusive. Here we describe a new mechanism: the recruitment of internal exonic sequences ('intronization') in *Caenorhabditis* species. The numbers of intronization events and introns gained by other mechanisms are similar, suggesting that intronization significantly contributes to recent intron creation in nematodes. Intronization is more common than the reverse process, loss of splicing of retained introns. Finally, these findings link alternative splicing with modern intron creation.

The elusive origins of introns and of alternative splicing

The origin of spliceosomal introns is one of molecular biology's longest standing mysteries. Very few cases of intron 'birth' have been thoroughly documented, although five mechanisms have been proposed to explain how introns are formed: (i) intron transposition [1–3]; (ii) transposon insertion [4,5]; (iii) tandem genomic duplication [6]; (iv) intron transfer between paralogs [7]; and (v) self-splicing type II intron insertion [6,8]. Furthermore, previously reported examples of seemingly clear intron origin have been called into question on further scrutiny [9,10].

Another major evolutionary puzzle concerns the origins and functional significance of alternative splicing [11–15]. Ubiquitous alternative splicing offers a plausible explanation for the organismal complexity of vertebrates and metazoans in general; however, the low level of evolutionary conservation of alternatively spliced exons between mouse and human suggests that much alternative splicing in mammals might be nonfunctional [16,17]. Intriguingly, the levels of evolutionary conservation of alternatively spliced exons seem to be much greater among *Caenorhabditis* nematodes than among mammals [13,17–19].

We explored the relationship between the gain and loss of introns and the evolution of alternative splicing. First, we compiled all 50 cases of retained introns (in which two alternative transcripts of a gene differ by inclusion or removal of an intron) that are currently annotated in the *Caenorhabditis elegans* genome and for which there is supporting cDNA/EST sequence (see online Supplementary Materials for full details). For each of these introns, we investigated orthologous sequences in five *Caenorhabditis* species (*C. elegans*, *C. briggsae*, *C. remanei*, *C. brenneri* and *C. japonica*) (Figure 1).

Corresponding author: Roy, S.W. (royscott@ncbi.nlm.nih.gov).