Endogenous Mechanisms for the Origins of Spliceosomal Introns

FRANCESCO CATANIA, XIANG GAO, AND DOUGLAS G. SCOFIELD

From the Department of Biology, Indiana University, 1001 East Third Street, Bloomington, IN 47405 (Catania and Gao); and the Department of Ecology and Evolutionary Biology, University of California, 621 Charles E. Young Drive South, Box 951606, Los Angeles, CA 90095-1606 (Scofield).

Address correspondence to Francesco Catania at the address above, or e-mail: fcatania@indiana.edu.

Abstract

Over 30 years since their discovery, the origin of spliceosomal introns remains uncertain. One nearly universally accepted hypothesis maintains that spliceosomal introns originated from self-splicing group-II introns that invaded the uninterrupted genes of the last eukaryotic common ancestor (LECA) and proliferated by "insertion" events. Although this is a possible explanation for the original presence of introns and splicing machinery, the emphasis on a high number of insertion events in the genome of the LECA neglects a considerable body of empirical evidence showing that spliceosomal introns can simply arise from coding or, more generally, nonintronic sequences within genes. After presenting a concise overview of some of the most common hypotheses and mechanisms for intron origin, we propose two further hypotheses that are broadly based on central cellular processes: 1) internal gene duplication and 2) the response to aberrant and fortuitously spliced transcripts. These two nonmutually exclusive hypotheses provide a powerful way to explain the establishment of spliceosomal introns in eukaryotes without invoking an exogenous source.

Key words: *group-II introns, internal gene duplication, intronization, spliceosomal introns*

Ever since the discovery of spliceosomal introns (Berget et al. 1977; Chow et al. 1977; Evans et al. 1977; Goldberg et al. 1977), an enormous number of multidisciplinary studies have shed light on the evolutionary features of these noncoding sequences and revealed peculiar links between splicing and a number of other cellular processes (Proudfoot et al. 2002; Reed and Hurt 2002). Yet, how (and why) spliceosomal introns first appeared remains an unsolved issue.

A Commonly Accepted View: Spliceosomal Introns Originate from Group-II Introns

Despite the several plausible attempts to explain the origin of spliceosomal introns (Belshaw and Bensasson 2006; Rodriguez-Trelles et al. 2006; Roy and Gilbert 2006), one single hypothesis appears nearly universally accepted, although essentially not verifiable. This hypothesis maintains that spliceosomal introns, together with the spliceosome (s)—a large ribonucleoprotein complex that is actively involved in splicing (Burge et al. 1999; Brow 2002; Kuhn and Kaufer 2003; Nilsen 2003)—originated from a class of self-splicing introns (group-II) (Sharp 1985; Cech 1986;

Jacquier 1990; Cavalier-Smith 1991; Palmer and Logsdon 1991). Briefly, group-II introns of α-proteobacterial endosymbionts (presumably primordial mitochondria) of the last eukaryotic common ancestor (LECA) are thought to have first invaded the previously uninterrupted nuclear genes and subsequently lost the ability of self-splicing, eventually coming to be spliced in "trans" by the newly invented spliceosomal machinery (Stoltzfus 1999). In line with this presumptive mode of origin, group-II introns and spliceosomal introns share several structural and functional similarities (Shukla and Padgett 2002; Valadkhan and Manley 2002), and plausible steps have been suggested to have allowed the conversion between the two classes of introns (Lynch and Richardson 2002). Although group-II introns provide a reasonable source for the original presence of spliceosomal introns and of the splicing machinery in the LECA, some aspects of the group-II introns origin hypothesis remain unclear, especially the conjectured widespread invasion of uninterrupted LECA genes. In particular, although group-II introns have been reported in the organelles of plants and lower eukaryotes (Bonen and Vogel 2001) as well as in eubacteria and archea (Dai and Zimmerly 2002, 2003; Rest and Mindell 2003; Toro 2003), neither functional copies nor relics of such sequences have ever been found in animal mitochondria or in eukaryotic

nuclei. A major loss of group-II introns subsequent to spliceosomal intron colonization cannot be entirely ruled out, but the current absence of group-II introns from eukaryotic nuclear genomes conflicts with the described evidence for putative recent intron gains (Coulombe-Huntington and Majewski 2007; Roy and Penny 2007; Omilian et al. 2008).

Origin of Spliceosomal Introns from Alternative Sources?

As opposed to widespread invasion of LECA genes by group-II introns, other hypotheses have been proposed that do not invoke an exogenous invading source. Transposable elements have been suggested as one possible endogenous source. After inserting into the coding sequence, a transposon would be removed from the pre-mRNA either by means of latent splice sites that may exist in the inserted element or after the acquisition of splicing signals. The very few spliceosomal introns that are known to derive from mobile elements (Giroux et al. 1994; Iwamoto et al. 1999) give limited support for the generality of this hypothesis. Moreover, excision of transposable elements that are inserted into genes is only rarely perfect (e.g., Giroux et al. 1994; Greco et al. 2005) and in most cases generates indels (e.g., Wessler 1989; Fridell et al. 1990; Giroux et al. 1994; Rushforth and Anderson 1996) so that the mutant gene either partially or completely retains the wild-type gene's phenotype only if the open reading frame (ORF) of the sequence downstream of the insertion site is not disrupted.

Spliceosomal introns have also been suggested to directly arise from random primordial and canonical ancestral gene sequences (the "split-gene model" and the "proto-splice site model," respectively). In particular, Senapathy (1986) observed that the distribution of reading-frame lengths in a random nucleotide sequence is exponential (but see Hoglund et al. 1990), with a 600-nt upper limit that is in line with a number of observed average exon sizes (Hawkins 1988; Yandell et al. 2006). Under the assumption that in the most primitive unicellular eukaryotes a selective pressure existed to generate longer coding sequences, the split-gene model (Senapathy 1986, 1988) maintains that random sequences that were populated with in-frame nonsense codons, and intervened between short reading frames, started to be excised by an already existing spliceosome. The model proposes not only that the sequences excised contained random clusters of in-frame nonsense codons but also that the splice junction signal sequences and the branchpoint sequence originate from nonsense codons (Senapathy 1988). One implication of this model is the existence of a nuclear scanning mechanism to recognize in-frame termination codons. Although some of the split-gene model's propositions have been challenged (Stoltzfus et al. 1995), the hypothesis that nonsense codons may have played a role in the origin of introns remains conceivable (see below and Senapathy 1988; Harris and Senapathy 1990).

In the proto-splice site model (Dibb 1991), introns arise at proto-splice sites, which contain the partially conserved exon consensus sequence [C|A]AG$^{\downarrow}$R adjacent to introns, where "$^{\downarrow}$" is the point of intron insertion (Dibb and Newman 1989). In this model, spliceosomal introns are originally coding and began as a result of the evolution of alternatively spliced isoforms of a transcript. In particular, initially inefficient splicing occurring at proto-splice sites would lead to the emergence of alternative transcript isoforms with subsequent mutations that stabilize beneficial isoforms. Splicing would become constitutive as a result of this selective filter and the new intronic sequences would drift so that any reinsertion in the mature transcript would be deleterious. To deal with the possible criticism that ancient proteins must have been very large if all introns originated from coding sequences, the model also maintains that only a limited number of introns would originate from the hypothetical process described above, and intron proliferation would largely proceed by insertion events driven by the spliceosomal machinery through the mechanism of reverse splicing (Tseng and Cheng 2008). Although a number of this model's observations appear to be consistent with the insertion of introns at proto-splice sites (Dibb and Newman 1989; Lee et al. 1991; Frugoli et al. 1998; Logsdon et al. 1998; Funke et al. 1999; Bhattacharya et al. 2000; Kent and Zahler 2000; Sverdlov et al. 2003; Coghlan and Wolfe 2004; Qiu et al. 2004; Sadusky et al. 2004; Sverdlov et al. 2004; Yoshihama et al. 2006; Omilian et al. 2008), no convincing sequence similarity has been ever observed between (putative) recent intron gains and available genomic sequences.

Finally, a related mechanism for intron origin that does not invoke exogenous sequences proposes that tandem duplication of a coding region containing an AGGT sequence could elicit splicing by providing both a 5′ donor and a 3′ acceptor sites (Rogers 1989). The validity of this model is supported by a recent study in humans where the authors found an excess of introns that contain perfect matching sequences at their boundaries (Zhuo et al. 2007) (see also Chabot et al. 2008; Roy and Irimia 2008), with most of these introns being young and undergoing alternative splicing. Interestingly, repeats and alternative splicing have been also shown to be associated with the origin of new exons in vertebrates (Zhang and Chasin 2006).

Empirical Observations in Favor of an Endogenous Origin of Spliceosomal Introns

Hypotheses for the endogenous origin of introns have been challenged, and sometimes discarded, on the basis both of current data and perhaps our biased perception of how gene structure has evolved over time, for example, the frequent assumption that parallel evolution occurs at very low frequency. Recent evidence for the parallel origin of unrelated introns at nearly the same insertion site in *Daphnia pulex* (Omilian et al. 2008) directly challenges this assumption.

Although we may not know for certain how the first introns originated, it is worth emphasizing that we do know, by empirical observations, how individual spliceosomal introns can originate. In plants, efficient splicing typically requires an elevated intron AT content (Goodall and Filipowicz 1989, 1991), probably because the latter favors the association of splicing-enhancing RNA-binding proteins (McCullough et al. 1993; Simpson and Filipowicz 1996). Interestingly, elevated AT content is sufficient for splicing of a nonintronic sequence via activation of flanking cryptic splice sites in a legumin J gene construct (Simpson and Brown 1993). Also, an AT-rich region in the 5′ end of the transferred DNA *rolA* gene of *Agrobacterium rhizogenes*, unsurprisingly unspliced in the bacterium, undergoes efficient splicing in *Arabidopsis* plants that are transgenic for this gene (Magrelli et al. 1994). In *Caenorhabditis elegans*, de novo introns can be created by recruitment of internal exonic sequences (Irimia et al. 2008), whereas in humans, de novo introns have been suggested to result from 1) fortuitous creation of splice sites (Courseaux and Nahon 2001); 2) activation of cryptic splice sites after the insertion of a transposable element (Sela et al. 2007); 3) tandem duplications (Zhuo et al. 2007); and 4) nucleotide substitutions that do not disrupt splice sites (Gromoll et al. 2007).

Origin of Spliceosomal Introns by Internal Gene Duplication and Intronization of Translatable Sequences

Along the same lines of the previously mentioned endogenous models, we propose here two nonmutually exclusive hypotheses that do not invoke an ancient exogenous source for intron origin but where introns constantly emerge 1) following events of internal gene duplication (we call this the "internal gene duplication" model) (Gao X, unpublished data); and 2) by "intronization" of translatable sequences (the intronization model) (Catania and Lynch 2008).

The former hypothesis is observation driven. Briefly, by surveying multiple eukaryotic genomes, we find that 8–17% of genes display internally duplicated sequences and that internal gene duplication is a steady-state birth-and-death process whose rates are comparable to those reported for events of whole-gene duplication (Lynch and Conery 2003). Surprisingly, we find that 7–30% of internally duplicated genes show new intron gains. These new introns derive only in small part from the direct duplication of prior introns within the same gene; the majority are represented by de novo introns that likely emerged from the duplicated sequence after both the spatial change and the acquisition of new/activation of latent splice sites (Gao X, unpublished data).

The second model attempts to integrate the diverse body of observations that have been reported to date on splicing and interrelated mRNA-associated processes. In the intronization model (Catania and Lynch 2008), the systems for mRNA surveillance, capping, and cleavage/polyadenylation interact to play pivotal roles in the physical establishment and distribution of spliceosomal introns along a transcript.

In brief, the model is based on the cell's ability to filter out aberrant transcripts, whereas sparing those mRNAs where the imperfect splice-site recognition of the spliceosome machinery has fortuitously removed ORF-disrupting mutations such as premature termination codons (PTCs). In the model, we propose also that cleavage/polyadenylation factors (CPFs) regularly access U-rich tracts along the mRNA during transcription but are antagonized (or interfered with) by splicing factors (SFs), when U-rich regions are located within an intron. Under this scenario, the interaction between SFs and CPFs modulates the likelihood of imperfect splice-site recognition, thereby defining the physical setting for the facilitation or inhibition of intron colonization (Catania and Lynch 2008).

The Crucial Role of Cellular Surveillance Systems in the Endogenous Origin of Introns

Our endogenous models rest on several forms of gene structural disruptions. Thus, we expect that transcript surveillance systems, such as nonsense-mediated decay (NMD), must play a significant role in the generation of new endogenous introns. NMD is a virtually ubiquitous cellular surveillance system in eukaryotes that recognizes and selectively degrades so-called "nonsense" transcripts that contain PTCs upstream of the true stop codon (Maquat 2004, 2006; Conti and Izaurralde 2005; Lynch et al. 2006). Translation of these nonsense transcripts would result in truncated polypeptides, and the failure of NMD surveillance has been implicated in a number of human diseases (Frischmeyer and Dietz 1999; Holbrook et al. 2004). Nonsense transcripts can occur via a variety of mechanisms, including point mutations, indels, transposition/viral insertions, and splicing mistakes and other errors during transcript processing (Mendell et al. 2004) as well as during production of alternatively spliced transcripts (Green et al. 2003; Lewis et al. 2003).

Regardless of the exact molecular mechanism by which degradation of transcripts is initiated (Isken and Maquat 2008), the near ubiquity of general NMD mechanisms in eukaryotes is consistent with the elevated frequency, as well as the likely negative selective importance, of the basic forms of gene structural disruptions that underlie our endogeneous models of intron origin. On appearance, transcripts featuring either internal duplications or de novo intronization are stabilized via NMD degradation of PTC-containing isoforms and the rapidity of fixation is dependent on the accumulation of additional structural modifications that stabilize non-PTC-containing isoforms.

Interconnection between Internal Gene Duplication and Intronization and the Role of Alternative Splicing

Although distinct, it is worth emphasizing that the two mechanisms for endogeneous intron origins we present are
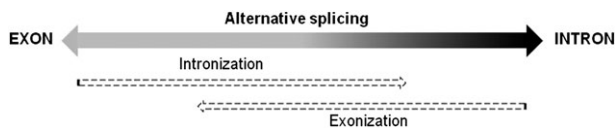
Figure 1. Creation of coding sequences from introns (Exonization) and creation of spliceosomal introns from coding exons (Intronization) are proposed to be reversible processes that occur not instantaneously but through a transient phase of gradual conversion (Alternative splicing).

not mutually exclusive. If fortuitously an internal duplication does not introduce a PTC, then a population of functional transcripts is maintained without NMD. However, if (a portion of) a duplicated exonic region acquires a PTC (see Letunic et al. 2002), then NMD would degrade the primary PTC-containing isoform and the mutated coding sequence could undergo intronization (Catania and Lynch 2008).

A joint action of intronization and internal gene duplication could be conjectured to contribute to maintain optimal protein length. Specifically, because of the frequent occurrence of events of internal gene duplication across several eukaryotes (Kondrashov and Koonin 2001; Letunic et al. 2002; Gao X, unpublished data), molecular mechanisms must exist that assist the removal of the newly created redundant sequences and thereby prevent proteins from growing extremely large. Intronization could represent one of these mechanisms. Once a duplicated sequence converts into an intron, that intron could be lost through the frequent but poorly understood process of intron loss, thus contributing to reduce the increase in gene length resulted by the prior event of duplication.

Finally, intronization of translatable sequences does not occur instantaneously but is predicted to undergo a transient phase involving alternative splicing (Catania and Lynch 2008). Under this view, intronization would not only complement the established process of exonization, where exons arise gradually from intronic sequences (Alekseyenko et al. 2007), but could also be coupled with internal gene duplication, in that young introns that are by-products of tandem duplication have been found to often undergo alternative splicing (Zhuo et al. 2007).

An interesting question associated with such a scenario—in which sequences may gradually and reversibly shift from an exonic to an intronic state over time (Figure 1)—is whether the time interval between sequence type conversion is comparable among eukaryotes. Although it remains unclear how levels of alternative splicing varies across lineages and what fraction of alternative isoforms are truly functional and thus maintained by natural selection, it is tempting to speculate that the duration of the processes of exonization and intronization may vary across species. In particular, the length of the transient phase of alternative splicing could be expected to be reduced in species with large N_e, where natural selection would more efficiently promote the fixation of optimal splicing signals—under the assump-

tion that maintaining alternative splicing is typically mildly deleterious. Under the aforementioned suggestion that exonic sequences may intronize and be subsequently lost, the relative fitness disadvantage that is typically associated with introns (Lynch 2002) may be offset to a greater or lesser degree by the simultaneously acquired selective advantage associated with the production of shorter but potentially more functional proteins. The latter idea is consistent with the observation that 1) lineages having a large N_e contain shorter protein length than lineages with relatively smaller N_e (e.g., Xu et al. 2006) and 2) introns abound in lineages with small N_e (Lynch and Conery 2003).

Conclusions

We have briefly outlined a number of observations, published and novel (Gao X, unpublished data), as well as a theoretical model (the intronization model, Catania and Lynch 2008), which 1) are consistent with the idea that spliceosomal introns may have originated endogenously and that 2) shed a new light both on the evolution of the eukaryotic gene structure and on the biological role of alternative splicing. Also, our endogeneous models of intron origin introduce additional mechanisms whereby gene structural evolution can be mediated by NMD (Lynch and Kewalramani 2003; Lynch et al. 2006; Scofield et al. 2007). Finally, we have conjectured that introns may act as a link between internal gene duplication and intronization, where intron loss (successive to intronization) contributes to counteract the expansion of gene length mediated by events of internal gene duplication.

Funding

References

Alekseyenko AV, Kim N, Lee CJ. 2007. Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes. RNA. 13:661–670.

Belshaw R, Bensasson D. 2006. The rise and falls of introns. Heredity. 96:208–213.

Berget SM, Moore C, Sharp PA. 1977. Spliced segments at 5′ terminus of adenovirus 2 late messenger-Rna. Proc Natl Acad Sci USA. 74:3171–3175.

Bhattacharya D, Lutzoni F, Reeb V, Simon D, Nason J, Fernandez F. 2000. Widespread occurrence of spliceosomal introns in the rDNA genes of ascomycetes. Mol Biol Evol. 17:1971–1984.

Bonen L, Vogel J. 2001. The ins and outs of group II introns. Trends Genet. 17:322–331.

Brow DA. 2002. Allosteric cascade of spliceosome activation. Annu Rev Genet. 36:333–360.

Burge CB, Tuschl T, Sharp PA. 1999. Splicing of precursors to mRNAs by the spliceosomes. In: Gesteland RF, Cech T, Atkins JF, editors. The RNA world. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. p. 525–560.

Catania F, Lynch M. 2008. Where do introns come from? PLoS Biol. 6:e283 doi:210.1371/journal.pbio.0060283.

Cavalier-Smith T. 1991. Intron phylogeny—a new hypothesis. Trends Genet. 7:145–148.

Cech TR. 1986. The generality of self-splicing Rna—relationship to nuclear messenger-Rna splicing. Cell. 44:207–210.

Chabot B, Elela SA, Zhuo D. 2008. Comment on "when good transcripts go bad: artifactual RT-PCR 'splicing' and genome analysis". Bioessays. 30:1256author reply 1257–1258

Chow LT, Gelinas RE, Broker TR, Roberts RJ. 1977. Amazing sequence arrangement at 5′ ends of adenovirus-2 messenger-Rna. Cell. 12:1–8.

Coghlan A, Wolfe KH. 2004. Origins of recently gained introns in Caenorhabditis. Proc Natl Acad Sci USA. 101:11362–11367.

Conti E, Izaurralde E. 2005. Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. Curr Opin Cell Biol. 17:316–325.

Coulombe-Huntington J, Majewski J. 2007. Characterization of intron loss events in mammals. Genome Res. 17:23–32.

Courseaux A, Nahon JL. 2001. Birth of two chimeric genes in the Hominidae lineage. Science. 291:1293–1297.

Dai LX, Zimmerly S. 2002. Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. Nucleic Acids Res. 30:1091–1102.

Dai LX, Zimmerly S. 2003. ORF-less and reverse-transcriptase-encoding group II introns in archaebacteria, with a pattern of homing into related group II intron ORFs. RNA. 9:14–19.

Dibb NJ. 1991. Proto-splice site model of intron origin. J Theor Biol. 151:405–416.

Dibb NJ, Newman AJ. 1989. Evidence that introns arose at proto-splice sites. Embo J. 8:2015–2021.

Evans RM, Fraser N, Ziff E, Weber J, Wilson M, Darnell JE. 1977. Initiation sites for Rna-transcription in Ad2 DNA. Cell. 12:733–740.

Fridell RA, Pret AM, Searles LL. 1990. A retrotransposon 412 insertion within an exon of the Drosophila melanogaster vermilion gene is spliced from the precursor RNA. Genes Dev. 4:559–566.

Frischmeyer PA, Dietz HC. 1999. Nonsense-mediated mRNA decay in health and disease. Hum Mol Genet. 8:1893–1900.

Frugoli JA, McPeek MA, Thomas TL, McClung CR. 1998. Intron loss and gain during evolution of the catalase gene family in angiosperms. Genetics. 149:355–365.

Funke RP, Kovar JL, Logsdon JM, Corrette-Bennett JC, Straus DR, Weeks DP. 1999. Nucleus-encoded, plastid-targeted acetolactate synthase genes in two closely related chlorophytes, Chlamydomonas reinhardtii and Volvox carteri: phylogenetic origins and recent insertion of introns. Mol Gen Genet. 262:12–21.

Giroux MJ, Clancy M, Baier J, Ingham L, McCarty D, Hannah LC. 1994. De novo synthesis of an intron by the maize transposable element dissociation. Proc Natl Acad Sci USA. 91:12150–12154.

Goldberg S, Schwartz H, Darnell JE. 1977. Evidence from UV transcription mapping in Hela-cells that heterogeneous nuclear-Rna is messenger-Rna precursor. Proc Natl Acad Sci USA. 74:4520–4523.

Goodall GJ, Filipowicz W. 1989. The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. Cell. 58:473–483.

Goodall GJ, Filipowicz W. 1991. Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. Embo J. 10:2635–2644.

Greco R, Ouwerkerk PB, Pereira A. 2005. Suppression of an atypically spliced rice CACTA transposon transcript in transgenic plants. Genetics. 169:2383–2387.

Green RE, Lewis BP, Hillman RT, Blanchette M, Lareau LF, Garnett AT, Rio DC, Brenner SE. 2003. Widespread predicted nonsense-mediated mRNA decay of alternatively-spliced transcripts of human normal and disease genes. Bioinformatics. 19(1):i118–i121.

Gromoll J, Lahrmann L, Godmann M, Muller T, Michel C, Stamm S, Simoni M. 2007. Genomic checkpoints for exon 10 usage in the luteinizing hormone receptor type 1 and type 2. Mol Endocrinol. 21:1984–1996.

Harris NL, Senapathy P. 1990. Distribution and consensus of branch point signals in eukaryotic genes—a computerized statistical-analysis. Nucleic Acids Res. 18:3015–3019.

Hawkins JD. 1988. A survey on intron and exon lengths. Nucleic Acids Res. 16:9893–9908.

Hoglund M, Sall T, Rohme D. 1990. On the origin of coding sequences from random open reading frames. J Mol Evol. 30:104–108.

Holbrook JA, Neu-Yilik G, Hentze MW, Kulozik AE. 2004. Nonsense-mediated decay approaches the clinic. Nat Genet. 36:801–808.

Irimia M, Rukov JL, Penny D, Vinther J, Garcia-Fernandez J, Roy SW. 2008. Origin of introns by 'intronization' of exonic sequences. Trends Genet. 24:378–381.

Isken O, Maquat LE. 2008. The multiple lives of NMD factors: balancing roles in gene and genome regulation. Nat Rev Genet. 9:699–712.

Iwamoto M, Nagashima H, Nagamine T, Higo H, Higo K. 1999. p-SINE1-like intron of the CatA catalase homologs and phylogenetic relationships among AA-genome Oryza and related species. Theor Appl Genet. 98:853–861.

Jacquier A. 1990. Self-splicing group-II and nuclear pre-messenger-Rna introns—how similar are they. Trends Biochem Sci. 15:351–354.

Kent WJ, Zahler AM. 2000. Conservation, regulation, synteny, and introns in a large-scale C. briggsae–C. elegans genomic alignment. Genome Res. 10:1115–1125.

Kondrashov FA, Koonin EV. 2001. Origin of alternative splicing by tandem exon duplication. Hum Mol Genet. 10:2661–2669.

Kuhn AN, Kaufer NF. 2003. Pre-mRNA splicing in Schizosaccharomyces pombe: regulatory role of a kinase conserved from fission yeast to mammals. Curr Genet. 42:241–251.

Lee VD, Stapleton M, Huang B. 1991. Genomic structure of Chlamydomonas-caltractin—evidence for intron insertion suggests a probable genealogy for the Ef-hand superfamily of proteins. J Mol Biol. 221:175–191.

Letunic I, Copley RR, Bork P. 2002. Common exon duplication in animals and its role in alternative splicing. Hum Mol Genet. 11:1561–1567.

Lewis BP, Green RE, Brenner SE. 2003. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci USA. 100:189–192.

Logsdon JM, Stoltzfus A, Doolittle WF. 1998. Molecular evolution: recent cases of spliceosomal intron gain? Curr Biol. 8:R560–R563.

Lynch M. 2002. Intron evolution as a population-genetic process. Proc Natl Acad Sci USA. 99:6118–6123.

Lynch M, Conery JS. 2003. The origins of genome complexity. Science. 302:1401–1404.

Lynch M, Hong X, Scofield DG. 2006. NMD and the evolution of eukaryotyic gene structure. In: Maquat LE, editor. Nonsense-mediated mRNA decay. Georgetown (TX): Landes Bioscience. p. 197–211.

Lynch M, Kewalramani A. 2003. Messenger RNA surveillance and the evolutionary proliferation of introns. Mol Biol Evol. 20:563–571.

Lynch M, Richardson AO. 2002. The evolution of spliceosomal introns. Curr Opin Genet Dev. 12:701–710.

Magrelli A, Langenkemper K, Dehio C, Schell J, Spena A. 1994. Splicing of the rolA transcript of *Agrobacterium rhizogenes* in *Arabidopsis*. Science. 266:1986–1988.

Maquat LE. 2004. Nonsense-mediated mRNA decay: a comparative analysis of different species. Curr Genomics. 5:175–190.

Maquat LE. 2006. Nonsense-mediated mRNA decay. Georgetown (TX): Landes Bioscience.

McCullough AJ, Lou H, Schuler MA. 1993. Factors affecting authentic 5′ splice site selection in plant nuclei. Mol Cell Biol. 13:1323–1331.

Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC. 2004. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat Genet. 36:1073–1078.

Nilsen TW. 2003. The spliceosome: the most complex macromolecular machine in the cell? Bioessays. 25:1147–1149.

Omilian AR, Scofield DG, Lynch M. 2008. Intron presence–absence polymorphisms in *Daphnia*. Mol Biol Evol. 25:2129–2139.

Palmer JD, Logsdon JM Jr. 1991. The recent origins of introns. Curr Opin Genet Dev. 1:470–477.

Proudfoot NJ, Furger A, Dye MJ. 2002. Integrating rnRNA processing with transcription. Cell. 108:501–512.

Qiu WG, Schisler N, Stoltzfus A. 2004. The evolutionary gain of spliceosomal introns: sequence and phase preferences. Mol Biol Evol. 21:1252–1263.

Reed R, Hurt E. 2002. A conserved mRNA export machinery coupled to pre-mRNA splicing. Cell. 108:523–531.

Rest JS, Mindell DP. 2003. Retroids in Archaea: phylogeny and lateral origins. Mol Biol Evol. 20:1134–1142.

Rodriguez-Trelles F, Tarrio R, Ayala FJ. 2006. Origins and evolution of spliceosomal introns. Ann Rev Genet. 40:47–76.

Rogers JH. 1989. How were introns inserted into nuclear genes. Trends Genet. 5:213–216.

Roy SW, Gilbert W. 2006. The evolution of spliceosomal introns: patterns, puzzles and progress. Nat Rev Genet. 7:211–221.

Roy SW, Irimia M. 2008. When good transcripts go bad: artifactual RT-PCR 'splicing' and genome analysis. Bioessays. 30:601–605.

Roy SW, Penny D. 2007. A very high fraction of unique intron positions in the intron-rich diatom *Thalassiosira pseudonana* indicates widespread intron gain. Mol Biol Evol. 24:1447–1457.

Rushforth AM, Anderson P. 1996. Splicing removes the *Caenorhabditis elegans* transposon Tc1 from most mutant pre-mRNAs. Mol Cell Biol. 16:422–429.

Sadusky T, Newman AJ, Dibb NJ. 2004. Exon junction sequences as cryptic splice sites: implications for intron origin. Curr Biol. 14:505–509.

Scofield DG, Hong X, Lynch M. 2007. Position of the final intron in full-length transcripts: determined by NMD? Mol Biol Evol. 24:896–899.

Sela N, Mersch B, Gal-Mark N, Lev-Maor G, Hotz-Wagenblatt A, Ast G. 2007. Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. Genome Biol. 8:R127.

Senapathy P. 1986. Origin of eukaryotic introns—a hypothesis, based on codon distribution statistics in genes, and its implications. Proc Natl Acad Sci USA. 83:2133–2137.

Senapathy P. 1988. Possible evolution of splice-junction signals in eukaryotic genes from stop codons. Proc Natl Acad Sci USA. 85:1129–1133.

Sharp PA. 1985. On the origin of Rna splicing and introns. Cell. 42:397–400.

Shukla GC, Padgett RA. 2002. A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome. Mol Cell. 9:1145–1150.

Simpson CG, Brown JW. 1993. Efficient splicing of an AU-rich antisense intron sequence. Plant Mol Biol. 21:205–211.

Simpson GG, Filipowicz W. 1996. Splicing of precursors to mRNA in higher plants: mechanism, regulation and sub-nuclear organisation of the spliceosomal machinery. Plant Mol Biol. 32:1–41.

Stoltzfus A. 1999. On the possibility of constructive neutral evolution. J Mol Evol. 49:169–181.

Stoltzfus A, Spencer DF, Zuker M, Logsdon JM, Doolittle WF. 1995. Introns and the origin of protein-coding genes—reply. Science. 268:1367–1369.

Sverdlov AV, Rogozin IB, Babenko VN, Koonin EV. 2003. Evidence of splice signal migration from exon to intron during intron evolution. Curr Biol. 13:2170–2174.

Sverdlov AV, Rogozin IB, Babenko VN, Koonin EV. 2004. Reconstruction of ancestral protosplice sites. Curr Biol. 14:1505–1508.

Toro N. 2003. Bacteria and Archaea Group II introns: additional mobile genetic elements in the environment. Environ Microbiol. 5:143–151.

Tseng CK, Cheng SC. 2008. Both catalytic steps of nuclear pre-mRNA splicing are reversible. Science. 320:1782–1784.

Valadkhan S, Manley JL. 2002. Intrinsic metal binding by a spliceosomal RNA. Nat Struct Biol. 9:498–499.

Wessler SR. 1989. The splicing of maize transposable elements from pre-mRNA – a minireview. Gene. 82:127–133.

Xu L, Chen H, Hu X, Zhang R, Zhang Z, Luo ZW. 2006. Average gene length is highly conserved in prokaryotes and eukaryotes and diverges only between the two kingdoms. Mol Biol Evol. 23:1107–1108.

Yandell M, Mungall CJ, Smith C, Prochnik S, Kaminker J, Hartzell G, Lewis S, Rubin GM. 2006. Large-scale trends in the evolution of gene structures within 11 animal genomes. PLoS Comput Biol. 2:e15.

Yoshihama M, Nakao A, Nguyen HD, Kenmochi N. 2006. Analysis of ribosomal protein gene structures: implications for intron evolution. PLoS Genet. 2:237–242.

Zhang XHF, Chasin LA. 2006. Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. Proc Natl Acad Sci USA. 103:13427–13432.

Zhuo D, Madden R, Elela SA, Chabot B. 2007. Modern origin of numerous alternatively spliced human introns from tandem arrays. Proc Natl Acad Sci USA. 104:882–886.

Corresponding Editor: Michael Lynch