BIOLOGY DIRECT

REVIEW Open Access

Identifying the mechanisms of intron gain: progress and trends

Paul Yenerall[1] and Leming Zhou[2,3]*

Abstract: Continued improvements in Next-Generation DNA/RNA sequencing coupled with advances in gene annotation have provided researchers access to a plethora of annotated genomes. Subsequent analyses of orthologous gene structures have identified numerous intron gain and loss events that have occurred both recently and in the very distant past. This research has afforded exceptional insight into the temporal and lineage-specific rates of intron gain and loss among various species throughout evolution. Numerous studies have also attempted to identify the molecular mechanisms of intron gain and loss. However, even after considerable effort, very little is known about these processes. In particular, the mechanism(s) of intron gain have proven exceptionally enigmatic and remain topics of considerable debate. Currently, there exists no definitive consensus as to what mechanism(s) may generate introns. Because many introns are known to affect gene expression, it is necessary to understand the molecular process(es) by which introns may be gained. Here we review the seven most commonly purported mechanisms of intron gain and, when possible, summarize molecular evidence for or against the occurrence of each of these mechanisms. Furthermore, we catalogue indirect evidence that supports the occurrence of each mechanism. Finally, because these proposed mechanisms fail to explain the mechanistic origin of many recently gained introns, we also look at trends that may aid researchers in identifying other potential mechanism(s) of intron gain.

Reviewers: This article was reviewed by Eugene Koonin, Scott Roy (nominated by W. Ford Doolittle), and John Logsdon.

Keywords: Intron, Intron gain, Intron evolution, Gene structure, Evolution, Mechanism

Background

Spliceosomal introns are segments of RNA that are excised by the spliceosome during the processing of pre-mRNA in eukaryotes. Although spliceosomal intron density varies widely among eukaryotes, no true eukaryote has ever been identified without a spliceosomal intron and some remnant of the spliceosome. Conversely, units of the spliceosome and/or spliceosomal introns have never been identified in any prokaryote [1,2]. Spliceosomal introns (herein referred to simply as introns) were originally believed to be "junk" DNA as they were not translated. However, since their initial discovery, numerous functional roles for introns have been elucidated, such as augmenting proteome diversity by enabling alternative splicing [3], enhancing gene expression [4-6] and harboring various *cis* and *trans* regulatory elements [7,8].

As researchers began to compare the structure of orthologous genes, it quickly became apparent that introns may be gained or lost throughout evolution [9]. Subsequent analyses have identified thousands of intron gains and losses [10-29]. These findings have sparked considerable interest into identifying the molecular mechanisms of intron gain and loss [11-15,18,20-24,27,30-35]. Two definitive mechanisms of intron loss, *Reverse Transcriptase-Mediated Intron Loss* (RTMIL) and genomic deletions, have been identified and are widely accepted [36,37]. The definitive mechanism(s) of intron gain, however, remain elusive and controversial. All together, there have been at least seven commonly purported mechanisms of intron gain: *Intron Transposition* [38], *Transposon Insertion* [39], *Tandem Genomic Duplication* [40], *Intron Transfer* [31], *Intron Gain during Double-Strand Break Repair (DSBR)* [27], *Insertion of a Group II Intron* [38] and *Intronization* [41,42].

* Correspondence: lmzhou@gmail.com
[2] Department of Health Information Management, University of Pittsburgh, Pittsburgh, PA 15260, USA
[3] Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA 15260, USA
Full list of author information is available at the end of the article

Although seven possible mechanisms of intron gain have been proposed, researchers have identified thousands of novel introns whose mechanistic origins defy these explanations [11,17,18,22,23,27,28,43]. Surprisingly, even many recently gained introns, which have the highest probability of revealing their mechanistic origin, do not appear to have arisen via any of these mechanisms [27,44]. These findings raise an intriguing question: do these proposed mechanisms of intron gain fail to describe the mechanistic origin of so many novel introns because they are not genuine mechanisms of intron gain, or are there other process(es) generating novel introns? Here we review the proposed mechanisms of intron gain and summarize any previously identified direct (molecular) and/or indirect (intron gains identified during genomic analyses with purportedly known mechanistic origins) evidence that supports or refutes the occurrence of each of these proposed mechanisms. Furthermore, we examine trends that may aid researchers in identifying other novel mechanism(s) of intron gain.

Review

The proposed mechanisms of intron gain

Intron transposition

The most commonly purported mechanism of intron gain is intron transposition [34]. Intron transposition has been proposed to occur when a spliced intron reverse splices into either its own mRNA or another mRNA at a previously intron-less position. This intron-containing mRNA is then reverse transcribed and the resulting intron-containing cDNA may then cause intron gain via three different routes: it may undergo “complete” or nearly complete recombination with its original genomic locus, in which recombination between the intron-containing cDNA and the original genomic locus occurs both in exonic and intronic regions (if the gene contains introns), thereby deleting any introns within the region of recombination and causing concurrent intron gain and loss; it may undergo “partial" recombination, in which recombination only occurs between the intron-containing cDNA and an exonic region, avoiding the deletion of introns; or it may retropose into the genome at a different locus and then transfer the gained intron via “partial” or “complete” recombination to the original locus. Partial recombination between the intron-containing cDNA and the original genomic locus is the most commonly discussed route of intron transposition and is illustrated in Figure 1a.

Intron transposition has frequently been favored among the seven proposed mechanisms of intron gain by researchers [34,37], likely because a transposed intron would already harbor the donor, acceptor, and branch point splicing sequences necessary for splicing [49]. However, the molecular process(es) by which intron transposition may occur, if it occurs at all, are unknown. To an extent, the most crucial and nebulous step in this process, reverse splicing, has been shown to occur [50]. However, this has only been demonstrated under non-physiological conditions using a mutant of Prp22 that fails to release the mRNA from the spliceosome. Furthermore, because a mutant was used that failed to release the mRNA, this process has only been shown to reverse splice introns back into their original position [50]. The process by which a previously spliced intron and spliceosome may bind (or remain bound), recognize an mRNA (based upon sequence motifs, protein complexes loaded onto the mRNA or random interactions), and begin nucleophilic attack and subsequent reverse splicing, is not known. Given that a recent study identified 95 introns with homologous sequences, of which only 11-24% may be novel introns [44], it is possible that interactions between proteins recruited to and deposited on the mRNA during and after splicing [51-53] may recruit spliceosomes “loaded” with previously spliced introns. Thus, reverse splicing may occur frequently, if not entirely, at positions in the mRNA that have already undergone splicing (i.e. only pre-existing intronic positions). This potential preference, if true, may help explain the finding of supposed parallel intron gains [27] if, following an initial intron gain, a different intron is transposed into this intronic position and both intronic sequences are maintained in the species. Alternatively, it is possible that reverse splicing occurs randomly at any position in an mRNA; however, the finding that the majority of suspected reverse splicing events occurred at previously intronic positions [44] argues against this explanation.

Even if intron transposition does occur, it has been suggested that any mechanism of intron gain that relies upon reverse transcriptase (RT) may not be a prevalent mechanism of intron gain. One potential problem is known as the “rate paradox” [34]. This may occur because the presumed most prevalent mechanism of intron loss, RTMIL [35], occurs via a process nearly identical to intron transposition, but does not require reverse splicing. Thus, the difference between the genome-wide rate of intron gain via intron transposition and intron loss via RTMIL should equal the rate of reverse splicing. However, as pointed out by Roy and Irimia, reverse splicing is believed to be an extremely rare process, as no reverse spliced introns have been found in any EST or cDNA sequences [34]. Furthermore, a recent statistical analysis has shown that it is unlikely that RT played a prominent role in intron gain throughout evolution [35].

Other proposed molecular mechanisms of intron gain that mimic intron transposition, such as spliceosomal retrohoming or reverse transcription template switching [34], may occur and may be responsible for a number of intron gains that have been identified with sequences homologous to other introns [25,44]. However, both of these models rely upon RT and therefore suffer from

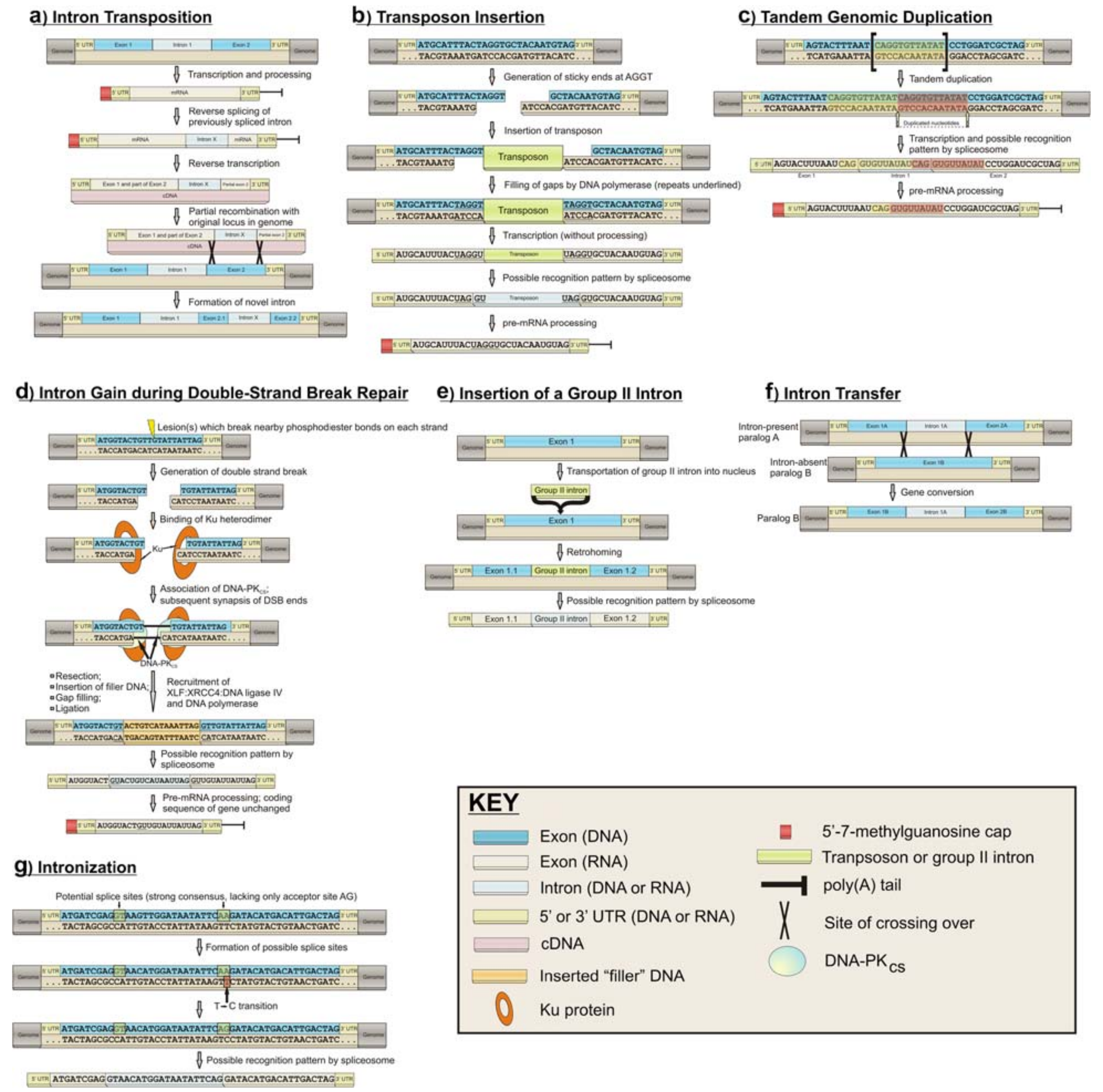
a) Intron Transposition
Genome
5' UTR
Exon 1
Intron 1
Exon 2
3' UTR
Transcription and processing
mRNA
Reverse splicing of previously spliced intron
Intron X
Reverse transcription
Exon 1 and part of Exon 2
Partial exon 2
cDNA
Partial recombination with original locus in genome
Formation of novel intron
Exon 2.1
Exon 2.2
b) Transposon Insertion
ATGCATTTACTAGGTGCTACAATGTAG
...TACGTAAATGATCCACGATGTTACATC...
Generation of sticky ends at AGGT
ATGCATTTACTAGGT
...TACGTAAATG
GCTACAATGTAG
ATCCACGATGTTACATC...
Insertion of transposon
Transposon
Filling of gaps by DNA polymerase (repeats underlined)
...TACGTAAATGATCCA
TAGGTGCTACAATGTAG
Transcription (without processing)
AUGCAUUUACUAGGU
UAGGUGCUACAAUGUAG
Possible recognition pattern by spliceosome
AUGCAUUUACUAG
GU
UAG
GUGCUACAAUGUAG
pre-mRNA processing
AUGCAUUUACUAGGUGCUACAAUGUAG
c) Tandem Genomic Duplication
AGTACTTTAAT
CAGGTGTTATAT
CCTGGATCGCTAG
...TCATGAAATTA
GTCCACAATATA
GGACCTAGCGATC...
Tandem duplication
AGTACTTTAATCAGGTGTTATATCAGGTGTTATATCCTGGATCGCTAG
...TCATGAAATTAGTCCACAATATAGTCCACAATATAGGACCTAGCGATC...
Duplicated nucleotides
Transcription and possible recognition pattern by spliceosome
AGUACUUUAAUCAG
GUGUUAUAUCAG
GUGUUAUAU
CCUGGAUCGCUAG
Exon 1
Intron 1
Exon 2
pre-mRNA processing
AGUACUUUAAUCAGGUGUUAUAUCCUGGAUCGCUAG
d) Intron Gain during Double-Strand Break Repair
Lesion(s) which break nearby phosphodiester bonds on each strand
ATGGTACTGTTGTATTATTAG
....TACCATGACATCATAATAATC....
Generation of double strand break
ATGGTACTGT
....TACCATGA
TGTATTATTAG
CATCCTAATAATC....
Binding of Ku heterodimer
Ku
Association of DNA-PKcs; subsequent synapsis of DSB ends
CATCATAATAATC....
DNA-PKcs
Resection;
Insertion of filler DNA;
Gap filling;
Ligation
Recruitment of XLF:XRCC4:DNA ligase IV and DNA polymerase
ATGGTACTGTACTGTCATAAATTAGGTTGTATTATTAG
...TACCATGACATGACAGTATTTAATCCATCATAATAATC...
Possible recognition pattern by spliceosome
AUGGUACT
GUACUGUCAUAAUUAG
GUUGUAUUAUUAG
Pre-mRNA processing; coding sequence of gene unchanged
AUGGUACTGUUGUAUUAUUAG
e) Insertion of a Group II Intron
Exon 1
Transportation of group II intron into nucleus
Group II intron
Retrohoming
Exon 1.1
Exon 1.2
Possible recognition pattern by spliceosome
f) Intron Transfer
Intron-present paralog A
Exon 1A
Intron 1A
Exon 2A
Intron-absent paralog B
Exon 1B
Gene conversion
Paralog B
Exon 2B
g) Intronization
Potential splice sites (strong consensus, lacking only acceptor site AG)
ATGATCGAGGTAAGTTGGATAATATTCAAGATACATGACATTGACTAG
...TACTAGCGCCATTGTACCTATTATAAGTTCTATGTACTGTAACTGATC...
Formation of possible splice sites
ATGATCGAGGTAACATGGATAATATTCAAGATACATGACATTGACTAG
T→C transition
ATGATCGAGGTAACATGGATAATATTCAGGATACATGACATTGACTAG
...TACTAGCGCCATTGTACCTATTATAAGTCCTATGTACTGTAACTGATC...
Possible recognition pattern by spliceosome
ATGATCGAG
GTAACATGGATAATATTCAG
GATACATGACATTGACTAG
KEY
Exon (DNA)
Exon (RNA)
Intron (DNA or RNA)
5' or 3' UTR (DNA or RNA)
cDNA
Inserted "filler" DNA
Ku protein
5'-7-methylguanosine cap
Tranpsoson or group II intron
poly(A) tail
Site of crossing over
DNA-PKCS

Figure 1 The seven proposed mechanisms of intron gain and loss. Introns shown are shorter in length than necessary for splicing strictly for illustrative purposes. **a**) Intron Transposition with "partial" recombination. Other routes of intron transposition (discussed in text) by which concurrent intron gains/losses may occur can be also envisioned. **b**) Transposon Insertion. Imprecise intron gain may also be envisioned (discussed in text). **c**) Tandem Genomic Duplication using duplicated AGGT sequences for splice sites. The segment to be duplicated is flanked by brackets. The template for the duplicated nucleotides is highlighted in yellow; the duplicated nucleotides are highlighted in red and underlined. Imprecise intron gain may also frequently occur (discussed in text). **d**) Intron Gain during Double-Strand Break Repair. Protein names are from mammals. Short direct repeats created by gap filling are underlined. The possible pathways of non-homologous end joining (NHEJ) are from proposals in [45-48]. **e**) Insertion of a Group II Intron. **f**) Intron Transfer. **g**) Intronization. Strong consensus donor and acceptor sites (following a single point mutation) are highlighted in yellow. The place at which the hypothetical point mutation occurs is highlighted in red.

some of the same problems as the canonical route of intron transposition, as discussed above. Regardless of the exact molecular process, indirect evidence exists which suggests that some form of intron transposition is likely a genuine mechanism of intron gain (Table 1).

Transposon insertion

Transposon insertions into genic regions are generally viewed as deleterious mutations. However, one underappreciated outcome of this process may be intron creation. Such an insertion may completely intronize the

Table 1 Intron gains identified with a purportedly known mechanistic origin

Proposed mechanism	*In vivo* demonstration	Indirect evidence	Number of events unambiguously identified	Specie(s) event identified in
Intron Transposition	No	Yes [25,44]	14	*Mycosphaerella*, *Oikopleura*
Transposon Insertion	No	Yes [23,25,54,55]	35	*Oryza*, *Drosophila*, *Oikopleura*, *Zea*
Tandem Genomic Duplication	Yes [56]	Yes [18,57]	188	*Arabidopsis*, *Oryza*, *Caenorhabditis*, *Drosophila*, mouse, human, *Cryptococcus*
Intron Gain during DSBR	No	Yes [22,23,27,58,59]	5	*Drosophila*, *Daphnia*, *Aspergillus*, *Bigelowiella*, human
Intron Transfer	No	Yes [22,31,44]	3	*Mycosphaerella*, *Aspergillus*, *Chironomus*,
Insertion of a Group II Intron	Likely does not occur [60]	No	0	
Intronization	No	Yes [41,42,61,62]	29	*Cryptococcus*, *Caenorhabditis*

Only intron gains with novel splice sites (i.e. not simple duplications of pre-existing introns) were included. Only novel introns with EST support were selected from [57]. For a more detailed treatment of events, see the MIGL database located at http://cpath.him.pitt.edu/intron/index.php (manuscript in preparation).

transposon without disrupting the coding sequence when a transposon inserts into the sequence AGGT, resulting in the duplication of this sequence on each side of the transposon (illustrated in Figure 1b). Three lines of evidence support the generation of introns via this model. One is that the sequence which the transposon inserts into, AGGT (also known as a protosplice site [63]), is believed to be a site of preferential intron gain [63-66]. Another is that the donor and acceptor splice sites created by this insertion, specifically the donor site AG|GT and the acceptor site AG|G (where "|" specifies splice junctions), adhere to the consensus donor and acceptor splice sites found in many organisms [67-71]. Finally, if these potential donor and acceptor splice sites are utilized efficiently by the spliceosome, any transposon may insert into the sequence AGGT in any gene without altering the genes coding sequence (demonstrated in Figure 1b). Alternatively, the transposon itself may harbor strong donor and acceptor splice sites near its boundaries or activate nearby latent splice sites, enabling its precise, or nearly precise, excision by the spliceosome. Indirect evidence of intron gain via transposon insertion has existed for nearly 20 years (see Table 1). However, what is not understood is exactly why these elements are spliced (if not by pure chance), or if any transposons preferentially cause intron gain, potentially due to target site preferences and/or the sequence of the transposon itself.

Tandem genomic duplication

Although widely underappreciated, the tandem genomic duplication of an exonic segment has recently emerged as a genuine and potentially prevalent mechanism of intron gain. Due to the similarity between consensus donor and acceptor splice sites, both of which closely resemble the sequence AGGT, the tandem genomic duplication of an exonic segment harboring the sequence AGGT generates two strong potential splice sites. If these splice sites are utilized by the spliceosome, the sequence between the original and duplicated AGGT will be spliced, affording the "precise" generation of an intron, i.e. the creation of an intron without alteration of the coding sequence of the gene (illustrated in Figure 1c). As discussed in the Transposon Insertion section above, use of AGGT as both the donor and acceptor splice site conforms to the most common splice site consensus sequences found in a variety of organisms [67-71], and AGGT has been found to be a site of preferential intron gain [64,65]. In contrast to precise intron gain, in which the coding sequence of the gene remains unaltered, tandem genomic duplication may also result in "imprecise" intron gain, in which the coding sequence is altered. This may occur when latent splice sites within the duplicated region are utilized, resulting in the addition of nucleotides to the coding sequence of the gene. Alternatively, the duplication may result in the activation of latent splice sites near the duplicated sequence, resulting in the removal of nucleotides from the coding sequence.

Unlike any other proposed mechanism of intron gain, *in vivo* evidence exists to support the occurrence of this mechanism. The ability of this mechanism to have produced a novel intron nearly 500 million years ago in the ancestor of jawed vertebrates was recently tested and verified *in vivo* [56]. Not only has this process been shown to be a feasible mechanism of intron creation *in vivo*, but a plethora of indirect evidence in support of this mechanism has also been identified in a number of eukaryotes (Table 1). It should be noted, however, that many of these intron gains were imprecise and resulted from the activation of latent splice sites within the duplicated segment [57].

Intron Gain during Double-Strand Break Repair (DSBR)

Double-strand breaks (DSBs) are genomic lesions in which nearby phosphodiester bonds are severed on both strands of the DNA double helix. Such lesions may be caused by ionizing radiation, reactive oxygen species, or cellular processes. Because the repair of these lesions is paramount to cell viability, organisms have evolved two ubiquitous, genetically distinct, well conserved processes to repair these breaks: homologous recombination and non-homologous end joining (NHEJ) [72]. The former results in the precise repair of the break and requires an undamaged template, such as a sister chromatid. Thus, homologous recombination occurs primarily during the S and G2 portions of the cell cycle [73]. NHEJ, on the other hand, requires no template. If the DNA ends are undamaged and complementary, NHEJ faithfully restores the break point junction. However, if the DNA ends are damaged or are not complementary, NHEJ may insert or delete nucleotides from the break point junction [45,46]. DSBR by NHEJ was recently implicated in intron gain when researchers identified short direct repeats flanking 43% of gained introns in *Daphnia* [27]. These repeats suggest that these introns were gained by the insertion of nucleotides during the repair of staggered DSBs by NHEJ [27] (illustrated in Figure 1d). Because NHEJ has been shown to preferentially insert mitochondrial DNA [74,75], further support for this model was garnered when the authors identified a gained intron that was homologous to the mitochondrial 16S ribosomal subunit [27]. How or why NHEJ preferentially uses mitochondrial DNA [74,75], why these sequences appear to integrate preferentially into genic regions [58], and why these sequences may function as introns, if not purely by chance, are topics that await further investigation.

Currently, indirect evidence for this proposed mechanism consists of gained introns with high similarity to mitochondrial DNA, rather than nuclear DNA (Table 1). In an attempt to determine the frequency by which introns without similarity to mitochondrial DNA may be gained via NHEJ, researchers have also quantified the number of repeats found flanking gained introns. In comparison to conserved introns, some studies have found gained introns to be enriched for repeats near their splice junctions [22], while others have not [23,44]. The current incongruence of gained introns to be preferentially flanked by repeats may be the result of a number of factors, such as differences in NHEJ among species [30,76-78] and cell types [79], insufficient sample sizes, or repeats and/or introns being inserted via other mechanisms. Alternatively, sequences of DNA inserted by NHEJ that are long enough to potentially form introns [80] may not frequently be flanked by direct repeats (see, for instance, inserts >30 bp long in [58,75,81,82], however, also see one insert in [83]). Finally, it should be noted that other studies have identified a number of gained introns that are flanked by repeats [27,28]. However, these numbers must be compared to the number of conserved introns flanked by repeats. This ensures that the number of repeats found flanking gained introns is significantly higher than the background level of repeats found to naturally flank introns. For instance, in Drosophila we found that 25% of gained introns were flanked by direct repeats ≥ 5 bp; however, 26% of conserved introns were flanked by repeats of the same size [23], indicating that while many introns are flanked by repeats in Drosophila, there is no bias for gained introns to be flanked by repeats in Drosophila.

Insertion of a Group II intron

Group II introns are self-splicing introns found in bacterial genomes and the organellular genomes of many eukaryotes. A number of remarkable similarities exist between group II introns and spliceosomal introns, such as their method of excision from a primary transcript, the sequence of their 5' and 3' ends, and structural similarities [84]. Even as our understanding of introns lay in its infancy, these similarities gave rise to the idea that spliceosomal introns and group II introns may be evolutionarily related [85]. Subsequent investigations have shown that it is likely that, following endosymbiosis between an α-proteobacteria and its archaeal host, many group II introns were transferred from the genome of the α-proteobacteria to the archaeal genome. The invasion and subsequent degradation of these elements may have imposed selective pressures which eventually gave rise to various hallmarks of eukaryotic cells, such as the nucleus [86,87], nonsense-mediated decay [60,88], and spliceosomal introns themselves [1,89,90].

In light of the findings that group II introns were likely the progenitors of spliceosomal introns, the retrohoming, or insertion, of a group II intron into a nuclear gene was proposed to cause recent spliceosomal intron gain [38] (Figure 1e). While it is widely believed that group II introns originally gave rise to spliceosomal introns, a recent *in vivo* assay demonstrated that the insertion of a group II intron into a nuclear gene nearly abolishes gene expression [60]. Thus, the evolution of nucleus-cytosol compartmentalization and nonsense-mediated decay following the initial invasion of group II introns (and subsequent creation of spliceosomal introns) may now impede intron gain via this mechanism. In accordance with these findings, indirect evidence of a recent intron gain via the insertion of group II intron has never been identified (Table 1). Therefore, it is unlikely that the insertion of a group II intron into a nuclear gene is a mechanism of recent spliceosomal intron gain. It is interesting to note, however, that the insertion of a group I intron into a nuclear gene does not appear to affect gene expression [60].

Intron transfer

Intron transfer has been hypothesized to result in intron gain when a paralog or pseudogene gains an intron and then transfers this intron via recombination to an intron-absent location in its sister paralog (illustrated in Figure 1f) [31]. Although indirect evidence has been found to support intron transfer (Table 1), this mechanism does not explain how the initial intron was gained, but rather explains how once gained an intron may propagate to other paralogs. It is possible that initially an imprecise intron gain occurs in a paralog or pseudogene and, following suppressor mutation(s), this intron is transferred via recombination to another paralog. Such a mechanism may allow introns to be initially gained imprecisely in a duplicated, unessential copy of a gene. Then, following suppressor mutation(s) in this neutrally evolving, extra intron-containing duplicate gene, recombination between the intron-containing region in the duplicated gene and the functional gene may result in precise or near precise intron gain in the essential copy of the gene. This process would allow an initially deleterious intron gain event in one paralog to result in non-deleterious intron gain in both paralogs without imposing a negative fitness cost to the host. If true, this suggests that intron gain rates should be higher in paralogous genes than genes without paralogs. This prediction has been confirmed in a broad range of species [15,91,92].

Intronization

Intronization is the process by which mutations create novel introns from formerly exonic sequence. Thus, unlike other proposed mechanisms of intron gain, this mechanism does not require the insertion or generation of DNA to create a novel intron. In the most commonly discussed route of intronization, mutations, generally point mutations, forge novel splice sites from exonic sequence [41,42], resulting in the formation of a novel intron (Figure 1g). An alternative route of intronization has been proposed to occur when, in an exonic segment flanked by latent splice sites, a premature termination codon is generated via mutations [93]. The spliceosome may then act upon these latent splice sites in a mysterious process known as nonsense-associated altered splicing [94-97], removing the premature termination codon from the transcript and saving it from destruction via nonsense-mediated decay. This differentially spliced product (i.e. lacking the premature termination codon) has then been proposed to persist until subsequent mutations facilitate efficient utilization of these latent splice sites by the spliceosome, resulting in the formation of a novel intron [93].

Only the most commonly discussed route of intronization, by which mutations forge novel splice sites, has garnered indirect evidence (Table 1). This may be because the alternative route of intronization is much more difficult to detect. Alternatively, this route may not occur. Regardless, nonsense-mediated decay may play some role in enabling intron gain, potentially by facilitating the persistence of an initially infrequently spliced gained intron that harbors a premature termination codon, as a recent study in Drosophila found that novel introns were enriched for in-frame stop codons [28], although it should be noted that similar results were not found in *Aspergillus* [22]. Furthermore, it is possible that the point mutations that generate potential splice sites during intronization may, at least initially, be utilized infrequently. Thus, the ability to detect intronization events relies heavily upon accurate gene annotation. Additionally, as these point mutations may not initially generate strong splicing signals, alternative splicing may occur using these or other latent splice sites, resulting in transient bouts of intronization. Such a process may confound both the process of genome annotation and analyses aimed at identifying intron gains and losses. Therefore, identification of many of these events has been limited to species with deep transcriptome profiling [41,42] or in analyses of recently created genes [61,62]. Indirect evidence of intronization may exist in fungi [22] as well as other species, but has yet to be identified. Further analyses armed with extensive RNA-seq data may shed further light on the prevalence of intronization in other species.

Trends in intron gain

Given that thousands of gained introns have been identified with no known mechanistic origin, it is apparent that the seven proposed mechanisms of intron gain fail to describe how the vast majority of novel introns have arisen. Because many gained introns have sequences that are not homologous to any endogenous sequences, it is likely that other mechanism(s) are also causing intron gain. Therefore, we must look at trends in intron gain that may lead researchers towards a different mechanistic explanation.

Perhaps the most tantalizing trend in intron gain is the role that transcription may play in intron gain. Studies in organisms with a dedicated germline have shown that genes that have experienced intron gain events are enriched for germline expression [23,24]. Other studies have shown that intron gain rates positively correlate with expression levels [25,98]. Furthermore, a recent study identified thousands of introns that appear to have been created by repeats, dubbed introner elements, in the *Micromonas* isolate CCMP1545 [99]. These repeats lack transposable element characteristics and were only found co-linear to transcribed DNA, suggesting that these repeats were conceived via a transcription-based mechanism. Much like the introner elements identified in *Micromonas*, another recent study in fungi identified elements that create introns, dubbed introner-like elements. These elements were also only found co-linear to transcribed DNA and also did not appear to be canonical transposable elements [100]. However, these elements were shown to have originated from a singular element, to be absent from species that have undergone intron transposition [25] and, while they were spliced efficiently, they were shown to rapidly degenerate into normal spliceosomal introns, indistinguishable from their original source [100]. Thus, throughout evolution, introner-like elements may have arisen in various species, caused brief episodes of massive intron gain, become silenced by an as of yet uncharacterized mechanism, and then rapidly degenerate, leaving behind no trace of the mechanism by which these introns arose [101]. If true, this would explain the varied and occasionally punctuated rates of intron gain found in some species [1,29]. Most importantly, unlike introner elements, introner-like elements are predicted to fold into stable RNA secondary structures, suggesting that these elements propagate via an RNA intermediate [100]. Taken together, these results suggest that either the act of transcription and/or the transcript itself may play an important and as of yet undefined role in intron gain. Given that RT does not appear to have played a major role in intron gain throughout evolution [34,35], and that the newly identified introner-like elements appear to propagate via an RNA intermediate and are only found in transcribed regions [100], it is more likely that the act of transcription and/or the transcript itself, without being converted to cDNA, enables intron gain. The exact process (es) by which transcription may facilitate or cause intron gain, if it occurs at all, warrants further investigation.

Frequently a positional bias has been used to support possible mechanisms of intron gain or loss. A bias for intron gains or losses in the 3' end of genes suggests that RT may have played a role in these events because RT transcribes from 3' to 5' and is known to frequently dissociate from templates before reaching their 5' ends [102,103]. Overall, there is discordance among species as to the positional bias for intron gain: in some species intron gains appear biased towards the 5' end of genes [17,23,28]; in others intron gains appear biased towards the 3' end of genes [10]; and others appear to have no detectable bias [13,25,35,44]. This discrepancy suggests that intron gain may not rely heavily upon any mechanism that suffers from a positional bias. Alternatively, mechanisms of intron gain that favor 5' or 3' gain may operate in some species but not in others.

Conclusion

Out of seven proposed mechanisms of intron gain, six have indirect evidence to support their occurrence (Table 1). The only proposed mechanism of recent intron gain that lacks any indirect evidence is the insertion of a group II intron. The insertion of a group II intron into a nuclear gene has also been shown to nearly abolish gene expression *in vivo* [60], suggesting that while group II introns were likely the progenitors of modern spliceosomal introns, they no longer create spliceosomal introns. Only one proposed mechanism of intron gain, tandem genomic duplication, has been shown to have been a genuine mechanism of intron gain *in vivo* [56]. Furthermore, this mechanism has a plethora of indirect evidence (Table 1), strongly suggesting that this mechanism is a prevalent and ubiquitous mechanism of intron gain in many species. The testing of other proposed mechanisms *in vivo*, in particular intron gain during DSBR, intron transfer and intronization, is feasible. Demonstration of these mechanisms *in vivo* is essential to solidify them as genuine mechanisms of intron gain. Further genomic analyses, especially those performed at the population level, may then quantify the relative contribution of each mechanism, potentially identifying species-specific biases that may help account for the varied rates of intron gain among species [1,29].

Significant progress has been made in identifying the mechanisms of intron gain. However, this field still lies in its infancy. Even with one definitive and five likely mechanisms of intron gain, the vast majority of gained introns lack a known mechanistic origin. Therefore, it is essential that other mechanisms are envisioned and tested, as it is likely that undiscovered mechanism(s) of intron gain exist. Germline expressed and highly expressed genes tend to accumulate introns [23-25,98]; yet, RT does not appear to have played a major role in intron gain [35]. Furthermore, a recent study identified intron creation via introner-like elements, which have been posited to have potentially created introns in species other than fungi and propagate via an RNA intermediate [100]. Hence, it is possible that an undiscovered mechanism of intron gain exists that relies upon either the act of transcription and/or the transcript itself. It is also possible that some other completely uncharacterized molecular mechanism is responsible for recent intron gains, or that novel introns are being obtained from unknown exogenous sources, such as viruses, bacteria or mobile genetic elements that have not yet been sequenced or identified. A combination of *in vivo* assays and genomic analyses performed at the population level, which will likely identify intron gains before extensive sequence divergence obscures their mechanistic origin, will likely prove the most fruitful avenues towards identifying and understanding the molecular processes underlying intron gain.

Abbreviations

DSBR: Double-strand break repair; NHEJ: Non-homologous end joining; RT: Reverse transcriptase; RTMIL: Reverse Transcriptase-Mediated Intron Loss; DSB: Double-strand break.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

PY drafted, illustrated and revised the manuscript. LZ revised the manuscript. All authors read and approved of the final version of the manuscript.

Reviewers' comments

Reviewer #1: Dr. Scott Roy (nominated by Dr. W. Ford Doolittle)

Reviewer's comments: The authors provide a very useful and timely review on the mechanisms of intron gain, a field in which a tremendous amount of progress has been made in the past few years. I have very few criticisms, and even on those points where I might have a difference of opinion in regards to emphasis or interpretation, I think the authors' perspective is well supported and balanced, and so overall I do not think any changes need to be made.

Authors' response: We thank you for your review and comments. Because most of these mechanisms lack definitive evidence to support how/if they occur, we agree that there exist various, and at times conflicting, opinions and interpretations upon how/if these proposed mechanisms occur. Thus, we are delighted that you feel our presentation was balanced and supported by primary literature.

Reviewer #2: Dr. Eugene Koonin

Reviewer's comments: This is a timely review of the important and fascinating problem of intron gain routes and mechanisms. The article is very well structured around the 7 distinct (proposed) routes of intron gain. My mild disappointments have to do with the excessive brevity of some of the sections. In particular, the section on Group II intron insertion is succinct to the point of being potentially misleading. I agree with the authors that there is no indication of recent intron gain via Group II intron insertion. However, this does not put into doubt the ultimate origin of spliceosomal introns from Group II introns whereas from the current version, the impression is that the authors refute that scenario. The situation can be easily remedied with a brief recapitulation of the evidence in support of the evolutionary links between self-splicing and spliceosomal introns.

Authors' response: We intentionally kept our review brief in order to meet the word limit of reviews in this journal (~3000 words). We apologize if, in our desire to write succinctly, we may have accidently expressed our beliefs and prior findings incorrectly, especially in regards to the origin of spliceosomal introns. As you mentioned, a significant amount of evidence exists to support the theory that group II introns, originating from an α-proteobacterium that would later go onto serve as the mitochondria, invaded the genome of their archaeal host, and in response to this invasion, selective pressures eventually gave rise to the nucleus, nonsense-mediated decay, and spliceosomal introns. We do note that in our original manuscript we directly stated that "group II introns are widely regarded as the progenitors of spliceosomal introns". However, the wording of this section may have been misleading. Thus, we have rewritten the section "Insertion of Group II Intron" to better exemplify this point and have also added a brief primer on the origin of spliceosomal introns.

Reviewer's comments: In addition, the authors do not explicitly address the connection between alternative splicing and intron gain mechanisms. Emergence of a new alternative splice form often involves ("part-time") intronization, so I think this belongs in the paper.

Authors' response: Alternative splicing and, in particular, nonsense-mediated decay likely play key roles in the evolution of gained introns by alleviating negative fitness costs imposed upon organisms following the initial intron gain event (or, more explicitly stated, the initial mutation). However, due to space limitations, we chose to focus only on the mechanisms that may initially cause the intron gain (the initial mutation). The forces at play following this mutation that fixate this mutation in a population/species deserve a more in-depth treatment than we could provide with the space allotted. Regardless, we have added a brief discussion of these topics to the "Intronization" section, as this mechanism may rely most heavily upon evolutionary processes to yield efficient introns.

Reviewer's comments: Again, a very timely and useful review but with a little more attention to details, it could become even better.

Authors' response: We thank you for your review and apologize if we may have neglected some topics in an effort to be succinct.

Reviewer #3: Dr. John Logsdon, University of Iowa

This reviewer provided no comments for publication.

Acknowledgements

We would like to thank two reviewers, Dr. Scott Roy and Dr. Eugene Koonin, for their great comments, which are very helpful for the further improvement of this review. This work was supported by NSF CPATH program, IIS-0938393.

Author details

[1]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA. [2]Department of Health Information Management, University of Pittsburgh, Pittsburgh, PA 15260, USA. [3]Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA 15260, USA.

Received: 23 May 2012 Accepted: 22 August 2012
Published: 10 September 2012

References

1. Rogozin IB, Carmel L, Csuros M, Koonin EV: **Origin and evolution of spliceosomal introns.** *Biol Direct* 2012, **7**:11.
2. Koonin EV: **Intron-dominated genomes of early ancestors of eukaryotes.** *J Hered* 2009, **100**:618–623.
3. Graveley BR: **Alternative splicing: increasing diversity in the proteomic world.** *Trends Genet* 2001, **17**:100–107.
4. Choi T, Huang M, Gorman C, Jaenisch R: **A generic intron increases gene expression in transgenic mice.** *Mol Cell Biol* 1991, **11**:3070–3074.
5. Le HH, Nott A, Moore MJ: **How introns influence and enhance eukaryotic gene expression.** *Trends Biochem Sci* 2003, **28**:215–220.
6. Valencia P, Dias AP, Reed R: **Splicing promotes rapid and efficient mRNA export in mammalian cells.** *Proc Natl Acad Sci USA* 2008, **105**:3386–3391.
7. Rearick D, Prakash A, McSweeny A, Shepard SS, Fedorova L, Fedorov A: **Critical association of ncRNA with introns.** *Nucleic Acids Res* 2011, **39**:2357–2366.
8. Mattick JS, Makunin IV: **Non-coding RNA.** *Hum Mol Genet* 2006, **15**(Spec No 1):R17–R29.
9. Rogers J: **Exon shuffling and intron insertion in serine protease genes.** *Nature* 1985, **315**:458–459.
10. Sverdlov AV, Babenko VN, Rogozin IB, Koonin EV: **Preferential loss and gain of introns in 3' portions of genes suggests a reverse-transcription mechanism of intron insertion.** *Gene* 2004, **338**:85–91.
11. Rogozin IB, Wolf YI, Sorokin AV, Mirkin BG, Koonin EV: **Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution.** *Curr Biol* 2003, **13**:1512–1517.
12. Roy SW, Fedorov A, Gilbert W: **Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain.** *Proc Natl Acad Sci USA* 2003, **100**:7158–7162.
13. Nielsen CB, Friedman B, Birren B, Burge CB, Galagan JE: **Patterns of intron gain and loss in fungi.** *PLoS Biol* 2004, **2**:e422.
14. Roy SW, Hartl DL: **Very little intron loss/gain in Plasmodium: intron loss/gain mutation rates and intron number.** *Genome Res* 2006, **16**:750–756.
15. Lin H, Zhu W, Silva JC, Gu X, Buell CR: **Intron gain and loss in segmentally duplicated genes in rice.** *Genome Biol* 2006, **7**:R41.
16. Stajich JE, Dietrich FS: **Evidence of mRNA-mediated intron loss in the human-pathogenic fungus Cryptococcus neoformans.** *Eukaryot Cell* 2006, **5**:789–793.
17. Coulombe-Huntington J, Majewski J: **Intron loss and gain in Drosophila.** *Mol Biol Evol* 2007, **24**:2842–2850.
18. Knowles DG, McLysaght A: **High rate of recent intron gain and loss in simultaneously duplicated Arabidopsis genes.** *Mol Biol Evol* 2006, **23**:1548–1557.
19. Coulombe-Huntington J, Majewski J: **Characterization of intron loss events in mammals.** *Genome Res* 2007, **17**:23–32.
20. Loh YH, Brenner S, Venkatesh B: **Investigation of loss and gain of introns in the compact genomes of pufferfishes (Fugu and Tetraodon).** *Mol Biol Evol* 2008, **25**:526–535.
21. Sharpton TJ, Neafsey DE, Galagan JE, Taylor JW: **Mechanisms of intron gain and loss in Cryptococcus.** *Genome Biol* 2008, **9**:R24.
22. Zhang LY, Yang YF, Niu DK: **Evaluation of models of the mechanisms underlying intron loss and gain in Aspergillus fungi.** *J Mol Evol* 2010, **71**:364–373.
23. Yenerall P, Krupa B, Zhou L: **Mechanisms of intron gain and loss in Drosophila.** *BMC Evol Biol* 2011, **11**:364.

24. Coghlan A, Wolfe KH: **Origins of recently gained introns in Caenorhabditis.** *Proc Natl Acad Sci USA* 2004, **101:**11362–11367.
25. Denoeud F, Henriet S, Mungpakdee S, Aury JM, Da SC, Brinkmann H, Mikhaleva J, Olsen LC, Jubin C, Canestro C, *et al*: **Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate.** *Science* 2010, **330:**1381–1385.
26. Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, Tokishita S, Aerts A, Arnold GJ, Basu MK, *et al*: **The ecoresponsive genome of Daphnia pulex.** *Science* 2011, **331:**555–561.
27. Li W, Tucker AE, Sung W, Thomas WK, Lynch M: **Extensive, recent intron gains in Daphnia populations.** *Science* 2009, **326:**1260–1262.
28. Farlow A, Meduri E, Dolezal M, Hua L, Schlotterer C: **Nonsense-mediated decay enables intron gain in Drosophila.** *PLoS Genet* 2010, **6:**e1000819.
29. Csuros M, Rogozin IB, Koonin EV: **A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes.** *PLoS Comput Biol* 2011, **7:**e1002150.
30. Farlow A, Meduri E, Schlotterer C: **DNA double-strand break repair and the evolution of intron density.** *Trends Genet* 2010, **27:**1–6.
31. Hankeln T, Friedl H, Ebersberger I, Martin J, Schmidt ER: **A variable intron distribution in globin genes of Chironomus: evidence for recent intron gain.** *Gene* 1997, **205:**151–160.
32. Fedorov A, Roy S, Fedorova L, Gilbert W: **Mystery of intron gain.** *Genome Res* 2003, **13:**2236–2241.
33. Tarrio R, Ayala FJ, Rodriguez-Trelles F: **Alternative splicing: a missing piece in the puzzle of intron gain.** *Proc Natl Acad Sci USA* 2008, **105:**7223–7228.
34. Roy SW, Irimia M: **Mystery of intron gain: new data and new models.** *Trends Genet* 2009, **25:**67–73.
35. Cohen NE, Shen R, Carmel L: **The role of reverse-transcriptase in intron gain and loss mechanisms.** *Mol Biol Evol* 2011, **29:**179–186.
36. Derr LK, Strathern JN: **A role for reverse transcripts in gene conversion.** *Nature* 1993, **361:**170–173.
37. Roy SW, Gilbert W: **The evolution of spliceosomal introns: patterns, puzzles and progress.** *Nat Rev Genet* 2006, **7:**211–221.
38. Sharp PA: **On the origin of RNA splicing and introns.** *Cell* 1985, **42:**397–400.
39. Crick F: **Split genes and RNA splicing.** *Science* 1979, **204:**264–271.
40. Rogers JH: **How were introns inserted into nuclear genes?** *Trends Genet* 1989, **5:**213–216.
41. Irimia M, Rukov JL, Penny D, Vinther J, Garcia-Fernandez J, Roy SW: **Origin of introns by 'intronization' of exonic sequences.** *Trends Genet* 2008, **24:**378–381.
42. Roy SW: **Intronization, de-intronization and intron sliding are rare in Cryptococcus.** *BMC Evol Biol* 2009, **9:**192.
43. Roy SW, Penny D: **Smoke without fire: most reported cases of intron gain in nematodes instead reflect intron losses.** *Mol Biol Evol* 2006, **23:**2259–2262.
44. Torriani SF, Stukenbrock EH, Brunner PC, McDonald BA, Croll D: **Evidence for extensive recent intron transposition in closely related fungi.** *Curr Biol* 2011, **21:**2017–2022.
45. Weterings E, Chen DJ: **The endless tale of non-homologous end-joining.** *Cell Res* 2008, **18:**114–124.
46. Lieber MR: **The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway.** *Annu Rev Biochem* 2010, **79:**181–211.
47. Mahaney BL, Meek K, Lees-Miller SP: **Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining.** *Biochem J* 2009, **417:**639–650.
48. Weterings E, van Gent DC: **The mechanism of non-homologous end-joining: a synopsis of synapsis.** *DNA Repair (Amst)* 2004, **3:**1425–1435.
49. Rodriguez-Trelles F, Tarrio R, Ayala FJ: **Models of spliceosomal intron proliferation in the face of widespread ectopic expression.** *Gene* 2006, **366:**201–208.
50. Tseng CK, Cheng SC: **Both catalytic steps of nuclear pre-mRNA splicing are reversible.** *Science* 2008, **320:**1782–1784.
51. Crabb TL, Lam BJ, Hertel KJ: **Retention of spliceosomal components along ligated exons ensures efficient removal of multiple introns.** *RNA* 2010, **16:**1786–1796.
52. Le HH, Izaurralde E, Maquat LE, Moore MJ: **The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon-exon junctions.** *EMBO J* 2000, **19:**6860–6869.
53. Kataoka N, Yong J, Kim VN, Velazquez F, Perkinson RA, Wang F, Dreyfuss G: **Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm.** *Mol Cell* 2000, **6:**673–682.
54. Iwamoto M, Maekawa M, Saito A, Higo H, Higo K: **Evolutionary relationship of plant catalase genes inferred from exon-intron structures: isozyme divergence after the separation of monocots and dicots.** *TAG Theoretical and Applied Genetics* 1998, **97:**9–19.
55. Giroux MJ, Clancy M, Baier J, Ingham L, McCarty D, Hannah LC: **De novo synthesis of an intron by the maize transposable element Dissociation.** *Proc Natl Acad Sci USA* 1994, **91:**12150–12154.
56. Hellsten U, Aspden JL, Rio DC, Rokhsar DS: **A segmental genomic duplication generates a functional intron.** *Nat Commun* 2011, **2:**454.
57. Gao X, Lynch M: **Ubiquitous internal gene duplication and intron creation in eukaryotes.** *Proc Natl Acad Sci USA* 2009, **106:**20818–20823.
58. Ricchetti M, Tekaia F, Dujon B: **Continued colonization of the human genome by mitochondrial DNA.** *PLoS Biol* 2004, **2:**e273.
59. Curtis BA, Archibald JM: **A spliceosomal intron of mitochondrial DNA origin.** *Curr Biol* 2010, **20:**R919–R920.
60. Chalamcharla VR, Curcio MJ, Belfort M: **Nuclear expression of a group II intron is consistent with spliceosomal intron ancestry.** *Genes Dev* 2010, **24:**827–836.
61. Szczesniak MW, Ciomborowska J, Nowak W, Rogozin IB, Makalowska I: **Primate and rodent specific intron gains and the origin of retrogenes with splice variants.** *Mol Biol Evol* 2010, **28:**33–37.
62. Yang Z, Huang J: **De novo origin of new genes with introns in Plasmodium vivax.** *FEBS Lett* 2011, **585:**641–644.
63. Dibb NJ, Newman AJ: **Evidence that introns arose at proto-splice sites.** *EMBO J* 1989, **8:**2015–2021.
64. Qiu WG, Schisler N, Stoltzfus A: **The evolutionary gain of spliceosomal introns: sequence and phase preferences.** *Mol Biol Evol* 2004, **21:**1252–1263.
65. Sverdlov AV, Rogozin IB, Babenko VN, Koonin EV: **Reconstruction of ancestral protosplice sites.** *Curr Biol* 2004, **14:**1505–1508.
66. Lee VD, Stapleton M, Huang B: **Genomic structure of Chlamydomonas caltractin. Evidence for intron insertion suggests a probable genealogy for the EF-hand superfamily of proteins.** *J Mol Biol* 1991, **221:**175–191.
67. Kupfer DM, Drabenstot SD, Buchanan KL, Lai H, Zhu H, Dyer DW, Roe BA, Murphy JW: **Introns and splicing elements of five diverse fungi.** *Eukaryot Cell* 2004, **3:**1088–1100.
68. Fahey ME, Higgins DG: **Gene expression, intron density, and splice site strength in Drosophila and Caenorhabditis.** *J Mol Evol* 2007, **65:**349–357.
69. Burset M, Seledtsov IA, Solovyev VV: **SpliceDB: database of canonical and non-canonical mammalian splice sites.** *Nucleic Acids Res* 2001, **29:**255–259.
70. Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S: **Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information.** *Nucleic Acids Res* 1996, **24:**3439–3452.
71. Mount SM, Burks C, Hertz G, Stormo GD, White O, Fields C: **Splicing signals in Drosophila: intron size, information content, and consensus sequences.** *Nucleic Acids Res* 1992, **20:**4255–4262.
72. Rothkamm K, Kruger I, Thompson LH, Lobrich M: **Pathways of DNA double-strand break repair during the mammalian cell cycle.** *Mol Cell Biol* 2003, **23:**5706–5715.
73. Shrivastav M, De Haro LP, Nickoloff JA: **Regulation of DNA double-strand break repair pathway choice.** *Cell Res* 2008, **18:**134–147.
74. Paques F, Haber JE: **Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae.** *Microbiol Mol Biol Rev* 1999, **63:**349–404.
75. Ricchetti M, Fairhead C, Dujon B: **Mitochondrial DNA repairs double-strand breaks in yeast chromosomes.** *Nature* 1999, **402:**96–100.
76. Pelczar P, Kalck V, Kovalchuk I: **Different genome maintenance strategies in human and tobacco cells.** *J Mol Biol* 2003, **331:**771–779.
77. Li G, Nelsen C, Hendrickson EA: **Ku86 is essential in human somatic cells.** *Proc Natl Acad Sci USA* 2002, **99:**832–837.
78. Lewis LK, Resnick MA: **Tying up loose ends: nonhomologous end-joining in Saccharomyces cerevisiae.** *Mutat Res* 2000, **451:**71–89.
79. Sharma S, Choudhary B, Raghavan SC: **Efficiency of nonhomologous DNA end joining varies among somatic tissues, despite similarity in mechanism.** *Cell Mol Life Sci* 2010, **68:**661–676.
80. Yu J, Yang Z, Kibukawa M, Paddock M, Passey DA, Wong GK: **Minimal introns are not "junk".** *Genome Res* 2002, **12:**1185–1189.
81. Varga T, Aplan PD: **Chromosomal aberrations induced by double strand DNA breaks.** *DNA Repair* 2005, **4:**1038–1046.
82. Gorbunova V, Levy AA: **Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions.** *Nucleic Acids Res* 1997, **25:**4650–4657.

83. Hazkani-Covo E, Covo S: **Numt-mediated double-strand break repair mitigates deletions during primate genome evolution.** *PLoS Genet* 2008, **4:**e1000237.
84. Lambowitz AM, Zimmerly S: **Mobile group II introns.** *Annu Rev Genet* 2004, **38:**1–35.
85. Cech TR: **The generality of self-splicing RNA: relationship to nuclear mRNA splicing.** *Cell* 1986, **44:**207–210.
86. Lopez-Garcia P, Moreira D: **Selective forces for the origin of the eukaryotic nucleus.** *Bioessays* 2006, **28:**525–533.
87. Martin W, Koonin EV: **Introns and the origin of nucleus-cytosol compartmentalization.** *Nature* 2006, **440:**41–45.
88. Lynch M, Kewalramani A: **Messenger RNA surveillance and the evolutionary proliferation of introns.** *Mol Biol Evol* 2003, **20:**563–571.
89. Bonen L, Vogel J: **The ins and outs of group II introns.** *Trends Genet* 2001, **17:**322–331.
90. Rodriguez-Trelles F, Tarrio R, Ayala FJ: **Origins and evolution of spliceosomal introns.** *Annu Rev Genet* 2006, **40:**47–76.
91. Castillo-Davis CI, Bedford TB, Hartl DL: **Accelerated rates of intron gain/loss and protein evolution in duplicate genes in human and mouse malaria parasites.** *Mol Biol Evol* 2004, **21:**1422–1427.
92. Babenko VN, Rogozin IB, Mekhedov SL, Koonin EV: **Prevalence of intron gain over intron loss in the evolution of paralogous gene families.** *Nucleic Acids Res* 2004, **32:**3724–3733.
93. Catania F, Lynch M: **Where do introns come from?** *PLoS Biol* 2008, **6:**e283.
94. Li B, Wachtel C, Miriami E, Yahalom G, Friedlander G, Sharon G, Sperling R, Sperling J: **Stop codons affect 5' splice site selection by surveillance of splicing.** *Proc Natl Acad Sci USA* 2002, **99:**5277–5282.
95. Wang J, Chang YF, Hamilton JI, Wilkinson MF: **Nonsense-associated altered splicing: a frame-dependent response distinct from nonsense-mediated decay.** *Mol Cell* 2002, **10:**951–957.
96. Cartegni L, Chew SL, Krainer AR: **Listening to silence and understanding nonsense: exonic mutations that affect splicing.** *Nat Rev Genet* 2002, **3:**285–298.
97. Wang J, Hamilton JI, Carter MS, Li S, Wilkinson MF: **Alternatively spliced TCR mRNA induced by disruption of reading frame.** *Science* 2002, **297:**108–110.
98. Carmel L, Rogozin IB, Wolf YI, Koonin EV: **Evolutionarily conserved genes preferentially accumulate introns.** *Genome Res* 2007, **17:**1045–1050.
99. Worden AZ, Lee JH, Mock T, Rouze P, Simmons MP, Aerts AL, Allen AE, Cuvelier ML, Derelle E, Everett MV, *et al*: **Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes Micromonas.** *Science* 2009, **324:**268–272.
100. van der Burgt A, Severing E, de Wit PJ, Collemare J: **Birth of New Spliceosomal Introns in Fungi by Multiplication of Introner-like Elements.** *Curr Biol* 2012, **22:**1260–1265.
101. Roy SW, Irimia M: **Genome evolution: where do new introns come from?** *Curr Biol* 2012, **22:**R529–R531.
102. Abbotts J, Bebenek K, Kunkel TA, Wilson SH: **Mechanism of HIV-1 reverse transcriptase. Termination of processive synthesis on a natural DNA template is influenced by the sequence of the template-primer stem.** *J Biol Chem* 1993, **268:**10312–10323.
103. Klarmann GJ, Schauber CA, Preston BD: **Template-directed pausing of DNA synthesis by HIV-1 reverse transcriptase during polymerization of HIV-1 sequences in vitro.** *J Biol Chem* 1993, **268:**9793–9802.

doi:10.1186/1745-6150-7-29
Cite this article as: Yenerall and Zhou: **Identifying the mechanisms of intron gain: progress and trends.** *Biology Direct* 2012 **7**:29.