REVIEW

Intron creation and DNA repair

Hermann Ragg

Abstract The genesis of the exon–intron patterns of eukaryotic genes persists as one of the most enigmatic questions in molecular genetics. In particular, the origin and mechanisms responsible for creation of spliceosomal introns have remained controversial. Now the issue appears to have taken a turn. The formation of novel introns in eukaryotes, including some vertebrate lineages, is not as rare as commonly assumed. Moreover, introns appear to have been gained in parallel at closely spaced sites and even repeatedly at the same position. Based on these discoveries, novel hypotheses of intron creation have been developed. The new concepts posit that DNA repair processes are a major source of intron formation. Here, after summarizing the current views of intron gain mechanisms, I review findings in support of the DNA repair hypothesis that provides a global mechanistic scenario for intron creation. Some implications on our perception of the mosaic structure of eukaryotic genes are also discussed.

Keywords Intron creation · DNA repair · Splicing

Introduction

Interruption of coding regions by spliceosomal introns is a key feature of most eukaryotic genes. Since the first discovery of the mosaic nature of genes [1–3], much progress has been made, particularly concerning biochemistry and molecular genetics of intron excision [4]. Two questions, however, have remained enigmatic. The first of these issues addresses the time and manner of spliceosomal intron emergence. Essentially all critically important components of two classes of spliceosomes have been detected in presumably early diverging eukaryotes, indicating that functional intron excision machineries were already present in the last common ancestor of eukaryotes [5, 6]. Similarities in the catalytic mechanism used for excision argue in favor of the hypothesis that spliceosomal introns arose from self-splicing group II introns (see [7, 8] for review); however, there are alternative views on the origin of spliceosomal introns and the splicing machinery (discussed in [9]). The second issue concerns dynamics and mechanisms underlying intron loss and gain. Many introns appear to have been maintained for hundreds of millions of years, as inferred from the observation that a large fraction of orthologs in vertebrates and some basal metazoans share the majority of intron positions [10, 11]. The discovery of intron-rich genes in supposedly early diverging lineages also seems to confirm that the genomes of some primordial eukaryotes were densely populated with introns. Introns may be lost at varying rates that depend on the evolutionary lineage [12, 13]. This is often explained by reverse transcription of mRNAs or partially spliced transcripts followed by gene conversion, though other processes have been claimed to be involved (reviewed in [7]). Though the mechanisms responsible for intron loss are far from clear, there are many more issues concerning intron gain that will be addressed in this article. Here it is argued that processes associated with repair of DNA double-strand breaks (DSBs) are a major source of intron formation.

H. Ragg (✉)
Department of Biotechnology, Faculty of Technology,
University of Bielefeld, 33501 Bielefeld, Germany
e-mail: hr@zellkult.techfak.uni-bielefeld.de

Recent findings demonstrate de novo intron creation

Intron gain and loss have been studied in various lineages of eukaryotes. Usually, intron creation was found to be considerably less frequent than intron loss, though some reports have advocated for an excess of intron acquisition [14–17]. These conflicting results are partly due to problems and misconceptions associated with identification of "true" intron gains (i.e., inadequate use of outgroups) and partly due to gene- and lineage-specific, real differences in intron gain rates [18]. Several recent reports, however, have now convincingly demonstrated that genes indeed have been colonized by novel introns and that de novo intron formation is still ongoing [19–21]. One of the most clear-cut examples of intron gain has been unearthed in the water flea, *Daphnia pulex* (*D. pulex*). Omilian et al. [22] found that the *rab4* gene of an endemic population of this aquatic crustacean contains two non-standard introns— shifted by one nucleotide in allelic variants—that are not present in any other *D. pulex* population examined. With the exception of one *D. pulex*/*D. pulicaria* hybrid, these introns also do not exist in closely related species of the same genus or in *rab4* homologs of insects, echinoderms, and vertebrates. Aside from brief oligonucleotide stretches, the novel introns neither share sequence similarity to each other nor are there matches to any other region of the water flea genome. As mentioned by the authors, the restricted geographic distribution and the lack of sequence diversity within the affected *D. pulex* population suggest a recent origin for these introns. Subsequent whole-genome analyses revealed multiple novel introns in *Daphnia*, some of which appear to have independently arisen several times at the same site [23].

Creation of novel introns in vertebrates, particularly in mammals, was thought to be very rare or even non-existent. Some recent reports, however, argue that in some lineages of vertebrates at least, de novo intron formation has occurred [24–27]. Remarkably, there are even two examples of genes each depicting two novel introns, possibly created in a concerted event or rapidly one after the other (on an evolutionary time scale) [27]. Examination of orthologs in basal vertebrates (cyclostomes) and/or other outgroups corroborated the evolutionarily young age of all non-standard introns, and in several cases ESTs or cDNA sequences confirmed that the presumptive introns are indeed spliced out. Frequently, the formation of novel introns did not cause insertions or deletions at the insertion point, thus leaving the size of the affected proteins unchanged. In some instances, however, intron creation was associated with gain or loss of a few amino acids at the target site, in accordance with similar observations in other lineages [21, 28]. There is also one report of modern introns in human [29], but the findings have been challenged due to problems potentially associated with the methodology used [30].

Classical models of intron gain

Though there is now a considerable body of evidence demonstrating that spliceosomal introns may arise de novo, the mechanism(s) of their formation has/have remained controversial. Several excellent recent reviews have discussed this topic in depth [7, 31–33], therefore only a brief summary and a critical evaluation of the most influential proposals are presented here (Table 1).

(1) Intron generation via transposon insertion

There are a few spliceosomal intron gains apparently associated with acquisition of mobile elements [34–36]. The rarity of such examples, however, argues against transposons as a major source of introns. A drawback of this mechanism is that spliceosomal excision of the inserted sequences will usually create insertions at the integration site, probably often interfering with the function of the protein.

(2) Reverse transcription of released introns followed by integration elsewhere in the genome

Insertion of reverse-transcribed introns into genes is an attractive proposal, since it would create introns that are perfectly equipped with the attributes required for efficient and "clean" excision. The reverse-transcription based mechanism implies the presence of parental intron sequences elsewhere in the genome or in the genome of close relatives. By analyzing intron presence/absence polymorphisms in *Caenorhabditis*, Coghlan and Wolfe detected several putative novel introns with sequence similarity to other introns in the same genes [37]. Intron loss or repeated insertions of transposons, however, may also explain the data [18]. Large-scale analyses of several genomes provided no evidence in support of reverse-

Table 1 Comparison of proposed intron gain mechanisms

Proposed mechanism of intron gain	Maintenance of protein size explained	Flanking repeats explained
DSB repair	Yes	Yes
Exon intronization	No (size loss)	No
Alu exonization	No (size increase)	No[a]
Reverse splicing	Yes	No
Transposon insertion	?	Yes

[a] Complete *Alu* repeats are usually flanked by direct repeats

transcriptase based intron transfer to new locations [38]. Population-genomic sequencing projects will probably offer the best way to evaluate the significance of this mechanism.

(3) Tandem duplication of AGGT tetramers in coding sequences

Expansion of directly repeated AGGT sequences with their embedded splice sites might lead to intron formation, as originally proposed by Rogers [39]. Currently, there is only a single example possibly in accord with this mechanism [40].

(4) Intronization of exonic sequences

Studying five *Caenorhabditis* species, Irimia et al. [32] compiled 16 introns evidently ensuing from intronization of exon sequences. In many of these cases, novel splice boundaries were generated by a single base pair change. RT-PCR analyses showed that spliced and original transcript forms are often co-expressed; the potentially deleterious effects originating from silencing of exon segments thus are mitigated. The investigation of Irimia et al. [32] clearly documents that simple point mutations can elicit novel introns in *Caenorhabditis*, this model, however, does not seem to be common in other lineages [41]. Since inherently associated with loss of amino acid stretches, this model also cannot explain formation of introns that leave the protein size unaltered (i.e., intron formation in gene segments that code for protein domains of low structural plasticity is problematic).

Another mechanism that is related to the exon intronization model also suffers from the fact that size changes of the encoded proteins ensue. Duplication of intragenic segments (encompassing intronic and flanking exon sequences) may lead to activation of latent splice sites [42], thus creating novel introns or exons (that are flanked by novel introns). Dissenting from the Rogers [39] model, here the splice signals of novel introns reside in dispersed, complex repeats rather than in expanded AGGT repetitions. Varying with the species, up to 30% of all genes with internally duplicated segments were found to contain novel introns, demonstrating the significance of this mechanism.

(5) Exonization of intron sequences

Exonization of complex repetitive elements residing in introns represents another facet of the 'fortuitous splice site creation' model [43, 44]. *Alu* repeats are a group of primate-specific retroposons of about 300 nucleotides in size, typically containing a poly (A)-stretch at their 3'-ends. When present in the antisense orientation, the poly(T)-sequence can serve as a splice-promoting polypyrimidine

tract that, together with point mutations producing AG/GT splice sites, may lead to a partially exonized *Alu* repeat preceded by a novel intron. In at least one case, RNA editing processes have been shown to contribute to *Alu* exonization [45]. Exonization of *Alu* repeats (and other repetitive elements) certainly represents one of the best supported models of intron formation with thousands of documented cases [43, 46]. Exonization of these repeats, however, will lead to sequence additions with potentially deleterious consequences for the affected protein. Retention of the ancestral transcript, brought about by alternative splicing or backup copies (paralogs), can dampen such effects. In summary, numerous examples demonstrate that appropriate sequence modifications in *Alu* elements can elicit formation of novel introns in primates, though at the risk of losing the ancestral protein.

Intron gain and DNA repair

Though each of the mechanisms discussed above can explain formation of introns, they either cannot easily rationalize all types of introns (Table 1) or the evidence in support of them is scarce. Recently, two groups suggested an alternative, arguing that DSBs might trigger formation of spliceosomal introns [23, 27]. Investigating intron dynamics in a superfamily of protease inhibitor genes, one of these groups found that all novel introns identified in the vertebrate lineage occurred in a fish clade that had undergone genome compaction. Since genome size reduction may be caused solely through loss of entire chromosomes and/or deletion of intrachromosomal DNA segments following creation of DSBs, the authors suspected that conditions favoring DSBs and the processes involved in their repair could foster intron formation [27]. In line with this proposal are other reports of intron gains associated with genome contraction. *Arabidopsis thaliana,* for instance, experienced extensive formation of novel introns. This well-characterized model plant also underwent DNA loss following at least one whole-genome duplication in the *Arabidopsis* lineage [19]. Intron gains were also detected in genes of bacterial origin (assumed to be intron-free) horizontally transferred to bdelloid rotifers [47]. These originally tetraploid invertebrates are in the process of losing genome segments and are extraordinarily resistant to desiccation and radiation-induced DSBs, probably due to efficient DNA repair [48–50]. Likewise, *Oikopleura dioica*, a tunicate, features a compressed genome densely populated with genes, many of which depict non-conserved intron patterns [17, 51].

Analyzing intron presence/absence polymorphisms on a genomic scale, Lynch's group found that a substantial fraction (43%) of novel introns in *Daphnia* are flanked by

short direct repeats, ranging from 5 to 12 base pairs in size [23]. These observations led the authors to suggest that intron gains in *Daphnia* resulted from the repair of DSBs accompanied by small insertions, since short DNA insertions flanked by short direct repeats are frequently observed in a subset of repaired DSBs. Direct repeats overlapping with the splice sites are also associated with several novel introns in *Drosophila* [21], suggesting that such sequences represent phylogenetically widespread attributes in a part of novel introns. Collectively, these findings are compatible with the notion that conditions that favor DSB formation and genome instability, such as replication fork collapse, genome compaction, desiccation, ionizing radiation, or cleavage by retrotranspositional endonucleases [52], can pave the way for intron formation.

Molecular basis of the DNA repair hypothesis

Genetic and genomic investigations have accumulated evidence implying DSB formation and DNA damage repair in intron formation. But is there a solid biochemical basis for the DNA repair hypothesis? In the following sections, I will discuss data and findings that suggest that DNA damage and the processes involved in DNA rejoining may, indeed, represent a global source of intron generation.

Double-strand breaks are hazardous genomic lesions that occur at a high rate in eukaryotes, even in the absence of ionizing radiation or genotoxic chemicals [53, 54]. Human cells, for instance, have been estimated to suffer ∼10 DSBs every time a cell divides [55]. There are two major groups of networked DSB repair pathways, homologous recombination (HR) and non-homologous end-joining (NHEJ) [56, 57]. In multicellular eukaryotes, NHEJ usually dominates [54], but the relative importance of these pathways may vary with the species considered. HR typically results in error-free correction of the lesion, however, when repetitive DNA segments are affected, sequence insertions (and deletions) also likely arise [58–60]. It is conceivable that, via non-allelic HR, expansion of repetitive sequences enclosing cryptic splice signals could lead to novel introns, though as yet there is no evidence for this.

Non-homologous end-joining enables joining of broken DNA molecules in the absence of a homologous template, though short sequence homologies between the ends may promote repair. Following Mahaney et al. [61], NHEJ may be graded into three central steps that include DSB sensing (detection of DSBs and tethering/protection of DNA ends), DNA end processing, and DNA ligation. Key players of DSB sensing in vertebrates are Ku70/80 (a heterodimer that binds to DSB termini and recruits other proteins of the NHEJ complex) and DNA-dependent protein kinase C (DNA-PKc). DNA end processing may include removal of damaged or non-ligatable groups, polishing of ends, and addition of nucleotides. An important enzyme of this step is the Artemis nuclease, however, additional proteins, such as DNA-PKc and DNA polymerases λ or μ (pol λ, pol μ) may be involved [62]. Joining of the processed ends is executed by the XRCC4/DNA ligase IV complex.

Due to the variability in composition and order in which the components of the NHEJ complex may act, the final outcome of the repair process can vary [63]. Thus, deletions, exact rejoining, or the addition of a few nucleotides may ensue from these activities. Some NHEJ events, however, involve capture of filler DNA leading to large insertions [64–67]. The relative frequencies of these events depend on the nature of DNA damage, the taxonomic position of the species concerned, and in experimental settings, the conditions used [64, 65, 68]. NHEJ-mediated insertion events also vary with respect to junction borders and sequence complexity of the DNA integrated. Frequently, acquisition of filler DNA is associated with resection of DSB termini that can involve donor and/or genomic DNA, and the integrated DNA sequences are often flanked by short direct or inverted repeats [63, 69].

For repair of DSBs, DNA of various origins may be used as filling material. Aside from numerous unique sequences, endogenous telomeric repeats and organellar DNA fragments are frequently used as patching material [70, 71]. Various experimental studies unearthed a potpourri of DNA sequences as inserts. Integrated sequences include simple repeats, such as $(GT)_n$ [64], $G(T)_{20}$ [68] or $(GAGAA)_9(AAAGG)_3(AAGGG)_3$ [66], fragments of expression vectors (applied to generate chromosomal DSBs), genomic DNA from bacteria (presumably contaminants of transfection vectors), deliberately added sequences, and DNA of unknown origin [66, 67]. Thus, it appears that DNA of any source, when available, can be recruited for NHEJ. Collectively, these findings suggest that NHEJ-mediated repair can explain several traits of novel introns, such as short (and large) intron size and the presence of short flanking repeats [21, 23].

Integration of *numts*: a paradigm of intron formation via DNA fragment insertion?

Integration of *numts* (nuclear DNA of mitochondrial origin) into the genome depicts numerous features that appear to mirror segment-mediated de novo intron formation. In their detailed analysis of *numt*-mediated DSB repair in primates, Hazkani-Covo and Covo [72] detected many types of fusions between organellar and nuclear DNA, two of their observations, however, are of particular interest. Firstly, more than 50% of *numt* insertions (49 of 90 events) did not depict any deletion of chromosomal DNA. The

authors suppose that, in the presence of filler DNA, end-processing repair pathways are competed out. The second important finding was that target site duplications of up to 11 nucleotides were observed in a minority of ligations, in accordance with some intron gains observed in *Daphnia* [23] and *Drosophila* [21]. In line with these observations, Li et al. [23] found that one of the novel introns found in the genome of water fleas was homologous to mitochondrial DNA.

Intron formation by DNA synthesis?

Though it appears that en bloc transfer of DNA segments of organellar, nuclear or any other origin to DSBs can provide the raw material for new introns, alternative ways of intron formation might exist. One such alternative is repair-mediated synthesis. Pol μ, a widely expressed member of X family DNA polymerases, has been shown to perform non-templated DNA synthesis, both in vitro (>40 nucleotides in the case of T addition [73]), and in vivo [74–76]. De novo synthesis thus might explain the lack of sequence similarity of novel introns to other regions in the genome, a characteristic feature of most novel introns. There are, however, some reservations challenging the role of this and related enzymes as exclusive, major actors in intron formation. Firstly, it is not clear, whether, in vivo, pol μ can synthesize non-instructed DNA segments that are long enough to function as introns (about 20–25 bases). Secondly, some eukaryotes that acquired novel introns, such as *D. melanogaster* or *Caenorhabditis*, do not harbor X family DNA polymerase genes [77]. Possibly, in these and other cases as yet little characterized alternative end-joining pathways might come into play [78].

Repair-mediated DNA insertion, splicing, and nonsense-mediated decay surveillance

Clearly, most of the repair-mediated DNA insertions will have deleterious consequences, when interlaced in coding regions (or other essential gene sequences). However, these harmful effects may be counteracted, on the RNA level, by two mechanisms: splicing and nonsense-mediated decay (NMD). Splicing involves signals that are either degenerate [79, 80] (standard splice sites: GU/AG) or fuzzy (C/U-rich region, splice enhancers, secondary structures); it is thus conceivable that, by chance, some of the novel sequences will accommodate the requirements imposed by the splice apparatus. Some of the insert-colonized genes will give rise to proteins of unaltered size, if the spliceosome can remove the integrated sequence entirely. Inexact excision, however, will produce insertions or deletions at the target site, depending on the positions of the splice sites actually used.

Nonsense-mediated decay is a powerful control system that, in the pioneering round of translation, mediates destruction of transcripts with premature stop codons (PTCs) [81], thereby preventing accumulation of aberrant proteins that may be toxic to cells. PTCs emerging from insertions are no exceptions, and consequently the corresponding transcripts are considered flawed and led to degradation. Removal of PTCs by splicing, in contrast, results in transcript accumulation. NMD-mediated suppression of the harmful consequences thus may greatly favor emergence of introns in sequences that are inefficiently spliced [21, 82]. Subsequent mutations affecting initially suboptimal splice signals may finally result in intron fixation. Recent investigations have indeed shown that novel introns in *Drosophila* are more likely to contain in-frame PTCs than conserved introns, indicating that NMD may become active in the case of splicing failure, thus enabling selection of better splice motifs [21].

Implications of the DNA repair hypothesis on our view of the mosaic nature of eukaryotic genes

The new developments in one of the most enigmatic fields of genetics may also change our view of splicing. Whatever the very origin of spliceosomal introns, it is now conceivable that eukaryotes have channeled splicing and the splice apparatus into a device that corrects some of the genomic scars left by the error-prone NHEJ repair. Splicing thus may serve as another gene surveillance and rescue system that acts on the RNA level. Though it is now evident that introns continue to emerge—in some lineages at unprecedented rate—other queries are still unresolved. For instance, it is not clear how the common attributes of introns, such as 5′ and 3′ splice signals, arose. Are they related to sequence preferences of the DSB repair machinery or are they relics of the very first introns?

The DNA repair model of intron formation is compatible with the existence of hot spots for intron gain [23]. The DNA and/or chromatin structures that favor "hits", however, remain to be elucidated. Repeated intron formation at identical positions also challenges that introns at concordant sites in orthologs of phylogenetically distant organisms are invariably offspring of a common ancestor. The new model of intron formation also addresses the question of time and mechanism(s) of primordial spliceosomal intron creation. NHEJ is an evolutionarily ancient tool present in various lineages of prokaryotes [83], whether this has any relevance for the emergence of the first spliceosomal introns remains to be investigated.

Outlook

Recent findings have demonstrated that spliceosomal introns have arisen de novo at appreciable rates in various lineages of eukaryotes and that introns still continue to emerge. As a novel mechanism of intron formation, DSB repair comes into focus. Evidence from genomic data in favor of this model is accumulating, experimental confirmation, however, is still lacking. The genetic and biochemical tools already available will undoubtedly be used to scrutinize the significance of this mechanism for intron formation. Next-generation sequencing technologies will provide a realistic picture of intron formation rates across various clades and within populations. The occurrence of hot spots of intron insertion still represents an enigma that awaits clarification.

References

1. Berget SM, Moore C, Sharp PA (1977) Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci USA 74:3171–3175
2. Chow LT, Gelinas RE, Broker TR, Roberts RJ (1977) An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. Cell 12:1–8
3. Evans RM, Fraser N, Ziff E, Weber J, Wilson M, Darnell JE (1977) The initiation sites for RNA transcription in Ad2 DNA. Cell 12:733–739
4. Wahl MC, Will CL, Lührmann R (2009) The spliceosome: design principles of a dynamic RNP machine. Cell 36:701–718
5. Collins L, Penny D (2005) Complex spliceosomal organization ancestral to extant eukaryotes. Mol Biol Evol 22:1053–1066
6. Russell AG, Charette JM, Spencer DF, Gray MW (2006) An early evolutionary origin for the minor spliceosome. Nature 443:863–866
7. Roy SW, Gilbert W (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. Nat Rev Genet 7:211–221
8. Koonin EV (2006) The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? Biol Direct 1:22
9. Penny D, Hoeppner MP, Poole AM, Jeffares DC (2009) An overview of the introns-first theory. J Mol Evol 69:527–540
10. Raible F, Tessmar-Raible K, Osoegawa K, Wincker P, Jubin C, Balavoine G, Ferrier D, Benes V, de Jong P, Weissenbach J, Bork P, Arendt D (2005) Vertebrate-type intron-rich genes in the marine annelid *Platynereis dumerilii*. Science 310:1325–1326
11. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, Jurka J, Genikhovich G, Grigoriev IV, Lucas SM, Steele RE, Finnerty JR, Technau U, Martindale MQ, Rokhsar DS (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317:86–94
12. Jeffares DC, Mourier T, Penny D (2006) The biology of intron gain and loss. Trends Genet 22:16–22
13. Carmel L, Wolf YI, Rogozin IB, Koonin EV (2007) Three distinct modes of intron dynamics in the evolution of eukaryotes. Genome Res 17:1034–1044
14. Logsdon JM Jr, Tyshenko MG, Dixon C, D-Jafari J, Walker VK, Palmer JD (1995) Seven newly discovered intron positions in the triose-phosphate isomerase gene: evidence for the introns-late theory. Proc Natl Acad Sci USA 92:8507–8511
15. Ragg H, Lokot T, Kamp PB, Atchley WR, Dress A (2001) Vertebrate serpins: construction of a conflict-free phylogeny by combining exon–intron and diagnostic site analyses. Mol Biol Evol 18:577–584
16. Qiu WG, Schisler N, Stoltzfus A (2004) The evolutionary gain of spliceosomal introns: sequence and phase preferences. Mol Biol Evol 21:1252–1263
17. Edvardsen RB, Lerat E, Maeland AD, Flåt M, Tewari R, Jensen MF, Lehrach H, Reinhardt R, Seo HC, Chourrout D (2004) Hypervariable and highly divergent intron–exon organizations in the chordate *Oikopleura dioica*. J Mol Evol 59:448–457
18. Roy SW, Penny D (2006) Smoke without fire: most reported cases of intron gain in nematodes instead reflect intron losses. Mol Biol Evol 23:2259–2262
19. Knowles DG, McLysaght A (2006) High rate of recent intron gain and loss in simultaneously duplicated *Arabidopsis* genes. Mol Biol Evol 23:1548–1557
20. 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
21. Farlow A, Meduri E, Dolezal M, Hua L, Schlötterer C (2010) Nonsense-mediated decay enables intron gain in *Drosophila*. PLoS Genet 6:e1000819
22. Omilian AR, Scofield DG, Lynch M (2008) Intron presence–absence polymorphisms in *Daphnia*. Mol Biol Evol 25:2129–2139
23. Li WL, Tucker AE, Sung W, Thomas WK, Lynch M (2009) Extensive, recent intron gains in *Daphnia* populations. Science 326:1260–1262
24. Schiöth HB, Haitina T, Fridmanis D, Klovins J (2005) Unusual genomic structure: melanocortin receptors in *Fugu*. Ann NY Acad Sci 1040:460–463
25. Moriyama S, Oda M, Yamazaki T, Yamaguchi K, Amiya N, Takahashi A, Amano M, Goto T, Nozaki M, Meguro H, Kawauchi H (2008) Gene structure and functional characterization of growth hormone in dogfish, *Squalus acanthias*. Zool Sci 25:604–613
26. Hussain A, Saraiva LR, Korsching SI (2009) Positive Darwinian selection and the birth of an olfactory receptor clade in teleosts. Proc Natl Acad Sci USA 106:4313–4318
27. Ragg H, Kumar A, Köster K, Bentele C, Wang Y, Frese MA, Prib N, Krüger O (2009) Multiple gains of spliceosomal introns in a superfamily of vertebrate protease inhibitor genes. BMC Evol Biol 9:208
28. Coulombe-Huntington J, Majewski J (2007) Intron loss and gain in *Drosophila*. Mol Biol Evol 24:2842–2850
29. 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
30. Roy SW, Irimia M (2008) When good transcripts go bad: artifactual RT-PCR 'splicing' and genome analysis. Bioessays 30:601–605 (comment in: Bioessays 30:1256; author reply 1257–1258)
31. Rodríguez-Trelles F, Tarrío R, Ayala FJ (2006) Origins and evolution of spliceosomal introns. Annu Rev Genet 40:47–76
32. 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
33. Roy SW, Irimia M (2009) Mystery of intron gain: new data and new models. Trends Genet 25:67–73
34. 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
35. 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

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

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

38. Fedorov A, Roy S, Fedorova L, Gilbert W (2003) Mystery of intron gain. Genome Res 13:2236–2241

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

40. Figueroa F, Ono H, Tichy H, O'Huigin C, Klein J (1995) Evidence for insertion of a new intron into an Mhc gene of perch-like fish. Proc Biol Sci 259:325–330

41. Roy SW (2009) Intronization, de-intronization and intron sliding are rare in *Cryptococcus*. BMC Evol Biol 9:192

42. Gao X, Lynch M (2009) Ubiquitous internal gene duplication and intron creation in eukaryotes. Proc Natl Acad Sci USA 106:20818–20823

43. Krull M, Brosius J, Schmitz J (2005) Alu-SINE exonization: en route to protein-coding function. Mol Biol Evol 22:1702–1711

44. Schmidt EE, Davies CJ (2007) The origins of polypeptide domains. Bioessays 29:262–270

45. Lev-Maor G, Sorek R, Levanon EY, Paz N, Eisenberg E, Ast G (2007) RNA-editing-mediated exon evolution. Genome Biol 8:R29

46. Sorek R (2007) The birth of new exons: mechanisms and evolutionary consequences. RNA 13:1603–1608

47. Gladyshev EA, Meselson M, Arkhipova IR (2008) Massive horizontal gene transfer in bdelloid rotifers. Science 320:1210–1213

48. Gladyshev E, Meselson M (2008) Extreme resistance of bdelloid rotifers to ionizing radiation. Proc Natl Acad Sci USA 105:5139–5144

49. Mark Welch DB, Mark Welch JL, Meselson M (2008) Evidence for degenerate tetraploidy in bdelloid rotifers. Proc Natl Acad Sci USA 105:5145–5149

50. Hur JH, Van Doninck K, Mandigo ML, Meselson M (2009) Degenerate tetraploidy was established before bdelloid rotifer families diverged. Mol Biol Evol 26:375–383

51. Seo HC, Kube M, Edvardsen RB, Jensen MF, Beck A, Spriet E, Gorsky G, Thompson EM, Lehrach H, Reinhardt R, Chourrout D (2001) Miniature genome in the marine chordate *Oikopleura dioica*. Science 294:2506 [Erratum in: Science 295:45 (2002)]

52. Gasior SL, Wakeman TP, Xu B, Deininger PL (2006) The human LINE-1 retrotransposon creates DNA double-strand breaks. J Mol Biol 357:1383–1393

53. Vilenchik MM, Knudson AG (2003) Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. Proc Natl Acad Sci USA 100:12866–12871

54. Lieber MR, Ma Y, Pannicke U, Schwarz K (2003) Mechanism and regulation of human non-homologous DNA end-joining. Nat Rev Mol Cell Biol 4:712–720

55. Haber JE (1999) DNA recombination: the replication connection. Trends Biochem Sci 24:271–275

56. Pardo B, Gómez-González B, Aguilera A (2009) DNA repair in mammalian cells: DNA double-strand break repair: how to fix a broken relationship. Cell Mol Life Sci 66:1039–1056

57. Hartlerode AJ, Scully R (2009) Mechanisms of double-strand break repair in somatic mammalian cells. Biochem J 423:157–168

58. Wells RD, Dere R, Hebert ML, Napierala M, Son LS (2005) Advances in mechanisms of genetic instability related to hereditary neurological diseases. Nucleic Acids Res 33:3785–3798

59. Shishkin AA, Voineagu I, Matera R, Cherng N, Chernet BT, Krasilnikova MM, Narayanan V, Lobachev KS, Mirkin SM (2009) Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. Mol Cell 35:82–92

60. Sasaki M, Lange J, Keeney S (2010) Genome destabilization by homologous recombination in the germ line. Nat Rev Mol Cell Biol 11:182–195

61. Mahaney BL, Meek K, Lees-Miller SP (2009) Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. Biochem J 417:639–650

62. Lieber MR, Lu H, Gu J, Schwarz K (2008) Flexibility in the order of action and in the enzymology of the nuclease, polymerases, and ligase of vertebrate non-homologous DNA end joining: relevance to cancer, aging, and the immune system. Cell Res 18:125–133

63. Lieber MR (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu Rev Biochem 79:181–211

64. Liang F, Han M, Romanienko PJ, Jasin M (1998) Homology-directed repair is a major double-strand break repair pathway in mammalian cells. Proc Natl Acad Sci USA 95:5172–5177

65. Haviv-Chesner A, Kobayashi Y, Gabriel A, Kupiec M (2007) Capture of linear fragments at a double-strand break in yeast. Nucleic Acids Res 35:5192–5202

66. Lin Y, Waldman AS (2001) Capture of DNA sequences at double-strand breaks in mammalian chromosomes. Genetics 158:1665–1674

67. Lin Y, Waldman AS (2001) Promiscuous patching of broken chromosomes in mammalian cells with extrachromosomal DNA. Nucleic Acids Res 29:3975–3981

68. Odersky A, Panyutin IV, Panyutin IG, Schunck C, Feldmann E, Goedecke W, Neumann RD, Obe G, Pfeiffer P (2002) Repair of sequence-specific ^{125}I-induced double-strand breaks by nonhomologous DNA end joining in mammalian cell-free extracts. J Biol Chem 277:11756–11764

69. Jensen-Seaman MI, Wildschutte JH, Soto-Calderón ID, Anthony NM (2009) A comparative approach shows differences in patterns of numt insertion during hominoid evolution. J Mol Evol 68:688–699

70. Ruiz-Herrera A, Nergadze SG, Santagostino M, Giulotto E (2008) Telomeric repeats far from the ends: mechanisms of origin and role in evolution. Cytogenet Genome Res 122:219–228

71. Leister D (2005) Origin, evolution and genetic effects of nuclear insertions of organelle DNA. Trends Genet 21:655–663

72. Hazkani-Covo E, Covo S (2008) Numt-mediated double-strand break repair mitigates deletions during primate genome evolution. PLoS Genet 4:e1000237

73. Gu J, Lu H, Tippin B, Shimazaki N, Goodman MF, Lieber MR (2007) XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. EMBO J 26:1010–1023 (Erratum in: EMBO J 26:3506–3507)

74. Moon AF, Garcia-Diaz M, Bebenek K, Davis BJ, Zhong X, Ramsden DA, Kunkel TA, Pedersen LC (2007) Structural insight into the substrate specificity of DNA polymerase μ. Nat Struct Mol Biol 14:45–53

75. Gozalbo-López B, Andrade P, Terrados G, de Andrés B, Serrano N, Cortegano I, Palacios B, Bernad A, Blanco L, Marcos MA, Gaspar ML (2009) A role for DNA polymerase μ in the emerging DJH rearrangements of the postgastrulation mouse embryo. Mol Cell Biol 29:1266–1275

76. Andrade P, Martín MJ, Juárez R, López de Saro F, Blanco L (2009) Limited terminal transferase in human DNA polymerase μ defines the required balance between accuracy and efficiency in NHEJ. Proc Natl Acad Sci USA 106:16203–16208

77. Burgers PM, Koonin EV, Bruford E, Blanco L, Burtis KC, Christman MF, Copeland WC, Friedberg EC, Hanaoka F, Hinkle DC, Lawrence CW, Nakanishi M, Ohmori H, Prakash L, Prakash S, Reynaud CA, Sugino A, Todo T, Wang Z, Weill JC, Woodgate

R (2001) Eukaryotic DNA polymerases: proposal for a revised nomenclature. J Biol Chem 276:43487–43490

78. Yu AM, McVey M (2010) Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. Nucleic Acids Res. doi:10.1093/nar/gkq379

79. Sheth N, Roca X, Hastings ML, Roeder T, Krainer AR, Sachidanandam R (2006) Comprehensive splice-site analysis using comparative genomics. Nucleic Acids Res 34:3955–3967

80. Lin CF, Mount SM, Jarmolowski A, Makalowski W (2010) Evolutionary dynamics of U12-type spliceosomal introns. BMC Evol Biol 10:47

81. Stalder L, Mühlemann O (2008) The meaning of nonsense. Trends Cell Biol 18:315–321

82. Catania F, Lynch M (2008) Where do introns come from? PLoS Biol 6:e283

83. Bowater R, Doherty AJ (2006) Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. PLoS Genet 2:e8