Mobile DNA

Evolution of group II introns

Steven Zimmerly* and Cameron Semper

Abstract

Present in the genomes of bacteria and eukaryotic organelles, group II introns are an ancient class of ribozymes and retroelements that are believed to have been the ancestors of nuclear pre-mRNA introns. Despite long-standing speculation, there is limited understanding about the actual pathway by which group II introns evolved into eukaryotic introns. In this review, we focus on the evolution of group II introns themselves. We describe the different forms of group II introns known to exist in nature and then address how these forms may have evolved to give rise to spliceosomal introns and other genetic elements. Finally, we summarize the structural and biochemical parallels between group II introns and the spliceosome, including recent data that strongly support their hypothesized evolutionary relationship.

Keywords: Ribozyme, Retroelement, Spliceosome, Molecular evolution, Mobile DNA

Review

Introduction

Investigating the evolution of mobile DNAs involves unique challenges compared to other evolutionary studies. The sequences of mobile DNAs are usually short and evolve rapidly, resulting in limited phylogenetic signals. The elements often transfer horizontally, which prevents the linkage of their evolution to that of their host organisms or other genes in the organism. Finally, many mobile elements themselves consist of multiple components that may have different evolutionary histories. All of these complicating factors apply to group II introns and must be considered when trying to understand their evolutionary history.

Group II intron retroelements consist of an RNA and a protein component. The RNA is a ribozyme (catalytic RNA) that is capable of self-splicing *in vitro*, while the intron-encoded protein (IEP)'s open reading frame (ORF) sequence is contained internally within the RNA sequence and encodes a reverse transcriptase (RT) protein [1-6]. The two components cooperate intricately to carry out a series of inter-related reactions that accomplish intron splicing and retromobility. In addition to the 2- to 3-kb retroelement form, group II introns have evolved into many variant forms and spread throughout all domains of life. They are present in bacteria, archaebacteria, mitochondria, and

chloroplasts but are notably excluded from nuclear genomes, with the exception of presumably inert sequences transferred to the nucleus as segments of mitochondrial DNA [7,8].

Group II introns have attracted considerable attention, in part due to their hypothesized relationship to eukaryotic pre-mRNA introns. The purpose of this review is to carefully consider the evidence available regarding the evolutionary history of group II introns. We present a summary of the multiple types of group II introns known to exist in nature and discuss a model for how the variant forms arose and subsequently evolved into spliceosomal introns and other elements.

Structure and properties of group II introns

The biochemical and genetic properties of group II introns have been described in depth elsewhere [1,3,5,6,9-14] and are summarized briefly here. Of the 2- to 3-kb intron sequence, the RNA component corresponds to approximately 500 to 900 bps, which are separated between the first approximately 600 bp and last approximately 100 bp of the intron sequence (red shading in Figure 1A). After transcription, the RNA folds into a complex structure that carries out splicing [12,14-18]. There is little conservation of primary sequence among all group II intron RNAs, but the introns fold into a common secondary structure that consists of six domains (Figure 1B). Domain I is very large and comprises about half of the ribozyme. Among other roles, it serves as a structural scaffold for

* Correspondence: zimmerly@ucalgary.ca
Department of Biological Sciences, University of Calgary, 2500 University
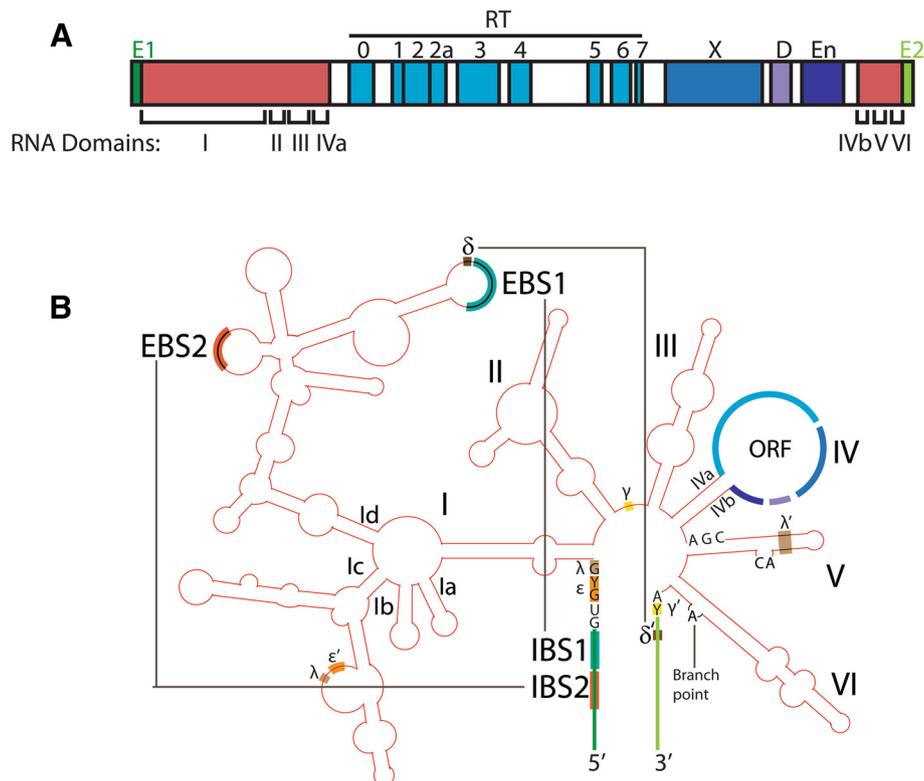Drive N.W., Calgary, Alberta T2N 1N4, Canada

Figure 1 Group II intron DNA sequence and RNA structure. (A) Genomic structure of a group II intron. The 2- to 3-kb sequence consists of RNA and protein portions. The intron RNA domains are depicted in red and demarcated with Roman numerals. Domains I to IVa are at the 5' end of the intron, while domains IVb to VI are at the 3' end. The IEP sequence is nested within the RNA's sequence and the domains are denoted by differently shaded blue boxes. The IEP contains a reverse transcriptase domain (RT) with motifs 0 to 7, a maturase domain (X, sometimes called X/thumb), a DNA-binding domain (D), and an endonuclease domain (En). Exons are shown in green. **(B)** Secondary structure of the unspliced RNA transcript. The intron RNA (red) folds into a structure of six domains, with the ORF encoded in a large loop of domain IV. The 5' and 3' exons are the green vertical lines at the bottom. Watson-Crick pairing interactions that are important for exon recognition are IBS1-EBS1, IBS2-EBS2, and δ-δ' (for IIA introns), which are shown with teal, orange, and brown shadings, respectively, and connected with black lines. For IIB and IIC introns, the 3' exon is recognized instead through an IBS3-EBS3 pairing (not shown). The ε-ε', λ-λ', and γ-γ' interactions are also indicated, because they have potential parallels in the spliceosome (Figure 5); other known tertiary interactions are omitted for simplicity. Both the RNA and DNA structures depicted correspond to the *L. lactis ltrB* intron. EBS, exon-binding site; IBS, intron-binding site; ORF, open reading frame.

the entire ribozyme and importantly recognizes and positions the exon substrates for catalysis [19-21]. Domain V is a small, highly conserved domain that contains the so-called catalytic triad AGC (or CGC for some introns), which binds two catalytically important metal ions [22,23]. Domain VI contains the bulged A motif that is the branch site during the splicing reaction. Splicing is accomplished by two transesterification reactions that produce ligated exons and excised intron lariat (Figure 2A) [24,25]. For some group II introns, the RNA component alone can self-splice *in vitro* under appropriate reaction conditions, typically with elevated concentrations of magnesium and/or salt.

The IEP is encoded within the loop of the RNA domain IV (Figure 1) and is translated from the unspliced precursor transcript. The IEP contains seven sequence blocks that are conserved across different types of RTs, as well as the X domain that is the thumb structure of

the RT protein but is not highly conserved in sequence (Figure 1A) [26-29]. Downstream of domain X are DNA binding (D) and endonuclease (En) domains, which are critical for retromobility [30-33].

Both the RNA and IEP are required for splicing and mobility reactions *in vivo*. The translated IEP binds to the unspliced intron structure via the RT and X domains, which results in RNA conformational adjustments leading to splicing (Figure 2A) [34-38]. The role of the IEP in splicing is known as maturase activity because it results in maturation of the mRNA. After splicing, the IEP remains bound to the lariat to form a ribonucleoprotein (RNP) that is the machinery that carries out a retromobility reaction [35,39].

For most group II introns, the mobility reaction is highly specific to a defined target sequence of approximately 20 to 35 bp known as the homing site. The mechanism of mobility is called target-primed reverse transcription
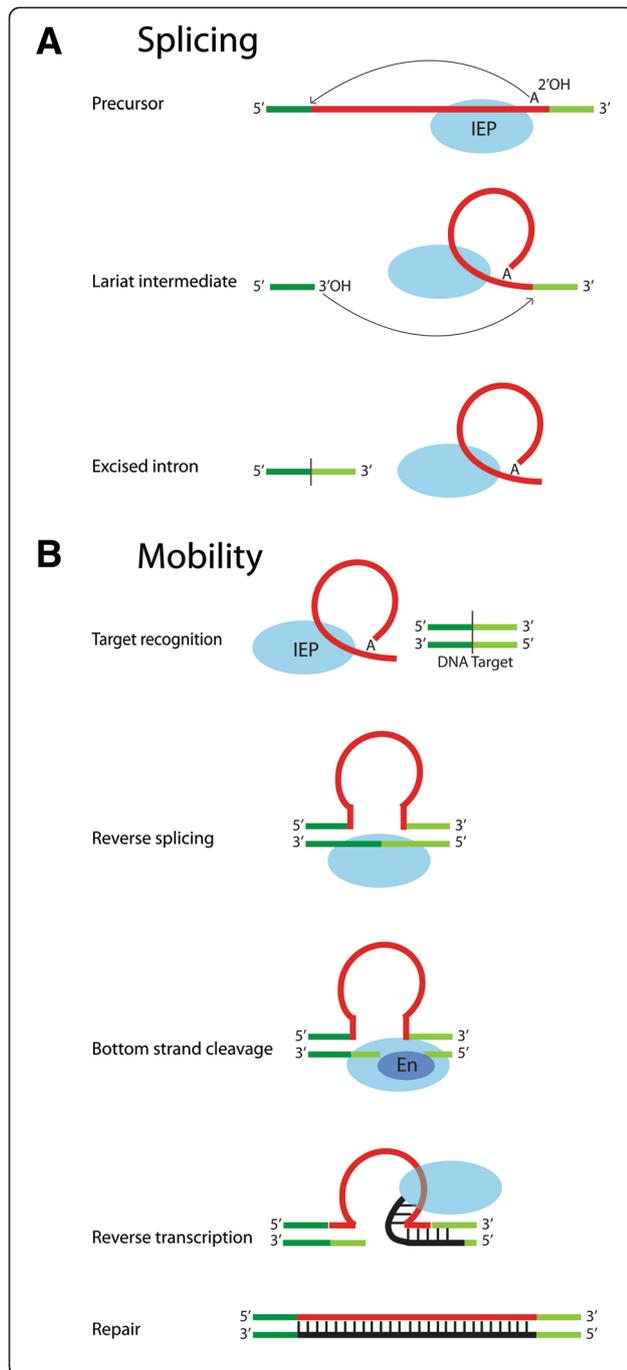
Figure 2 Group II intron activities. (A) The splicing reaction. Splicing is intrinsically RNA-catalyzed and occurs for naked RNA *in vitro*; however, under physiological conditions, the IEP is required as well. The IEP binds to the RNA structure to enable it to adopt its catalytic conformation and accomplish splicing. In the first transesterification step of splicing, the 2′ OH of the branch site adenosine initiates nucleophilic attack on the 5′ splice junction, yielding cleaved 5′ exon and a lariat-3′ exon intermediate. In the second transesterification, the 3′ OH of the 5′ exon attacks the 3′ splice site to form ligated exons and intron lariat. The IEP remains tightly bound to the lariat to form a mobility-competent RNP particle. **(B)** The mobility reaction, known as target-primed reverse transcription (TPRT). The RNP product of splicing recognizes the DNA target site and reverse splices into the top strand. The En domain cleaves the bottom strand and the free 3′ OH is the primer for reverse-transcription. Host repair activities, which vary across organisms, complete the process. IEP, intron-encoded protein.

(TPRT) [6,10,31,40-44]. The RNP first recognizes and unwinds the two strands of the target, and the intron RNA reverse splices into the top strand of the DNA (Figure 2B). The reaction is the reverse of splicing but utilizes DNA exons rather than RNA exons, and so part of the target site specificity comes from the intron-binding site 1 (IBS1)-exon-binding site 1 (EBS1), IBS2-EBS2, and δ-δ′ pairings between the intron RNA and DNA exons. The IEP facilitates reverse splicing analogously as it does in the forward splicing reaction, that is, it helps the ribozyme fold into its catalytic conformation. In addition, the IEP contributes to target site specificity through interactions of its D domain with the DNA exons. The bottom strand of the target DNA is cleaved by the En domain, either 9 or 10 bp downstream of the insertion site to create a 3′OH that is the primer for reverse transcription of the inserted intron [31,45]. Repair processes convert the inserted sequence to double-stranded DNA, although the repair activities involved differ across host organisms [46-48].

Relevant to this review is a key distinction in the character of group II introns in bacteria compared to introns in mitochondria and chloroplasts. In bacteria, the introns behave mainly as mobile DNAs that survive by constant movement to new genomic sites, whereas in organelles, they are less mobile [5,49,50]. This can be inferred from genome sequences because the majority of intron copies in bacteria are truncated or inactivated, and many are surrounded by other mobile DNAs [49,51]. Most bacterial introns are located outside of housekeeping genes so that their splicing does not greatly affect the host biology. On the other hand, in organelles group II, introns are almost always located in housekeeping genes, which necessitates that they splice efficiently [1,15]. Organellar introns are rarely truncated and frequently have lost mobility properties altogether to become splicing-only entities. As opposed to bacterial introns, organellar introns have taken up a more stable residence in genomes, potentially assuming roles in gene regulation because their splicing factors are under nuclear control (below).

Major classes of group II introns

The varieties of group II introns can be classified either according to their RNA or IEP components. Group II introns were initially classified as IIA or IIB based on the RNA sequence and secondary structure characteristics of introns in mitochondrial and chloroplast genomes [15]. A third variation of RNA structure was subsequently identified in bacteria, IIC [52,53]. These three classes each exhibit considerable variation, especially IIB introns, and classes can be further subdivided (for example, IIB1 and IIB2) [15,54]. The most prominent difference among IIA, IIB, and IIC ribozymes is the mechanism of exon recognition, because each class uses a distinct combination of pairing interactions to recognize the 5′ and 3′ exons (that is, different combinations of IBS1-EBS1, IBS2-EBS2, IBS3-EBS3, and δ-δ′ pairings [15,17,19,21,55]).

Alternatively, group II introns can be classified according to phylogenetic analysis of their IEP amino acid sequences. Eight IEP classes have been defined: mitochondrial-like (ML), chloroplast-like (CL), A, B, C, D, E, and F [28,50,56]. The two classification systems are useful for different purposes. Classes IIA, IIB, and IIC apply to all introns regardless of whether they encode an IEP, whereas the IEP-based classes are more specific and correspond to phylogenetic clades. The correspondence between the ribozyme and IEP classifications is shown in Table 1. IIA and IIB introns are found in bacteria, mitochondria, and chloroplasts, while IIC introns are only present in bacteria [15,49,53,57]. Among IEP-classified introns, all forms are found in bacteria, whereas only ML and CL introns are found in mitochondria and chloroplasts (Table 2). There is some relation between IEP classes and host organisms. For example, within bacteria, CL2 introns are almost exclusively found in Cyanobacteria, while class B introns are found exclusively in Firmicutes [50,51].

Table 1 Correspondence between RNA- and IEP-based classes

IEP-based classes[a]	RNA structure-based classes[b]		
	IIA	IIB	IIC
A	X		
B		X	
C			X
D		X	
E		X	
F		X	
ML	X		
CL		X	

[a]Classes of introns based on phylogenetic groupings of the IEPs. [b]Classes of introns based on ribozyme secondary structure characteristics. CL, chloroplast-like; ML, mitochondrial-like.

Table 2 Distribution of intron classes in different organisms and organelles

	Eubacteria	Archaebacteria	Mitochondria	Chloroplasts
RNA-based classes[a]				
IIA	X		X	X
IIB	X	X	X	X
IIC	X			
IEP-based classes[b]				
CL	X	X	X	X
ML	X		X	X
A	X			
B	X			
C	X			
D	X	X		
E	X	X		
F	X			

[a]Classes of introns based on the ribozyme structural characteristics. [b]Classes of introns based on phylogenetic groupings of the IEPs. CL, chloroplast-like; IEP, intron-encoded protein; ML, mitochondrial-like.

Intron variations that deviate from the 'standard' retroelement form

Reconstructing the evolution of group II introns requires an accounting of all known intron forms and their distribution. Here, we describe the range of variants that differ from the 'standard' retroelement form diagrammed in Figure 1.

Introns lacking En domains in the IEP Approximately a quarter of group II intron IEPs in organelles and over half in bacteria lack an En domain [44,50,51], including all introns of classes C, D, E, and F and a minority of CL introns (Figure 3B). The En domain belongs to the prokaryotic family of H-N-H nucleases [30,58], suggesting that the En domain was appended to an ancestral IEP that had only RT and X domains. If true, then at least some of the lineages of En-minus introns (classes C, D, E, F) represent a form of group II introns that predated acquisition of the En domain.

With regard to mobility mechanisms, En-minus introns are unable to form the bottom strand primer and require an alternative pathway. It has been shown for these introns that the primer is provided by the leading or lagging strand of the replication fork during DNA replication [33,59-62]. Some En-minus introns (namely, IIC/class C) use a different specificity in selecting DNA target sites. Rather than recognizing a homing site of 20 to 35 bp, IIC introns insert at the DNA motifs of intrinsic transcriptional terminators, while a smaller fraction inserts at the *attC* motifs of integrons (imperfect inverted repeat
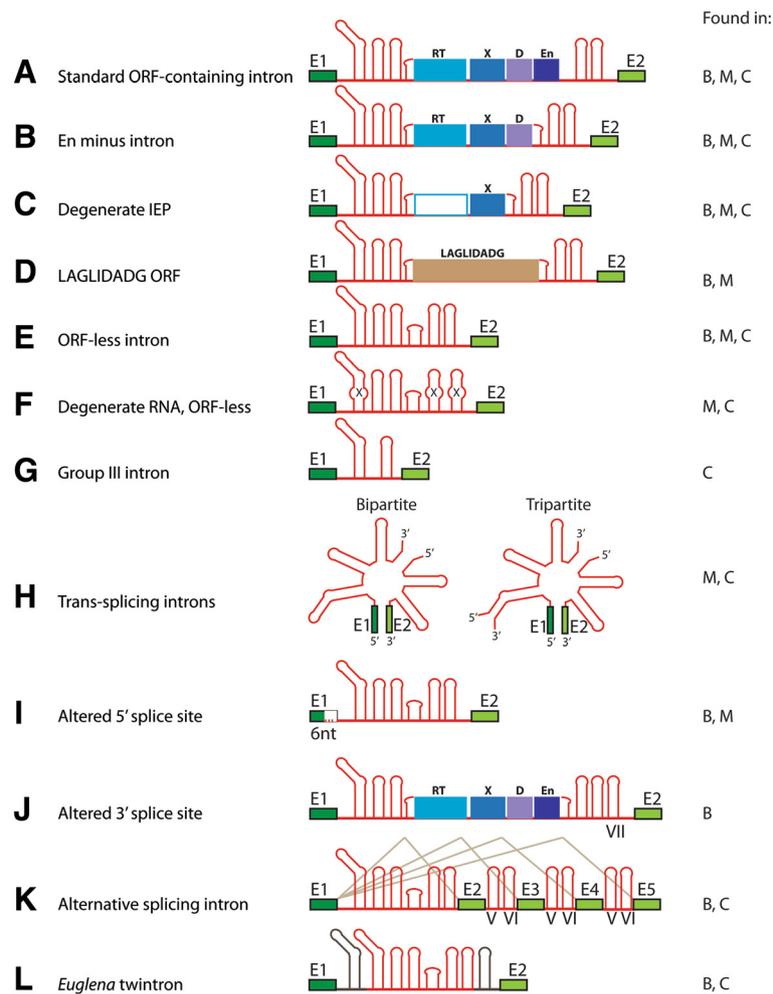
Figure 3 Variations in group II intron forms. RNA domains are depicted as stem-loops in red, ORF domains in blue or tan, and exons in green. The right column indicates whether the variants are found in bacteria (B), mitochondria (M), or chloroplasts (C). **(A)** Full-length retroelement form with standard RNA and IEP domains. Example: the IIA intron Ll.LtrB of *Lactococcus lactis*. ORF, open reading frame; RT, reverse transcriptase. **(B)** Intron lacking the endonuclease domain (found in all introns of classes C, D, E, and F and some of class CL). Example: the IIC intron B.h.I1. **(C)** Intron in which the IEP has lost RT motifs while maintaining the domain X/thumb domain required for maturase function. Example: the chloroplast IIA intron *trn*KI1, which encodes the ORF MatK. IEP, intron-encoded protein. **(D)** Intron encoding a LAGLIDADG homing endonuclease. Example: *Grifola frondosa* SSUI1 rRNA intron (fungi). **(E)** ORF-less, self-splicing intron. Example: *S. cerevisiae* aI5g. **(F)** ORF-less intron with a degenerated RNA sequence. Example: tobacco *petD*I1. **(G)** Group III intron. Example: *Euglena gracilis rps*11 **(H)** Trans-splicing group II introns. Examples: tobacco *nad*1I1 (bipartite) and *Chlamydomonas psa*AI1 (tripartite). **(I)** Altered 5′ splice site. Example: *Grifola frondosa* SSUI1 rRNA intron. **(J)** Altered 3′ splice site. Example: *Bacillus cereus* B.c.I4. **(K)** Alternatively splicing group II intron. Example: *Clostridium tetani* C.te.I1. **(L)** Twintron. Example: *Euglena gracilis rps*3.

sequences that are recognized by the integron's integrase) [49,52,63-69].

Introns with 'degenerated' IEPs that have lost RT activity Among mitochondrial and chloroplast introns, many IEPs have lost critical RT domain residues (for example, the active site motif YADD) or lost alignability altogether to some of the conserved RT motifs (for example, *trn*KI1 in plant chloroplasts, *nad*1I4 in plant mitochondria, and *psb*CI4 in *Euglena* chloroplasts) (Figure 3C) [27,28,70,71]. These divergent IEPs have undoubtedly lost RT activity and presumably have lost mobility function as well, although the splicing (maturase) function likely endures [27].

A well-studied example is the chloroplast IIA intron *trn*KI1, which is located in an essential tRNA^Lys gene. The IEP encoded by this intron, MatK, aligns with other RTs only across motifs 5 to 7, with the upstream sequence being unalignable with motifs 0 to 4; however, domain X sequence is clearly conserved, suggesting the maintenance of the maturase function [27,44]. MatK has been shown biochemically to bind to multiple chloroplast IIA introns,

supporting the hypothesis that it has evolved a more general maturase activity that facilitates splicing of multiple IIA introns in plant chloroplasts [70,72].

In bacteria, degenerations of the IEP sequences are rare because the great majority of non-truncated intron copies are active retroelement forms. The only known example is O.i.I2 of *Oceanobacillus iheyensis*, which encodes an IEP of the ML class that lacks the YADD and other motifs. The fact that the ORF has not accumulated stop codons suggests that it retains maturase activity, particularly because its exons encode the DNA repair protein RadC [50].

Introns with LAGLIDADG ORFs A small set of group II introns do not encode RT ORFs but instead encode proteins of the family of LAGLIDADG homing endonucleases (LHEs) and are presumably mobile through a distinct pathway that relies on the LHE (Figure 3D). LHEs in group II introns were first identified in several fungi, although an example has since been identified in the giant sulfur bacterium *Thiomargarita namibiensis* [73-76]. LHEs are a well-studied class of mobility proteins associated with group I introns, and they promote mobility by introducing double-stranded DNA breaks at alleles that lack the introns [2]. Consistent with this role, the LAGLIDADG ORFs in group II introns of the fungi *Ustilago* and *Leptographium* were shown biochemically to cleave intronless target sequences [77,78]. However, the IEP of *Leptographium* did not promote splicing of the host intron, as sometimes occurs for some group I intron-encoded LHEs [77,79]. To date, all identified LHE-encoding group II introns in both mitochondria and bacteria belong to the IIB1 subclass and are located in rRNA genes [73,80].

Introns without IEPs Group II introns without IEPs have lost retromobility properties and exist as splicing-only elements (Figure 3E). They are present in both bacteria and organelles but are especially prevalent in mitochondrial and chloroplast genomes [15]. For example, in plant angiosperms, there are approximately 20 ORF-less group II introns in each mitochondrial and chloroplast genome [70,71,81,82]. These plant organellar introns have been inherited vertically for over 100 million years of angiosperm evolution, consistent with their lack of a mobility-promoting IEP. Because the introns are situated in housekeeping genes in each organelle, efficient splicing is enabled by many splicing factors supplied by the host cells (below). In organellar genomes of fungi, protists, and algae, ORF-less group II introns are also common but less prevalent than in plants. Many of these introns contain remnants of IEP sequences, pointing to a sporadic and ongoing process of loss of the IEP and retromobility [53,83-86].

In bacteria, ORF-less group II introns are rare. Among the known examples, the ORF-less introns nearly always reside in genomes containing related introns whose IEPs may act in *trans* on the ORF-less introns [50]. Splicing function in *trans* has in fact been demonstrated experimentally for an IEP in a cyanobacterium [87]. The sole known exception to this pattern is the C.te.I1 intron in *Clostridium tetani*, for which no IEP-related gene is present in its sequenced genome. C.te.I1 self-splices robustly *in vitro*, and it was speculated that the intron might not require splicing factors *in vivo* [88,89]. This example lends plausibility to possibility that the ribozyme form of group II introns may exist and evolve in bacteria apart from the retroelement form; however, this would be rare because C.te.I1 is the only example of this type among over 1,500 known copies of group II introns in bacteria [90].

Introns with 'degenerated' ribozymes Many group II introns in mitochondria and chloroplasts have defects in conserved ribozyme motifs, such as mispaired DV or DVI helices or large insertions or deletions in catalytically important regions (Figure 3F) [15,44,71,91,92]. For such introns, secondary structure prediction with confidence is difficult or impossible, and these introns have presumably lost the ability to self-splice. Consistent with this inference, no plant mitochondrial or chloroplast group II intron has been reported to self-splice *in vitro*.

For introns with compromised ribozyme structures, splicing relies heavily on host-encoded splicing factors [71,93,94]. The catalogue of host-encoded factors is diverse and organism-specific. In yeast mitochondria, the ATP-dependent helicase MSS116 is a splicing factor for multiple self-splicing group I and group II introns [95]. In plant mitochondria and chloroplasts, an array of nuclear-encoded splicing factors has been identified [71,94,96]. Splicing in chloroplasts involves at least 16 proteins that contain motifs of five families of RNA-binding motifs (CRM, PPR, APO, PORR, and TERF families). Some splicing factors (for example, CRS1) are specific to a single chloroplast intron (*atpF*I1), whereas others (for example, CFM2, MatK) aid in splicing multiple introns, which are usually structurally related [97-100]. The situation is similar in mitochondria, where 11 proteins have been identified [71,101]. Additionally, there are four nuclear-encoded, IEP-derived maturases (nMat-1a, nMat-1b, nMat-2a, nMat-2b) that are imported into organelles and are involved in splicing of multiple mitochondrial and possibly chloroplast introns [71,102-105].

These examples illustrate that group II introns have repeatedly lost their splicing capability in organelles. To compensate, cellular splicing factors have evolved independently in different organisms to enable efficient splicing of the introns that lie in housekeeping genes. Similar

to the case of ORF-less group II introns, there has been a conversion from retromobility to splicing-only function, and splicing is under the control of the host nuclear genome.

Group III introns The most extreme examples of degenerated RNA structures are group III introns, found in *Euglena gracilis* chloroplasts (Figure 3G) [106]. These introns are approximately 90 to 120 nt in length and sometimes contain only DI and DVI motifs. *Euglena* chloroplasts are replete with >150 group III and degenerated group II introns, many located in essential genes. Because group III introns lack a DV structure, it is thought that a generalized machinery consisting of *trans*-acting RNAs and/or proteins facilitate their excision from cellular mRNAs.

Trans-splicing introns Some group II intron sequences in plant mitochondria and chloroplasts have been split through genomic rearrangements into two or more pieces that are encoded in distant segments of the genome (Figure 3H) [71,107,108]. The intron pieces are transcribed separately and then associate physically to form a tertiary structure that resembles a typical group II intron. The majority of *trans*-splicing introns are split into two pieces with the break point located in DIV. However, the *Oenethera nad5*I3 and *Chlamydomonas psaA*I1 are tripartite, containing breaks in both DI and DIV [108,109]. These and other *trans*-splicing introns require multiple splicing factors for efficient processing. In the case of *psaA*I1 in *Chlamydomonas reinhardtii* chloroplasts, as many as twelve proteins are required in the *trans*-splicing reaction [110,111]. For some introns, the evolutionary timing of the genomic rearrangement can be specified. The *nad1*I1 intron is *cis*-splicing in horsetail, but *trans*-splicing in fern and angiosperms, indicating that the genomic rearrangement occurred after horsetail split from the fern/angiosperm lineage over 250 million years ago [112,113]. No *trans*-splicing introns have yet been reported in bacteria.

Altered 5′ and 3′ splice sites While the vast majority of group II introns splice at specific junction sequences at the boundaries of the introns (5′GUGYG…AY3′), a number of group II introns have attained plasticity that allows them to splice at other points (Figure 3I). A set of fungal rRNA introns was identified that splice 1 to 33 nt upstream of the GUGYG motif. The alteration in splicing property was attributed to specific ribozyme structural changes, including an altered IBS1-EBS1 pairing, and loss of the EBS2 and branch site motifs [74]. These changes were inferred to have evolved independently multiple times. All of the introns are of the IIB1 subclass and the majority encodes a LAGLIDADG

IEP [74]. Interestingly, a similar situation was found for the bacterial intron C.te.I1 of *C. tetani*, which exhibits analogous structural deviations and splices eight nucleotides upstream of the GUGYG motif [89]. Alterations of the 3′ splice site have also been reported. About a dozen class B introns are known that contain insertions at the 3′ end of the intron, called domain VII, which result in a shift of splicing to approximately 50 to 70 nt downstream of the canonical 3′AY boundary sequence at the end of domain VI (Figure 3J) [114-116].

Alternative splicing The fact that group II introns can utilize 5′ and 3′ splice sites separated from the 5′ GUGYG and AY3′ sequences allows for the possibility of alternative splicing. The first report of this was in *Euglena* chloroplasts, where several group III introns spliced *in vivo* using noncognate 5′ or 3′ splice sites [117,118]. The frequencies of these splicing events, however, were low, being detected by RT-PCR, and the resultant proteins were truncated due to frame shifts and stop codons, which together raise the possibility that this is a natural error rate in splicing rather than regulated alternative splicing *per se*.

In bacteria, alternative splicing at the 3′ splice site was found for B.a.I2 of *Bacillus anthracis*. In that case, two *in vivo*-utilized sites are located 4 nt apart (each specified by a γ-γ′ and IBS3-EBS3 pairing), which result in two protein products, one consisting of the upstream exon ORF alone and the other a fusion of upstream and downstream ORFs [119]. In a more dramatic example, the *C. tetani* intron C.te.I1 utilizes four 3′ splice sites, each specified by a different DV/VI repeat. Each resulting spliced product is a distinct fusion protein between the 5′ exon-encoded ORF and one of four downstream exon-encoded ORFs [88]. The latter example resembles alternative splicing in eukaryotes because several protein isoforms are produced from a single genetic locus (Figure 3K).

Twintrons A twintron is an intron arrangement in which one group II intron is nested inside another intron as a consequence of an intron insertion event (Figure 3L). For a twintron to splice properly, often the inner intron must be spliced out before the outer intron RNA can fold properly and splice [118,120,121]. Twintrons are common in *Euglena* chloroplasts where they were first described, and where approximately 30 of its 160 introns are in twintron arrangements [106]. Several twintrons are known in bacteria; however, splicing of these twintrons does not appear to greatly impact cellular gene expression, because the twintrons are intergenic or outside of housekeeping genes [51,122]. Twintrons in the archaebacterium *Methanosarcina acetivorans* have a particularly complex arrangement [123]. There are up to five introns in a nested configuration but no coding ORFs in the flanking exons. Based on

the boundary sequences of the introns, it can be concluded that the introns have undergone repeated cycles of site-specific homing into the sequences of other group II introns. These repeated insertions are balanced by deletions of intron copies through homologous recombination. For these introns, the twintron organizations do not affect host gene expression but provide a perpetual homing site in the genome for group II introns.

Molecular phylogenetic evidence for the evolution of group II introns

While there has been much speculation about intron evolution, it remains difficult to obtain direct evidence for specific models. For group II introns, clear phylogenetic conclusions can only be drawn when analyzing closely related introns. This is because only closely related sequences allow the extensive alignments needed for robust phylogenetic signals. Such analyses have indicated multiple cases of horizontal transfers among organisms. Some of the inferred examples are as follows: from an unknown cyanobacterial source to *Euglena* chloroplasts [124]; from unknown sources into a cryptophyte (red alga; *Rhodomonas salina*) [125] or a green alga (*Chlamydomonas*) [126]; between mitochondrial genomes of diatoms and the red alga *Chattonella* [127]; and from the mitochondrion of an unknown yeast to *Kluyveromyces lactis* [127,128]. In bacteria, it was concluded that group II introns from multiple classes have transferred horizontally into *Wolbacchia* endosymbionts, because the resident introns are of different classes [129]. More broadly, horizontal transfers among bacteria appear to be relatively common because many bacteria contain introns of multiple classes [51,130,131].

Beyond identification of horizontal transfers, unfortunately, global phylogenetic analyses result in poor phylogenetic signals because the number of characters available (that is, those that are unambiguously alignable for all introns) decrease to at most approximately 230 aa for the ORF and approximately 140 nt for the RNA [57]. With such reduced-character data sets, clades are clearly identified in bacteria corresponding to classes A, B, C, D, E, F, ML, and CL [28,50,56,132]; however, relationships among the clades are not well supported. Notably, when IEPs of organellar introns are included in trees along with bacterial introns, the organellar IEPs cluster with the ML and CL clades of bacteria, indicating that introns of mitochondrial and chloroplast genomes originated from the ML and CL lineages of bacteria [28]. A global analysis with all known organellar and bacterial intron IEPs is not possible because of extreme sequence divergence of many organellar introns.

The limited phylogenetic resolution for group II introns was attributed to several potential factors [57]. First, the amino acid data sets had substantial levels of saturation (that is, repeated changes per amino acid), which decreased the signal-to-noise ratio. Second, the sequences of some clades had extreme base composition biases that could distort the results (for example, GC-rich genomes have biased amino acid composition that can cause artifacts; this is especially true for class B introns). In addition, there were problematic taxon-sampling effects (differences in trees depending on which intron sequences were included). These complications underscore the difficulty of obtaining rigorous evidence for the evolution of group II introns and the need for exercising caution in drawing interpretations and conclusions. In the future, identifying the basis for these effects may allow for compensation and optimization that may produce more satisfying conclusions.

Coevolution of ribozyme and IEP and the retroelement ancestor hypothesis

Over a decade ago, it was noticed that there is a general pattern of coevolution among group II intron IEPs and their RNA structures [53,133]. Specifically, each phylogenetically supported IEP clade corresponds to a distinct RNA secondary structure. Coevolution of RNA and IEP should not be surprising given the intimate biochemical interactions between ribozyme and protein during the splicing and mobility reactions. However, coevolution clearly has not occurred for group I ribozymes and their IEPs. Group I introns have been colonized by four families of IEPs, and there is evidence for a constant cycle of ORF gain and loss from group I ribozymes [134-137].

The principle of coevolution is a central principle to deciphering the history of group II introns. Importantly, it simplifies the reconstruction from two independent histories to a single history. Based on the pattern of coevolution, a model was set forth to explain the history of group II introns, which was called the retroelement ancestor hypothesis [53,133]. The model holds that group II introns diversified into the major extant lineages as retroelements in bacteria, and not as independent ribozymes. Subsequently, the introns migrated to mitochondria and chloroplasts, where many introns became splicing-only elements.

Phylogenetic analyses have in general supported the initial observation of coevolution, because both RNA and IEP trees define the same clades of introns, thereby excluding extensive exchanges between ribozymes and the different classes of IEPs [57]. However, caveats remain. The most obvious one is the fact that some group II introns encode LHE proteins rather than RT proteins. The invasion of group II ribozymes by LHE's occurred at least once in bacteria and multiple times in fungal mitochondria [74,76]. So far, these exceptions are limited in number and do not significantly undermine the overall pattern of coevolution. A second caveat comes from topology tests between the IEP and RNA trees which indicated a conflict [57] (topology tests are mathematical

techniques for evaluating and comparing different trees). As noted in that study, the conflict could be explained by either discordant evolution (reassortment of IEPs and ribozymes) or convergence of RNA or IEP sequences that masks their true evolutionary relationships. While the source of the conflict was not resolved, more recent data support the latter reason (L. Wu, S. Zimmerly, unpublished).

A model for the evolution of group II introns

Diversification within Eubacteria The retroelement ancestor model continues to be consistent with available data and is elaborated here to show how it can explain the emergence of the known forms and distribution of group II introns (Figure 4). The ancestral group II intron is hypothesized to have been a retroelement in Eubacteria that consisted of a ribozyme and intron-encoded RT component and had both mobility and self-splicing properties. The earliest introns would have behaved as selfish DNAs [49], which then differentiated in Eubacteria into several retroelement lineages (A, B, C, D, E, F, ML, CL). The IEP initially would have consisted of a simple RT, similar to RTs of classes C, D, E, and F, while the En domain was acquired subsequently from H-N-H nucleases present in Eubacteria [30,58]. The En domain would have provided the benefit of enhanced mobility properties and/or allowed the introns to exploit new biological niches.

Of the three target specificities known for bacterial introns (insertion into homing sites, after terminator motifs, and into *attC* sites) [64,65], any of these specificities could have been used by the ancestor, although homing is by far the most prevalent specificity, occurring for all lineages but class C. Horizontal transfers would have driven the dissemination of group II introns across species. Some group II introns took up residence in housekeeping genes, particularly in cyanobacteria and for CL and ML lineages [51,138,139]. These introns would have had to splice efficiently to avoid inhibiting expression of the host genes. Limited numbers of introns deviated from the 'standard' retroelement form, including ORF-less introns, introns with degenerate IEPs, twintrons, and alternatively splicing introns. Most of these lost mobility properties but maintained splicing ability. Some introns adapted altered mechanisms of 5′ and 3′ exon recognition and altered 5′ or 3′ intron termini [71,72,74,89,116,117,119,123].

Migration to archaebacteria and organelles Introns belonging to the lineages CL, D, and E migrated from Eubacteria to archaebacteria [51,123]. The direction of migration can be inferred from the lower number and diversity of introns in archaebacteria compared to Eubacteria. Introns of the CL and ML lineages migrated

from Eubacteria to mitochondria and chloroplasts. The introns could have been contained within the original bacterial endosymbionts that produced each organelle or been introduced by subsequent migrations. Horizontal transfers of introns among mitochondrial and chloroplast genomes created a diversity of IIA and IIB introns in both organellar genomes [124-128].

Diversification within organelles Within mitochondria and chloroplasts, the character of group II introns changed to become more genomically stable and less selfish. The introns took up residence in housekeeping genes, which necessitated efficient splicing, and which was enabled by host-encoded splicing factors [71,93-96]. While many group II introns maintained retromobility, many more degenerated in their RNA and/or IEP structures or lost the IEPs entirely, leading to immobile introns. In plants, the introns proliferated greatly to copy numbers of approximately 20 per organelle, with nearly all IEPs being lost. At least two IEPs migrated from the plant mitochondrial genome to the nucleus to encode four splicing factors that are imported to the mitochondria and possibly chloroplasts for organellar intron splicing [71,85].

In fungi, a small fraction of ORF-less introns acquired an IEP of the LAGLIDADG family, which permitted mobility through the homing endonuclease mechanism. In mitochondria and chloroplasts, introns sporadically became *trans*-splicing due to genomic rearrangements that split intron sequences [71,107-109,112,113]. In *Euglena* chloroplasts, the introns degenerated on a spectacular scale to become group III introns. The earliest euglenoids are inferred to be intron-poor while the later branching euglenoids harbor more introns, pointing to a process of intron proliferation within *Euglena* chloroplasts [140,141].

Caveats It should be kept in mind that this model is contingent upon the available sequence data. One cautionary note is that our picture of group II introns in bacteria may be skewed, because for the data available the introns were identified bioinformatically in genomes based on the RT ORF. This may result in some oversight of ORF-less group II introns; however, the numbers of those introns do not appear to be large. In a systematic search of bacterial genomes for domain V motifs, nearly all introns identified were retroelement forms [50]. There was one example uncovered of a group II intron with a degenerate IEP, and only a few ORF-less introns, all in genomes with closely related introns where an IEP may act in *trans* on the ORF-less intron. A single independent, ORF-less group II intron was found out of 225 genomes surveyed. Hence, it seems safe to predict that relatively few ORF-less introns have been overlooked in
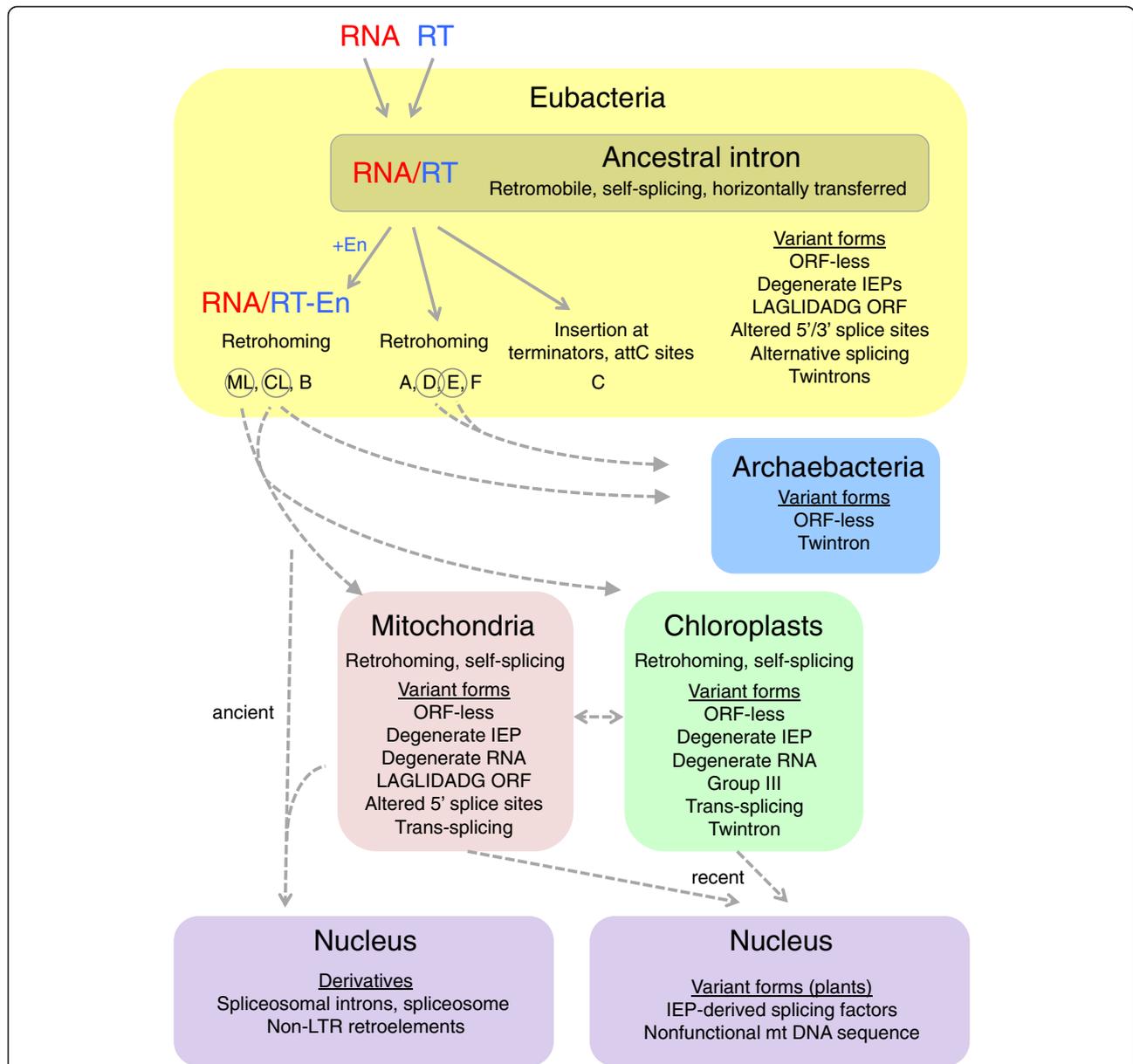
Figure 4 Global model for group II intron evolution. An ancient reverse transcriptase combined with a structured RNA to form a group II intron retroelement. This ancestral form was present in Eubacteria and had properties of splicing and retromobility. The retroelement form differentiated into eight lineages, of which ML, CL, and B acquired an endonuclease domain. All lineages but class C (IIC) introns were mobile by retrohoming into site-specific target sequences. Introns of three lineages transferred to archaebacteria, while introns of two lineages transferred to mitochondria and chloroplasts. Variant forms of group II introns were produced in each location as noted. Prior to the LECA, group II introns invaded the nucleus where they developed into the spliceosome and non-LTR retroelements. Much later in plants, group II introns transferred to the nucleus, where the IEPs developed into splicing factors that are imported into mitochondria and/or chloroplasts to help splice organellar group II introns. See text for full description. IEP, intron-encoded protein; LTR, long terminal repeat; ORF, open reading frame; RT, reverse transcriptase.

bacteria, unless they have domain V structures unlike those of known group II introns.

Origin of group II introns

If the ancestor of extant group II introns was a retroelement, where did that retroelement come from? The simplest scenario is that pre-existing ribozyme and RT components combined into a single element, creating a new mobile DNA. An interesting alternative possibility is that a self-splicing RNA might have arisen at the boundaries of a retroelement to prevent host damage by the mobile DNA [142].

There are many potential sources for the ancestral RT component, because a myriad of uncharacterized RTs

exist in bacterial genomes, most of which could potentially correspond to forms that were co-opted by the primordial group II intron [143]. Because there is little evidence that bacterial RTs other than group II introns are proliferative elements, it is possible that the property of mobility emerged only after the RT became associated with the RNA component.

Similarly, there are many structured RNAs in bacteria that could have given rise to the ancestral group II ribozyme, including noncoding RNAs, riboswitches, or even a fragment of the ribosome [144-146]. The primordial RNA component would not necessarily have been self-splicing like modern group II introns, but upon associating with the RT, it would have generated a simple retroelement, which then became specialized and/or optimized to become the efficient retroelement that was then the ancestor of the different lineages. Although the topic of the ultimate origin of group II introns is interesting to consider, any model will be speculative.

Which class of modern group II introns best represents the ancestral group II intron retroelement? It is often claimed in the literature that IIC introns are the most primitive form of group II introns [13,14,18,147]. While this idea is consistent with the small size of IIC introns, it is only weakly supported by phylogenetic data. The study cited provides a posterior probability of only 77% in Bayesian analysis in support of the conclusion (and <50% with neighbor-joining or maximum parsimony methods), whereas 95% is the usual standard for making conclusions with Bayesian analysis [148]. In more recent phylogenetic analyses, IIC introns are also seen often as the earliest branching of group II introns, albeit with weak or inconsistent support [57]. Interestingly, additional classes of group II introns have been uncovered more recently in sequence data, and some of these are as good or better candidates for most ancestral intron (L. Wu, S. Zimmerly, unpublished).

Structural parallels between group II introns, spliceosomal introns and the spliceosome

Major parallels The concept that group II introns were the ancestors of spliceosomal introns emerged shortly after the discovery of multiple intron types (spliceosomal, group I, group II introns) [149-151]. Since then, mechanistic and structural evidence has accumulated to the point that few if any skeptics remain. This is a shift from the early years when it was argued that mechanistic constraints could have resulted in convergent evolution of mechanisms and features [152].

The major similarities and parallels for the two intron types are summarized here. In terms of splicing mechanisms, the overall pathways for group II and spliceosomal introns are identical, with two transesterifications and a lariat intermediate (Figure 2A). The chemistry of the two splicing steps share characteristics with regard to their sensitivities to Rp and Sp thiosubstitutions. A Rp thiosubstitution (that is, sulfur atom substituted for the Rp non-bridging oxygen) at the reacting phosphate group inhibits both steps of the reaction for both group II and spliceosomal introns, whereas Sp substitutions do not, suggesting that different active sites are used for the two reactions [153-156]. This contrasts with data for group I introns, for which Rp substitutions inhibited only the first splicing step, and Sp substitutions inhibited only the second step, which is consistent with reversal of a reaction step at a common active site [157,158]. The shared sensitivities for the reactions of group II and spliceosomal introns suggest that similar active sites are used for the two types of introns, with the group II-like active site being maintained during evolution of spliceosomal introns.

Structurally, there are many parallels between group II intron RNAs and spliceosomal snRNAs, which run the gamut from being clearly analogous to being speculative. The most obvious parallel is the branch site motif that presents the 2′OH of a bulged A to the 5′ splice site for the first step of splicing. For group II introns, the bulged A is contained within a helix of domain VI; in the spliceosome the same bulged structure is formed by the pairing of the U2 snRNA to the intron's branch point sequence (Figure 5) [159]. Intron boundary sequences are also quite similar and presumably function analogously, being 5′ GU-AY 3′ for group II introns and 5′ GU-AG 3′ for spliceosomal introns (Figure 5). The first and last nucleotides of each intron have been reported to form physical interactions that are essential for an efficient second step of splicing [160-162].

For group II introns, the active site is in domain V, with two catalytically important metal ions being coordinated by the AGC catalytic triad and the AY bulge [147]. A similar structure is formed in the spliceosome by pairings between the U2 and U6 snRNAs, which bear an AGC motif and AU bulge (Figure 5) [23]. The equivalence between the two active sites has been supported experimentally through the substitution of the DV sequence of a group II intron for the analogous positions in the snRNAs of the minor spliceosome (in that case the U12-U6atac snRNA pairing rather than U2-U6) [163]. The substitution demonstrates that the group II intron sequence can assume a functional structure at the putative active site of the spliceosome. More recently, the equivalence of the two active sites was taken to a new level using thiosubstitution and metal rescue experiments, in which a thiosubstitution inhibits a splicing step, but is rescued by metal ions that coordinate sulfur better than magnesium does. These experiments demonstrated that the AGC and bulged AU motifs of the U6-U2 active site coordinate catalytic metal ions as predicted from the crystal structure of the group IIC intron [164].
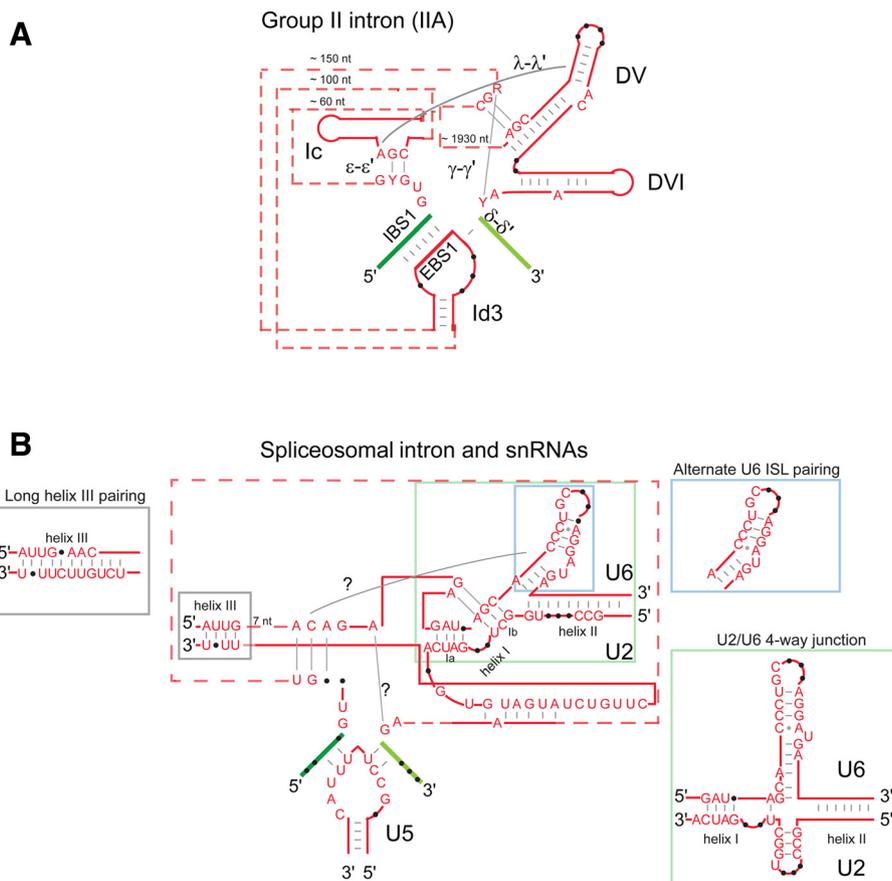
Figure 5 Structural comparison of group II introns, spliceosomal introns, and snRNAs. (A) Group IIA intron. EBS, exon-binding site; DV, domain V; DVI, domain VI; IBS, intron-binding site. **(B)** Pairings between U2, U5, and U6 snRNAs and the intron and exons. For both panels, intron sequences and snRNA sequences are shown in red, with exons shown in green. Base pairs are indicated by gray dashes and unpaired nucleotides as black dots. The size of sequences represented by dotted red lines are indicated in nucleotides. For group II introns, selected nucleotide positions critical for splicing are shown, while the sequences shown for snRNAs correspond to the 95% consensus for the U2, U5, and U6 snRNAs sequences present in Rfam [203]. The blue square inset shows an alternative secondary structure model for the ISL of U6, which is less compatible with DV of group II introns but is formed for naked snRNAs. The green square indicates an alternative four-way junction structure, also formed by naked snRNAs. Question marks indicate the interactions found in group II introns for which no equivalent interactions are reported in snRNAs. See text for a full description.

A further active site parallel comes from the discovery in the group II crystal structure of a triple helix between the AGC base pairs in domain V and two bases of the J2/3 strand (Figure 5A) [147]. This structure is hypothesized to be recapitulated in the active site of the spliceosome, with an AG of the ACAG*A*GA motif forming the triple base pairs with the *AG*C of the U6-U2 helix (Figure 5B). Experiments for the yeast spliceosome using covariation-rescue and cross-linking methods support the hypothesized triple base pairs in the spliceosome and lend further support for this active site parallel [165].

A final clear parallel between group II introns and spliceosomal introns was revealed by the crystal structure of a portion of the Prp8 protein, a 280-kDa protein (in yeast) located at the heart of the spliceosome. A region of Prp8 cross-links to the 5′ and 3′ exons and also to the intron's branch site, indicating its proximity to the spliceosome's active site. Surprisingly, the crystal structure of a major portion of yeast Prp8 revealed that the cross-linking portion is composed of a reverse transcriptase domain fold [166]. In fact, the existence of an RT domain in Prp8 had been previously predicted correctly based on sensitive sequence pattern profiles [167]. Thus, the active site region of the spliceosome appears to contain remnants of both an ancestral ribozyme (snRNA pairings) and an ancestral group II RT (Prp8), which together strongly support the idea that the eukaryotic spliceosome and nuclear pre-mRNA introns are highly elaborate derivatives of ancient, retromobile group II introns.

Less clear yet plausible parallels Additional parallels between group II intron and spliceosomal intron RNAs

are credible but less clear. The loop 1 structure of U5 snRNA is predicted to be analogous the EBS1 loop of group II introns, a substructure that forms base pairs with the 5′ exon of group II introns, thereby delivering the 5′ exon to the active site (Figure 1A). Supporting the parallel, the loop 1 structure of U5 forms cross-links with both the 5′ and 3′ exon boundary sequences [168]. An experiment supporting functional equivalence demonstrated that the EBS1 stem-loop of the bI1 intron of yeast mitochondria could be deleted and then rescued with a stem-loop supplied in *trans* that had either the native bI1 stem-loop sequence or the loop 1 sequence of the U5 snRNA [169]. However, because the function of the EBS1 loop sequence is to form base pairs with the exon's IBS1, and the U5 loop sequence is fortuitously capable of base pairing with the IBS1 of bI1 (but not other group II introns), the significance of the experiment is less clear. Interestingly, while the EBS1 loop sequence of IIB and IIC introns pairs with only the 5′ exon, the EBS1 loop of IIA introns pairs with both 5′ and 3′ exons (IBS1-EBS1 and δ-δ′ interactions; Figure 1), making the putative parallel more similar for IIA introns than for IIB or IIC introns [170].

The 2-bp ε-ε′ interaction of group II introns has been proposed to be equivalent to an experimentally detected pairing between the U6 snRNA and a sequence near the 5′ end of the intron (Figures 1 and 5) [12,171-173]. While the analogy is reasonable, the U6 pairing was initially reported as 3 bp and later evidence suggested it to be up to 6 bp [174,175]; it remains unclear whether or to what extent the two pairings are analogous structurally and functionally.

Finally, the λ-λ′ interaction of group II introns is a three-way interaction that connects the ε-ε′ interaction (and hence the 5′ end of the intron) to the distal stem of domain V (Figures 1 and 5). The parallel in snRNAs is proposed to be a triple base pair between a subset of nucleotides in the ACAGAGA motif and the internal stem-loop (ISL) helix of U6. While this structural parallel remains a possibility, it appears difficult for the ACAGAGA motif to simultaneously form the ε-ε′-like and λ-λ′-like interactions.

Missing or questionable structural parallels It is important not to ignore features that are not shared between group II and spliceosomal introns, in the rush to pronounce the two types of introns equivalent. Each type of intron has features not found or reported in the other. For example, the γ-γ′ interaction of group II introns is a Watson-Crick base pair between a J2/3 nucleotide and the last position of the intron, but it has not been reported for spliceosomal introns (Figures 1 and 5). The putatively equivalent nucleotides in the

snRNAs would be a residue of the ACA*G*AGA box and the last nucleotide (G) of the intron.

Two critical pairings that occur in the spliceosome but not in group II introns are temporal pairings formed during spliceosome assembly but not catalysis [176]. The U1 snRNA pairs to the 5′ end of the intron during splice site recognition and assembly, only to be replaced before catalysis by a pairing between U6 and the 5′ end of the intron. Similarly, the extensive pairings between the U6 and U4 snRNAs occur during spliceosome assembly but are disrupted and replaced by the U6-U2 pairing. Both of these transient RNA-RNA pairings can be predicted to have arisen during the evolutionary advent of the spliceosome, for the purposes of assembly and/or regulation.

On the other hand, Helices Ia and III of the U2-U6 structure (Figure 5) occur during catalysis, but have no equivalent in group II introns, and perhaps even conflict with the structural organization of group II intron RNAs. Helix Ia introduces a spacer between the catalytic AGC motif, the branch site motif and triple helix motif, potentially introducing a structural incompatibility between spliceosomal and group II introns. In any case, group II introns do not have an equivalent helix Ia structure. More problematic is Helix III, which is not present in group II introns, and appears to conflict with proposed structural parallels for the ACAGAGA sequence. In [175], it was proposed that helix III is shortened to approximately 4 bp during catalysis, but might form more fully during assembly. Again, because this established helix has no group II intron equivalent, it may have originated during evolution of the spliceosome.

A modest discrepancy involves the secondary structure of the ISL of U6 and the DV structure of group II introns. The secondary structure of the ISL is usually drawn with an AU bulge opposite an unpaired C (blue square, Figure 5) [177]. However, chemical modification protection data with purified, activated spliceosomes instead suggested an alternative structure that is more similar to group II introns. The alternative structure does not form for naked snRNAs, but it may form in the context of the spliceosome [163,175]. Another perplexing difference between intron types is the break of the catalytic helix into helices 1b and the ISL.

Finally, it is notable that secondary structure models for snRNA pairings have changed over the years, and there are proposed differences in snRNA pairings for yeast versus mammalian snRNAs, despite the fact that the relevant sequences are identical [178-182]. NMR structural analysis of the naked U2-U6 sequences revealed a four-way junction structure (Figure 5B) [180], which was subsequently supported by genetic data in yeast [183]. The four-way junction was proposed to form for the first step, with the three-way junction forming for the second step. However, there is no evidence for

the four-way junction structure in the mammalian spliceosome, most recently based on RNA modification protection data of purified, activated U5-U6-U2 spliceosomes [175].

The pathway for the evolution of spliceosomal introns from group II introns

Because virtually all eukaryotic genomes contain introns and spliceosomes, with the few exceptions attributed to losses [184-186], the spliceosome was necessarily present in the last eukaryotic common ancestor (LECA). Thus, evolution of ancestral group II introns to the spliceosome would have occurred prior to the LECA. Evidence from genome comparisons indicates that the LECA contained a multitude of introns [187]. Indeed, it is doubtful that such a complex machinery as the spliceosome would have arisen on account of a few introns.

Models for the conversion of group II introns to the spliceosome are not well refined, and multiple scenarios are possible [188-191]. At some point prior to the LECA, group II introns likely invaded the nuclear genome and proliferated as mobile DNAs. The invading group II intron(s) could have come from the genome of the alpha-proteobacterium that became the mitochondrial endosymbiont or alternatively could have been transferred from a bacterium to the nuclear genome after establishment of the mitochondrion. Rampant intron propagation would leave many introns interrupting essential genes, which would require the maintenance of splicing to ensure cell viability. Consequently, the cell evolved splicing factors to facilitate and eventually control splicing of the introns. Debilitating mutations in ribozyme sequences would occur easily through point mutations, leading to many copies of splicing-deficient introns in the genome. On the other hand, discarding such defective introns by precise deletions of entire introns would be rare. The cell could have solved this problem by evolving a general splicing machinery that acts in *trans*, leaving the introns free to lose all their ribozyme structures except for certain boundary sequences. The end result was the transfer of splicing catalysis from individual ribozyme units scattered throughout the genome to a single *trans*-acting RNP machinery that could act on all intron copies.

Because the modern spliceosome is ostensibly a elaborate derivative of a mobile group II intron RNP, it follows that at a time point prior to the LECA, the ribozyme structure of group II introns fragmented into the U2, U5, and U6 snRNA components of the spliceosome. In addition, the RT protein expanded in length through domain accretion, with the fusion of an RNase H domain, MPN/JAB1 (nuclease) domain, and possibly other domains that form portions of the modern 280-kDa Prp8 protein [167,192]. Additional protein splicing factors such as Sm and SR proteins were incorporated into the spliceosomal machinery.

The U1 and U4 snRNAs and snRNPs were added as new regulatory or facilitating activities, since they do not have equivalents in group II introns.

One intriguing model for the emergence of the spliceosome predicts that proliferation of mobile group II introns was the driving force for invention of the nuclear membrane [188,193]. The model is based on the likelihood that splicing would have been slow compared to transcription and translation processes. In an uncompartmentalized cell, translation would therefore occur before mRNAs were fully spliced, yielding nonfunctional proteins. By separating transcription and translation, the nuclear membrane ensured that only fully spliced transcripts were translated.

Several studies have experimentally addressed evolutionary issues of group II introns. One series of studies sought to reproduce the fragmentation of a group II ribozyme into a *trans*-splicing intron-in-pieces. It was shown that a retromobile IIA intron could be split into multiple functional *trans*-splicing RNA transcripts, with the break points distributed throughout the sequence and not only in domain IV as occurs for nearly all natural *trans*-splicing introns [189,194,195]. In a separate series of studies, the question was addressed as to why group II introns do not function optimally in nuclear genomes, where they are apparently excluded in functional form in nature. It was found that the introns spliced in the cytoplasm rather than the nucleus and that transcripts were subject to nonsense-mediated decay (NMD) and poor translation. Further dissection showed that transcripts were mislocalized to foci in the cytoplasm and that the excised intron lariat formed RNA-RNA pairings with spliced mRNAs that inhibited their translation. It was inferred that these phenomena demonstrate an incompatibility of group II introns with eukaryotic cellular organization and may have been responsible for the ejection of group II introns from nuclear genomes during evolution [190,196,197].

What other elements did group II introns evolve into?

In addition to spliceosomal introns, group II introns are believed to be the ancestors of non-LTR retroelements, a major class of mobile DNAs in eukaryotes [31]. The RTs of group II introns and non-LTR retroelements are related phylogenetically and share sequence motifs 0 and 2a, which are absent from other RTs except diversity-generating retroelements (DGRs) (2a), retroplasmids (2a), and possibly retrons (2a) [143,191,198,199]. Moreover, the retromobility mechanisms of group II and non-LTR elements are similar, with both called target-primed reverse transcription because they involve cleavage of the DNA target to produce a primer for reverse transcription [31,200]. As mobile group II introns were present in the nucleus prior to the LECA, it is plausible that some invading group II introns produced the non-LTR family

retroelements in the nucleus through the loss of their ribozyme and splicing functions but retention of mobility functions.

Moreover, it is clear that group II introns spawned other RT-containing units. A subset of CRISPR/Cas elements contain an RT gene, either as a free-standing ORF or fused to a *cas1* gene (denoted G2L1 and G2L2 (group II-like 1 and 2) [143,201]). By sequence, these RTs might be mistaken for group II introns except that no ribozyme RNA structure is present [143]. The *cas1* gene encodes a nuclease that helps integrate short sequences of phage or plasmid into CRISPR arrays, lending cellular immunity to DNAs containing those sequences [202]. The RT genes found within CRISPR/Cas systems are almost certainly derived from group II intron retroelements due to their close sequence similarity. It seems likely that they use a mechanism related to TPRT to integrate the new protospacer sequences into CRISPR arrays.

Three additional types of group II-related RTs exist in bacteria, denoted G2L3, G2L4, and G2L5 [143]. These are not associated with CRISPR/Cas systems and also lack ribozyme structures. It is unknown whether these RTs are part of mobile DNAs or participate in as yet unidentified functions.

Conclusions

Group II introns are compact and versatile retroelements that have successfully colonized genomes across all domains of life and have given rise to many variant forms. Current data are consistent with the model that the retroelement form (that is, the form diagrammed in Figure 1) was the ancestor of extant group II introns and was the driver for their spread and survival. The evolutionary success of group II introns may be linked to the multifunctionality of their splicing and mobility reactions, which allowed them to spread as selfish DNAs, and then derivatize into adaptable forms that shed either splicing or mobility properties. Interestingly, there is much overlap in variant forms of group II introns found in bacterial and organellar genomes (ORF-less introns, twintrons, altered 5′ splice sites, alternative splicing, degenerate IEP sequences, LAGLIDADG IEPs; Figure 4), which suggests that these derivative forms represent general ways that group II introns can differentiate. The low numbers of derivatives in bacteria suggest that the nonmobile derivatives do not persist long in bacterial genomes, whereas derivatized introns in organelles may persist indefinitely as splicing-only elements, and potentially provide benefits of gene regulation through nuclear control of their splicing.

With regard to the evolutionary pathway of group II introns into spliceosomal introns, important insights over the past 2 years have largely erased doubts about the long-standing hypothesis that the spliceosome descended from group II introns. Indeed, there are no credible competing hypotheses for the origin of the spliceosome. Still, the specifics of the pathway and the full scope of mechanistic parallels remain to be resolved. Additional insight may be forthcoming from structural elucidations of the spliceosome and comparisons to group II intron structures, as well as genomic comparisons of early branching eukaryotes, which may give information about introns in the LECA and potentially suggest evolutionary intermediates or pathways. Overall, the elucidation of group II intron biology, structure, and evolution remains an important facet in understanding the evolution and dynamics of eukaryotic genomes.

Abbreviations
D: DNA endonuclease domain of the group II intron-encoded protein; DI-DVI: Group II intron domains I-VI; EBS: Exon-binding site; IBS: Intron-binding site; IEP: Intron-encoded protein; LECA: Last eukaryotic common ancestor; LHE: LAGLIDADG homing endonuclease; ORF: Open reading frame; RT: Reverse transcriptase domain of the group II intron-encoded protein; TRPT: Target-primed reverse transcription; X: Maturase domain of the group II intron-encoded protein.

References
1. Bonen L, Vogel J. The ins and outs of group II introns. Trends Genet. 2001;17:322–31.
2. Belfort M, Derbyshire V, Parker MM, Cousineau B, Lambowitz AM. Mobile introns: pathways and proteins. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington, DC: ASM Press Publishers; 2002. p. 761–83.
3. Lehmann K, Schmidt U. Group II introns: structure and catalytic versatility of large natural ribozymes. Crit Rev Biochem Mol Biol. 2003;38:249–303.
4. Zimmerly S. Mobile introns and retroelements in bacteria. In: Mullany P, editor. The dynamic bacterial genome. Cambridge University Press; 2005. p. 121-150.
5. Toro N, Jimenez-Zurdo JI, Garcia-Rodriguez FM. Bacterial group II introns: not just splicing. FEMS Microbiol Rev. 2007;31:342–58.
6. Lambowitz AM, Zimmerly S. Group II introns: mobile ribozymes that invade DNA. Cold Spring Harb Perspect Biol. 2011;3:a003616.
7. Knoop V, Brennicke A. Promiscuous mitochondrial group II intron sequences in plant nuclear genomes. J Mol Evol. 1994;39:144–50.
8. Lin X, Kaul S, Rounsley S, Shea TP, Benito MI, Town CD, et al. Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana. Nature. 1999;402:761–8.
9. Michel F, Ferat JL. Structure and activities of group II introns. Annu Rev Biochem. 1995;64:435–61.
10. Lambowitz AM, Zimmerly S. Mobile group II introns. Annu Rev Genet. 2004;38:1–35.
11. Fedorova O, Zingler N. Group II introns: structure, folding and splicing mechanism. Biol Chem. 2007;388:665–78.
12. Michel F, Costa M, Westhof E. The ribozyme core of group II introns: a structure in want of partners. Trends Biochem Sci. 2009;34:189–99.

13. Pyle AM. The tertiary structure of group II introns: implications for biological function and evolution. Crit Rev Biochem Mol Biol. 2010;45:215–32.

14. Marcia M, Somarowthu S, Pyle AM. Now on display: a gallery of group II intron structures at different stages of catalysis. Mob DNA. 2013;4:14.

15. Michel F, Umesono K, Ozeki H. Comparative and functional anatomy of group II catalytic introns - a review. Gene. 1989;82:5–30.

16. Pyle AM, Lambowitz AM. Group II introns: ribozymes that splice RNA and invade DNA. In: Gesteland RF, Cech TR, Atkins JF, editors. The RNA world. Cold Springs Harbor, New York: Cold Springs Harbor Laboratory Press; 2006. p. 469–505.

17. Toor N, Rajashankar K, Keating KS, Pyle AM. Structural basis for exon recognition by a group II intron. Nat Struct Mol Biol. 2008;15:1221–2.

18. Robart AR, Chan RT, Peters JK, Rajashankar KR, Toor N. Crystal structure of a eukaryotic group II intron lariat. Nature. 2014;514:193–7.

19. Michel F, Jacquier A. Long-range intron-exon and intron-intron pairings involved in self-splicing of class II catalytic introns. Cold Spring Harb Symp Quant Biol. 1987;52:201–12.

20. Jacquier A, Michel F. Base-pairing interactions involving the 5′ and 3′-terminal nucleotides of group II self-splicing introns. J Mol Biol. 1990;213:437–47.

21. Costa M, Michel F, Westhof E. A three-dimensional perspective on exon binding by a group II self-splicing intron. EMBO J. 2000;19:5007–18.

22. Konforti BB, Abramovitz DL, Duarte CM, Karpeisky A, Beigelman L, Pyle AM. Ribozyme catalysis from the major groove of group II intron domain 5. Mol Cell. 1998;1:433–41.

23. Gordon PM, Piccirilli JA. Metal ion coordination by the AGC triad in domain 5 contributes to group II intron catalysis. Nat Struct Biol. 2001;8:893–8.

24. Schmelzer C, Schweyen RJ. Self-splicing of group II introns *in vitro*: mapping of the branch point and mutational inhibition of lariat formation. Cell. 1986;46:557–65.

25. van der Veen R, Arnberg AC, van der Horst G, Bonen L, Tabak HF, Grivell LA. Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing *in vitro*. Cell. 1986;44:225–34.

26. Michel F, Lang BF. Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. Nature. 1985;316:641–3.

27. Mohr G, Perlman PS, Lambowitz AM. Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. Nucleic Acids Res. 1993;21:4991–7.

28. Zimmerly S, Hausner G, Wu X. Phylogenetic relationships among group II intron ORFs. Nucleic Acids Res. 2001;29:1238–50.

29. Blocker FJ, Mohr G, Conlan LH, Qi L, Belfort M, Lambowitz AM. Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. RNA. 2005;11:14–28.

30. Gorbalenya AE. Self-splicing group I and group II introns encode homologous (putative) DNA endonucleases of a new family. Protein Sci. 1994;3:1117–20.

31. Zimmerly S, Guo H, Perlman PS, Lambowitz AM. Group II intron mobility occurs by target DNA-primed reverse transcription. Cell. 1995;82:545–54.

32. San Filippo J, Lambowitz AM. Characterization of the C-terminal DNA-binding/DNA endonuclease region of a group II intron-encoded protein. J Mol Biol. 2002;324:933–51.

33. Jimenez-Zurdo JI, Garcia-Rodriguez FM, Barrientos-Duran A, Toro N. DNA target site requirements for homing *in vivo* of a bacterial group II intron encoding a protein lacking the DNA endonuclease domain. J Mol Biol. 2003;326:413–23.

34. Moran JV, Mecklenburg KL, Sass P, Belcher SM, Mahnke D, Lewin A, et al. Splicing defective mutants of the COXI gene of yeast mitochondrial DNA: initial definition of the maturase domain of the group II intron al2. Nucleic Acids Res. 1994;22:2057–64.

35. Wank H, SanFilippo J, Singh RN, Matsuura M, Lambowitz AM. A reverse transcriptase/maturase promotes splicing by binding at its own coding segment in a group II intron RNA. Mol Cell. 1999;4:239–50.

36. Matsuura M, Noah JW, Lambowitz AM. Mechanism of maturase-promoted group II intron splicing. EMBO J. 2001;20:7259–70.

37. Noah JW, Lambowitz AM. Effects of maturase binding and Mg2+ concentration on group II intron RNA folding investigated by UV cross-linking. Biochemistry. 2003;42:12466–80.

38. Cui X, Matsuura M, Wang Q, Ma H, Lambowitz AM. A group II intron-encoded maturase functions preferentially in cis and requires both the reverse transcriptase and X domains to promote RNA splicing. J Mol Biol. 2004;340:211–31.

39. Saldanha R, Chen B, Wank H, Matsuura M, Edwards J, Lambowitz AM. RNA and protein catalysis in group II intron splicing and mobility reactions using purified components. Biochemistry. 1999;38:9069–83.

40. Guo H, Zimmerly S, Perlman PS, Lambowitz AM. Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. EMBO J. 1997;16:6835–48.

41. Cousineau B, Smith D, Lawrence-Cavanagh S, Mueller JE, Yang J, Mills D, et al. Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. Cell. 1998;94:451–62.

42. Mohr G, Smith D, Belfort M, Lambowitz AM. Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. Genes Dev. 2000;14:559–73.

43. Singh NN, Lambowitz AM. Interaction of a group II intron ribonucleoprotein endonuclease with its DNA target site investigated by DNA footprinting and modification interference. J Mol Biol. 2001;309:361–86.

44. Robart AR, Zimmerly S. Group II intron retroelements: function and diversity. Cytogenet Genome Res. 2005;110:589–97.

45. Matsuura M, Saldanha R, Ma H, Wank H, Yang J, Mohr G, et al. A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. Genes Dev. 1997;11:2910–24.

46. Eskes R, Yang J, Lambowitz AM, Perlman PS. Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. Cell. 1997;88:865–74.

47. Smith D, Zhong J, Matsuura M, Lambowitz AM, Belfort M. Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming. Genes Dev. 2005;19:2477–87.

48. Contreras LM, Huang T, Piazza CL, Smith D, Qu G, Gelderman G, et al. Group II intron-ribosome association protects intron RNA from degradation. RNA. 2013;19:1497–509.

49. Dai L, Zimmerly S. Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. Nucleic Acids Res. 2002;30:1091–102.

50. Simon DM, Clarke NA, McNeil BA, Johnson I, Pantuso D, Dai L, et al. Group II introns in eubacteria and archaea: ORF-less introns and new varieties. RNA. 2008;14:1704–13.

51. Candales MA, Duong A, Hood KS, Li T, Neufeld RA, Sun R, et al. Database for bacterial group II introns. Nucleic Acids Res. 2012;40:D187–90.

52. Granlund M, Michel F, Norgren M. Mutually exclusive distribution of IS1548 and GBSi1, an active group II intron identified in human isolates of group B streptococci. J Bacteriol. 2001;183:2560–9.

53. Toor N, Hausner G, Zimmerly S. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. RNA. 2001;7:1142–52.

54. Martinez-Abarca F, Toro N. Group II introns in the bacterial world. Mol Microbiol. 2000;38:917–26.

55. Toro N. Bacteria and archaea group II introns: additional mobile genetic elements in the environment. Environ Microbiol. 2003;5:143–51.

56. Toro N, Molina-Sanchez MD, Fernandez-Lopez M. Identification and characterization of bacterial class E group II introns. Gene. 2002;299:245–50.

57. Simon DM, Kelchner SA, Zimmerly S. A broadscale phylogenetic analysis of group II intron RNAs and intron-encoded reverse transcriptases. Mol Biol Evol. 2009;26:2795–808.

58. Shub DA, Goodrich-Blair H, Eddy SR. Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. Trends Biochem Sci. 1994;19:402–4.

59. Ichiyanagi K, Beauregard A, Lawrence S, Smith D, Cousineau B, Belfort M. Retrotransposition of the Ll.LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. Mol Microbiol. 2002;46:1259–72.

60. Zhong J, Lambowitz AM. Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription. EMBO J. 2003;22:4555–65.

61. Ichiyanagi K, Beauregard A, Belfort M. A bacterial group II intron favors retrotransposition into plasmid targets. Proc Natl Acad Sci U S A. 2003;100:15742–7.

62. Munoz-Adelantado E, San Filippo J, Martinez-Abarca F, Garcia-Rodriguez FM, Lambowitz AM, Toro N. Mobility of the Sinorhizobium meliloti group II intron RmInt1 occurs by reverse splicing into DNA, but requires an unknown reverse transcriptase priming mechanism. J Mol Biol. 2003;327:931–43.

63. Sunde M. Class I, integron with a group II intron detected in an Escherichia coli strain from a free-range reindeer. Antimicrob Agents Chemother. 2005;49:2512–4.

64. Robart AR, Seo W, Zimmerly S. Insertion of group II intron retroelements after intrinsic transcriptional terminators. Proc Natl Acad Sci U S A. 2007;104:6620–5.

65. Quiroga C, Roy PH, Centron D. The S.ma.I2 class C group II intron inserts at integron attC sites. Microbiology. 2008;154:1341–53.

66. Quiroga C, Centron D. Using genomic data to determine the diversity and distribution of target site motifs recognized by class C-attC group II introns. J Mol Evol. 2009;68:539–49.

67. Leon G, Roy PH. Potential role of group IIC-attC introns in integron cassette formation. J Bacteriol. 2009;191:6040–51.

68. Leon G, Roy PH. Group IIC intron mobility into attC sites involves a bulged DNA stem-loop motif. RNA. 2009;15:1543–53.

69. Leon G, Quiroga C, Centron D, Roy PH. Diversity and strength of internal outward-oriented promoters in group IIC-attC introns. Nucleic Acids Res. 2010;38:8196–207.

70. Zoschke R, Nakamura M, Liere K, Sugiura M, Borner T, Schmitz-Linneweber C. An organellar maturase associates with multiple group II introns. Proc Natl Acad Sci U S A. 2010;107:3245–50.

71. Brown GG, Colas Des Francs-Small C, Ostersetzer-Biran O. Group II intron splicing factors in plant mitochondria. Front Plant Sci. 2014;5:35.

72. Vogel J, Hubschmann T, Borner T, Hess WR. Splicing and intron-internal RNA editing of trnK-matK transcripts in barley plastids: support for MatK as an essential splice factor. J Mol Biol. 1997;270:179–87.

73. Toor N, Zimmerly S. Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. RNA. 2002;8:1373–7.

74. Li CF, Costa M, Bassi G, Lai YK, Michel F. Recurrent insertion of 5′-terminal nucleotides and loss of the branchpoint motif in lineages of group II introns inserted in mitochondrial preribosomal RNAs. RNA. 2011;17:1321–35.

75. Monteiro-Vitorello CB, Hausner G, Searles DB, Gibb EA, Fulbright DW, Bertrand H. The Cryphonectria parasitica mitochondrial rns gene: plasmid-like elements, introns and homing endonucleases. Fungal Genet Biol. 2009;46:837–48.

76. Salman V, Amann R, Shub DA, Schulz-Vogt HN. Multiple self-splicing introns in the 16S rRNA genes of giant sulfur bacteria. Proc Natl Acad Sci U S A. 2012;109:4203–8.

77. Mullineux ST, Costa M, Bassi GS, Michel F, Hausner G. A group II intron encodes a functional LAGLIDADG homing endonuclease and self-splices under moderate temperature and ionic conditions. RNA. 2010;16:1818–31.

78. Pfeifer A, Martin B, Kamper J, Basse CW. The mitochondrial LSU rRNA group II intron of Ustilago maydis encodes an active homing endonuclease likely involved in intron mobility. PLoS One. 2012;7:e49551.

79. Saldanha R, Mohr G, Belfort M, Lambowitz AM. Group I and group II introns. FASEB J. 1993;7:15–24.

80. Mullineux ST, Willows K, Hausner G. Evolutionary dynamics of the mS952 intron: a novel mitochondrial group II intron encoding a LAGLIDADG homing endonuclease gene. J Mol Evol. 2011;72:433–49.

81. Unseld M, Marienfeld JR, Brandt P, Brennicke A. The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. Nat Genet. 1997;15:57–61.

82. Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S. Complete structure of the chloroplast genome of Arabidopsis thaliana. DNA Res. 1999;6:283–90.

83. Turmel M, Otis C, Lemieux C. The mitochondrial genome of Chara vulgaris: insights into the mitochondrial DNA architecture of the last common ancestor of green algae and land plants. Plant Cell. 2003;15:1888–903.

84. Yin LF, Hu MJ, Wang F, Kuang H, Zhang Y, Schnabel G, et al. Frequent gain and loss of introns in fungal cytochrome b genes. PLoS One. 2012;7:e49096.

85. Guo W, Mower JP. Evolution of plant mitochondrial intron-encoded maturases: frequent lineage-specific loss and recurrent intracellular transfer to the nucleus. J Mol Evol. 2013;77:43–54.

86. Hafez M, Majer A, Sethuraman J, Rudski SM, Michel F, Hausner G. The mtDNA rns gene landscape in the Ophiostomatales and other fungal taxa: twintrons, introns, and intron-encoded proteins. Fungal Genet Biol. 2013;53:71–83.

87. Meng Q, Wang Y, Liu XQ. An intron-encoded protein assists RNA splicing of multiple similar introns of different bacterial genes. J Biol Chem. 2005;280:35085–8.

88. McNeil BA, Simon DM, Zimmerly S. Alternative splicing of a group II intron in a surface layer protein gene in Clostridium tetani. Nucleic Acids Res. 2014;42:1959–69.

89. McNeil BA, Zimmerly S. Novel RNA structural features of an alternatively splicing group II intron from Clostridium tetani. RNA. 2014;20:855–66.

90. Abebe M, Candales MA, Duong A, Hood KS, Li T, Neufeld RA, et al. A pipeline of programs for collecting and analyzing group II intron retroelement sequences from GenBank. Mob DNA. 2013;4:28.

91. Carrillo C, Chapdelaine Y, Bonen L. Variation in sequence and RNA editing within core domains of mitochondrial group II introns among plants. Mol Gen Genet. 2001;264:595–603.

92. Li-Pook-Than J, Bonen L. Multiple physical forms of excised group II intron RNAs in wheat mitochondria. Nucleic Acids Res. 2006;34:2782–90.

93. Barkan A. Intron splicing in plant organelles. In: Daniell H, Chase C, editors. Molecular biology and biotechnology of plant organelles. Dordrecht, The Netherlands: Kluwer Academic Publishers; 2004. p. 281–308.

94. Germain A, Hotto AM, Barkan A, Stern DB. RNA processing and decay in plastids. Wiley Interdiscip Rev RNA. 2013;4:295–316.

95. Niemer I, Schmelzer C, Borner GV. Overexpression of DEAD box protein pMSS116 promotes ATP-dependent splicing of a yeast group II intron in vitro. Nucleic Acids Res. 1995;23:2966–72.

96. Hammani K, Barkan A. An mTERF domain protein functions in group II intron splicing in maize chloroplasts. Nucleic Acids Res. 2014;42:5033–42.

97. Jenkins BD, Kulhanek DJ, Barkan A. Nuclear mutations that block group II RNA splicing in maize chloroplasts reveal several intron classes with distinct requirements for splicing factors. Plant Cell. 1997;9:283–96.

98. Ostersetzer O, Cooke AM, Watkins KP, Barkan A. CRS1, a chloroplast group II intron splicing factor, promotes intron folding through specific interactions with two intron domains. Plant Cell. 2005;17:241–55.

99. Asakura Y, Barkan A. Arabidopsis orthologs of maize chloroplast splicing factors promote splicing of orthologous and species-specific group II introns. Plant Physiol. 2006;142:1656–63.

100. Asakura Y, Bayraktar OA, Barkan A. Two CRM protein subfamilies cooperate in the splicing of group IIB introns in chloroplasts. RNA. 2008;14:2319–32.

101. Zmudjak M, Colas Des Francs-Small C, Keren I, Shaya F, Belausov E, Small I, et al. mCSF1, a nucleus-encoded CRM protein required for the processing of many mitochondrial introns, is involved in the biogenesis of respiratory complexes I and IV in Arabidopsis. New Phytol. 2013;199:379–94.

102. Mohr G, Lambowitz AM. Putative proteins related to group II intron reverse transcriptase/maturases are encoded by nuclear genes in higher plants. Nucleic Acids Res. 2003;31:647–52.

103. Nakagawa N, Sakurai N. A mutation in At-nMat1a, which encodes a nuclear gene having high similarity to group II intron maturase, causes impaired splicing of mitochondrial NAD4 transcript and altered carbon metabolism in Arabidopsis thaliana. Plant Cell Physiol. 2006;47:772–83.

104. Keren I, Bezawork-Geleta A, Kolton M, Maayan I, Belausov E, Levy M, et al. AtnMat2, a nuclear-encoded maturase required for splicing of group-II introns in Arabidopsis mitochondria. RNA. 2009;15:2299–311.

105. Keren I, Tal L, Colas Des Francs-Small C, Araujo WL, Shevtsov S, Shaya F. nMAT1, a nuclear-encoded maturase involved in the trans-splicing of nad1 intron 1, is essential for mitochondrial complex I assembly and function. Plant J. 2012;71:413–26.

106. Copertino DW, Hallick RB. Group II and group III introns of twintrons: potential relationships with nuclear pre-mRNA introns. Trends Biochem Sci. 1993;18:467–71.

107. Bonen L. Trans-splicing of pre-mRNA in plants, animals, and protists. FASEB J. 1993;7:40–6.

108. Bonen L. Cis- and trans-splicing of group II introns in plant mitochondria. Mitochondrion. 2008;8:26–34.

109. Knoop V, Altwasser M, Brennicke A. A tripartite group II intron in mitochondria of an angiosperm plant. Mol Gen Genet. 1997;255:269–76.

110. Choquet Y, Goldschmidt-Clermont M, Girard-Bascou J, Kuck U, Bennoun P, Rochaix JD. Mutant phenotypes support a trans-splicing mechanism for the expression of the tripartite psaA gene in the C reinhardtii chloroplast. Cell. 1988;52:903–13.

111. Goldschmidt-Clermont M, Choquet Y, Girard-Bascou J, Michel F, Schirmer-Rahire M, Rochaix JD. A small chloroplast RNA may be required for trans-splicing in Chlamydomonas reinhardtii. Cell. 1991;65:135–43.

112. Malek O, Knoop V. Trans-splicing group II introns in plant mitochondria: the complete set of cis-arranged homologs in ferns, fern allies, and a hornwort. RNA. 1998;4:1599–609.

113. Qiu YL, Palmer JD. Many independent origins of trans splicing of a plant mitochondrial group II intron. J Mol Evol. 2004;59:80–9.

114. Stabell FB, Tourasse NJ, Ravnum S, Kolsto AB. Group II intron in Bacillus cereus has an unusual 3′ extension and splices 56 nucleotides downstream of the predicted site. Nucleic Acids Res. 2007;35:1612–23.

115. Stabell FB, Tourasse NJ, Kolsto AB. A conserved 3′ extension in unusual group II introns is important for efficient second-step splicing. Nucleic Acids Res. 2009;37:3202–14.

116. Tourasse NJ, Stabell FB, Kolsto AB. Structural and functional evolution of group II intron ribozymes: insights from unusual elements carrying a 3′ extension. N Biotechnol. 2010;27:204–11.

117. Jenkins KP, Hong L, Hallick RB. Alternative splicing of the Euglena gracilis chloroplast roaA transcript. RNA. 1995;1:624–33.

118. Copertino DW, Hallick RB. Group II twintron: an intron within an intron in a chloroplast cytochrome b-559 gene. EMBO J. 1991;10:433–42.

119. Robart AR, Montgomery NK, Smith KL, Zimmerly S. Principles of 3′ splice site selection and alternative splicing for an unusual group II intron from Bacillus anthracis. RNA. 2004;10:854–62.

120. Copertino DW, Christopher DA, Hallick RB. A mixed group II/group III twintron in the Euglena gracilis chloroplast ribosomal protein S3 gene: evidence for intron insertion during gene evolution. Nucleic Acids Res. 1991;19:6491–7.

121. Hong L, Hallick RB. A group III intron is formed from domains of two individual group II introns. Genes Dev. 1994;8:1589–99.

122. Nakamura Y, Kaneko T, Sato S, Ikeuchi M, Katoh H, Sasamoto S, et al. Complete genome structure of the thermophilic cyanobacterium Thermosynechococcus elongatus BP-1. DNA Res. 2002;9:123–30.

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

124. Sheveleva EV, Hallick RB. Recent horizontal intron transfer to a chloroplast genome. Nucleic Acids Res. 2004;32:803–10.

125. Odom OW, Shenkenberg DL, Garcia JA, Herrin DL. A horizontally acquired group II intron in the chloroplast psbA gene of a psychrophilic Chlamydomonas: in vitro self-splicing and genetic evidence for maturase activity. RNA. 2004;10:1097–107.

126. Khan H, Archibald JM. Lateral transfer of introns in the cryptophyte plastid genome. Nucleic Acids Res. 2008;36:3043–53.

127. Kamikawa R, Masuda I, Demura M, Oyama K, Yoshimatsu S, Kawachi M, et al. Mitochondrial group II introns in the raphidophycean flagellate Chattonella spp suggest a diatom-to-Chattonella lateral group II intron transfer. Protist. 2009;160:364–75.

128. Hardy CM, Clark-Walker GD. Nucleotide sequence of the COX1 gene in Kluyveromyces lactis mitochondrial DNA: evidence for recent horizontal transfer of a group II intron. Curr Genet. 1991;20:99–114.

129. Leclercq S, Giraud I, Cordaux R. Remarkable abundance and evolution of mobile group II introns in Wolbachia bacterial endosymbionts. Mol Biol Evol. 2011;28:685–97.

130. Dai L, Zimmerly S. The dispersal of five group II introns among natural populations of Escherichia coli. RNA. 2002;8:1294–307.

131. Tourasse NJ, Kolsto AB. Survey of group I and group II introns in 29 sequenced genomes of the Bacillus cereus group: insights into their spread and evolution. Nucleic Acids Res. 2008;36:4529–48.

132. Toro N, Martinez-Abarca F. Comprehensive phylogenetic analysis of bacterial group II intron-encoded ORFs lacking the DNA endonuclease domain reveals new varieties. PLoS One. 2013;8:e55102.

133. Fontaine JM, Goux D, Kloareg B, Loiseaux-de GS. The reverse-transcriptase-like proteins encoded by group II introns in the mitochondrial genome of the brown alga Pylaiella littoralis belong to two different lineages which apparently coevolved with the group II ribosyme lineages. J Mol Evol. 1997;44:33–42.

134. Goddard MR, Burt A. Recurrent invasion and extinction of a selfish gene. Proc Natl Acad Sci U S A. 1999;96:13880–5.

135. Burt A, Koufopanou V. Homing endonuclease genes: the rise and fall and rise again of a selfish element. Curr Opin Genet Dev. 2004;14:609–15.

136. Haugen P, Simon DM, Bhattacharya D. The natural history of group I introns. Trends Genet. 2005;21:111–9.

137. Haugen P, Wikmark OG, Vader A, Coucheron DH, Sjottem E, Johansen SD. The recent transfer of a homing endonuclease gene. Nucleic Acids Res. 2005;33:2734–41.

138. Adamidi C, Fedorova O, Pyle AM. A group II intron inserted into a bacterial heat-shock operon shows autocatalytic activity and unusual thermostability. Biochemistry. 2003;42:3409–18.

139. Chee GJ, Takami H. Housekeeping recA gene interrupted by group II intron in the thermophilic Geobacillus kaustophilus. Gene. 2005;363:211–20.

140. Pombert JF, James ER, Janouskovec J, Keeling PJ. Evidence for transitional stages in the evolution of euglenid group II introns and twintrons in the Monomorphina aenigmatica plastid genome. PLoS One. 2012;7:e53433.

141. Wiegert KE, Bennett MS, Triemer RE. Tracing patterns of chloroplast evolution in euglenoids: contributions from Colacium vesiculosum and Strombomonas acuminata (Euglenophyta). J Eukaryot Microbiol. 2013;60:214–21.

142. Curcio MJ, Belfort M. Retrohoming: cDNA-mediated mobility of group II introns requires a catalytic RNA. Cell. 1996;84:9–12.

143. Simon DM, Zimmerly S. A diversity of uncharacterized reverse transcriptases in bacteria. Nucleic Acids Res. 2008;36:7219–29.

144. Weinberg Z, Perreault J, Meyer MM, Breaker RR. Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. Nature. 2009;462:656–9.

145. Weinberg Z, Wang JX, Bogue J, Yang J, Corbino K, Moy RH, et al. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. Genome Biol. 2010;11:R31.

146. Storz G, Vogel J, Wassarman KM. Regulation by small RNAs in bacteria: expanding frontiers. Mol Cell. 2011;43:880–91.

147. Toor N, Keating KS, Taylor SD, Pyle AM. Crystal structure of a self-spliced group II intron. Science. 2008;320:77–82.

148. Rest JS, Mindell DP. Retroids in archaea: phylogeny and lateral origins. Mol Biol Evol. 2003;20:1134–42.

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

150. Cech TR. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. Cell. 1986;44:207–10.

151. Sharp PA. Five easy pieces. Science. 1991;254:663.

152. Weiner AM. mRNA splicing and autocatalytic introns: distant cousins or the products of chemical determinism? Cell. 1993;72:161–4.

153. Moore MJ, Sharp PA. Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature. 1993;365:364–8.

154. Maschhoff KL, Padgett RA. The stereochemical course of the first step of pre-mRNA splicing. Nucleic Acids Res. 1993;21:5456–62.

155. Padgett RA, Podar M, Boulanger SC, Perlman PS. The stereochemical course of group II intron self-splicing. Science. 1994;266:1685–8.

156. Podar M, Perlman PS, Padgett RA. The two steps of group II intron self-splicing are mechanistically distinguishable. RNA. 1998;4:890–900.

157. Rajagopal J, Doudna JA, Szostak JW. Stereochemical course of catalysis by the Tetrahymena ribozyme. Science. 1989;244:692–4.

158. McSwiggen JA, Cech TR. Stereochemistry of RNA cleavage by the Tetrahymena ribozyme and evidence that the chemical step is not rate-limiting. Science. 1989;244:679–83.

159. Valadkhan S. The role of snRNAs in spliceosomal catalysis. Prog Mol Biol Transl Sci. 2013;120:195–228.

160. Chanfreau G, Jacquier A. Interaction of intronic boundaries is required for the second splicing step efficiency of a group II intron. EMBO J. 1993;12:5173–80.

161. Parker R, Siliciano PG. Evidence for an essential non-Watson-Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. Nature. 1993;361:660–2.

162. Ruis BL, Kivens WJ, Siliciano PG. The interaction between the first and last intron nucleotides in the second step of pre-mRNA splicing is independent of other conserved intron nucleotides. Nucleic Acids Res. 1994;22:5190–5.

163. Shukla GC, Padgett RA. A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome. Mol Cell. 2002;9:1145–50.

164. Fica SM, Tuttle N, Novak T, Li NS, Lu J, Koodathingal P, et al. RNA catalyses nuclear pre-mRNA splicing. Nature. 2013;503:229–34.

165. Fica SM, Mefford MA, Piccirilli JA, Staley JP. Evidence for a group II intron-like catalytic triplex in the spliceosome. Nat Struct Mol Biol. 2014;21:464–71.

166. Galej WP, Oubridge C, Newman AJ, Nagai K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. Nature. 2013;493:638–43.

167. Dlakic M, Mushegian A. Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase. RNA. 2011;17:799–808.

168. Newman AJ, Norman C. U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. Cell. 1992;68:743–54.

169. Hetzer M, Wurzer G, Schweyen RJ, Mueller MW. Trans-activation of group II intron splicing by nuclear U5 snRNA. Nature. 1997;386:417–20.

170. O'Keefe RT, Norman C, Newman AJ. The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast. Cell. 1996;86:679–89.

171. Lesser CF, Guthrie C. Mutations in U6 snRNA that alter splice site specificity: implications for the active site. Science. 1993;262:1982–8.

172. Kandels-Lewis S, Seraphin B. Involvement of U6 snRNA in 5′ splice site selection. Science. 1993;262:2035–9.

173. Boudvillain M, de Lencastre A, Pyle AM. A tertiary interaction that links active-site domains to the 5′ splice site of a group II intron. Nature. 2000;406:315–8.

174. Rhode BM, Hartmuth K, Westhof E, Luhrmann R. Proximity of conserved U6 and U2 snRNA elements to the 5′ splice site region in activated spliceosomes. EMBO J. 2006;25:2475–86.

175. Anokhina M, Bessonov S, Miao Z, Westhof E, Hartmuth K, Luhrmann R. RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. EMBO J. 2013;32:2804–18.

176. Staley JP, Guthrie C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell. 1998;92:315–26.

177. Madhani HD, Guthrie C. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. Cell. 1992;71:803–17.

178. Reddy R, Henning D, Epstein P, Busch H. Primary and secondary structure of U2 snRNA. Nucleic Acids Res. 1981;9:5645–58.

179. Guthrie C, Patterson B. Spliceosomal snRNAs. Annu Rev Genet. 1988;22:387–419.

180. Sashital DG, Cornilescu G, McManus CJ, Brow DA, Butcher SE. U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction. Nat Struct Mol Biol. 2004;11:1237–42.

181. Dunn EA, Rader SD. Secondary structure of U6 small nuclear RNA: implications for spliceosome assembly. Biochem Soc Trans. 2010;38:1099–104.

182. Will CL, Luhrmann R. Spliceosome structure and function. Cold Spring Harb Perspect Biol. 2011;3(7):a003707. doi:10.1101/cshperspect.a003707.

183. Mefford MA, Staley JP. Evidence that U2/U6 helix I promotes both catalytic steps of pre-mRNA splicing and rearranges in between these steps. RNA. 2009;15:1386–97.

184. Irimia M, Roy SW. Origin of spliceosomal introns and alternative splicing. Cold Spring Harb Perspect Biol. 2014;6(6):a016071. doi:10.1101/cshperspect.a016071.

185. Andersson JO, Sjogren AM, Horner DS, Murphy CA, Dyal PL, Svard SG, et al. A genomic survey of the fish parasite Spironucleus salmonicida indicates genomic plasticity among diplomonads and significant lateral gene transfer in eukaryote genome evolution. BMC Genomics. 2007;8:51.

186. Lane CE, van den Heuvel K, Kozera C, Curtis BA, Parsons BJ, Bowman S, et al. Nucleomorph genome of Hemiselmis andersenii reveals complete intron loss and compaction as a driver of protein structure and function. Proc Natl Acad Sci U S A. 2007;104:19908–13.

187. Koonin EV. Intron-dominated genomes of early ancestors of eukaryotes. J Hered. 2009;100:618–23.

188. Martin W, Koonin EV. Introns and the origin of nucleus-cytosol compartmentalization. Nature. 2006;440:41–5.

189. Belhocine K, Mak AB, Cousineau B. Trans-splicing of the Ll.LtrB group II intron in Lactococcus lactis. Nucleic Acids Res. 2007;35:2257–68.

190. Chalamcharla VR, Curcio MJ, Belfort M. Nuclear expression of a group II intron is consistent with spliceosomal intron ancestry. Genes Dev. 2010;24:827–36.

191. Xiong Y, Eickbush TH. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 1990;9:3353–62.

192. Grainger RJ, Beggs JD. Prp8 protein: at the heart of the spliceosome. RNA. 2005;11:533–57.

193. Koonin EV. The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? Biol Direct. 2006;1:22.

194. Belhocine K, Mak AB, Cousineau B. Trans-splicing versatility of the Ll.LtrB group II intron. RNA. 2008;14:1782–90.

195. Ritlop C, Monat C, Cousineau B. Isolation and characterization of functional tripartite group II introns using a Tn5-based genetic screen. PLoS One. 2012;7:e41589.

196. Qu G, Dong X, Piazza CL, Chalamcharla VR, Lutz S, Curcio MJ, et al. RNA-RNA interactions and pre-mRNA mislocalization as drivers of group II intron loss from nuclear genomes. Proc Natl Acad Sci U S A. 2014;111:6612–7.

197. Doolittle WF. The trouble with (group II) introns. Proc Natl Acad Sci U S A. 2014;111:6536–7.

198. Eickbush TH. Origin and evolutionary relationships of retroelements. In: Morse SS, editor. Evolutionary biology of viruses. New York: Raven; 1994. p. 121–57.

199. Malik HS, Burke WD, Eickbush TH. The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol. 1999;16:793–805.

200. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell. 1993;72:595–605.

201. Kojima KK, Kanehisa M. Systematic survey for novel types of prokaryotic retroelements based on gene neighborhood and protein architecture. Mol Biol Evol. 2008;25:1395–404.

202. van der Oost J, Westra ER, Jackson RN, Wiedenheft B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat Rev Microbiol. 2014;12:479–92.

203. Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, et al. Rfam 11.0: 10 years of RNA families. Nucleic Acids Res. 2013;41:D226–32.