Published in final edited form as:

Microbiol Spectr. 2015 February ; 3(1): MDNA3-0050-2014. doi:10.1128/
microbiolspec.MDNA3-0050-2014.

Mobile Bacterial Group II Introns at the Crux of Eukaryotic Evolution

Alan M. Lambowitz^{1,*} and Marlene Belfort^{2,*}

¹Institute for Cellular Molecular Biology and, Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712

²Department of Biological Sciences and, RNA Institute, University at Albany, State University of New York, Albany, NY 12222

SUMMARY

This review focuses on recent developments in our understanding of group II intron function, the relationships of these introns to retrotransposons and spliceosomes, and how their common features have informed thinking about bacterial group II introns as key elements in eukaryotic evolution. Reverse transcriptase-mediated and host factor-aided intron retrohoming pathways are considered along with retrotransposition mechanisms to novel sites in bacteria, where group II introns are thought to have originated. DNA target recognition and movement by target-primed reverse transcription infer an evolutionary relationship among group II introns, non-LTR retrotransposons, such as LINE elements, and telomerase. Additionally, group II introns are almost certainly the progenitors of spliceosomal introns. Their profound similarities include splicing chemistry extending to RNA catalysis, reaction stereochemistry, and the position of two divalent metals that perform catalysis at the RNA active site. There are also sequence and structural similarities between group II introns and the spliceosome's small nuclear RNAs (snRNAs) and between a highly conserved core spliceosomal protein Prp8 and a group II intron-like reverse transcriptase. It has been proposed that group II introns entered eukaryotes during bacterial endosymbiosis or bacterial-archaeal fusion, proliferated within the nuclear genome, necessitating evolution of the nuclear envelope, and fragmented giving rise to spliceosomal introns. Thus, these bacterial self-splicing mobile elements have fundamentally impacted the composition of extant eukaryotic genomes, including the human genome, most of which is derived from close relatives of mobile group II introns.

Introduction

Group II introns are remarkable mobile retroelements that use the combined activities of an autocatalytic RNA and an intron-encoded reverse transcriptase (RT) to propagate efficiently within genomes. But perhaps their most noteworthy feature is the pivotal role they are thought to have played in eukaryotic evolution. Mobile group II introns are ancestrally related to nuclear spliceosomal introns, retrotransposons and telomerase, which collectively comprise more than half of the human genome. Additionally, group II introns are postulated

*Correspondence: lambowitz@austin.utexas.edu or mbelfort@albany.edu.

to have been a major driving force in the evolution of eukaryotes themselves, including for the emergence of the nuclear envelope to separate transcription from translation.

In this review, we focus on recent developments in our understanding of group II intron function, the relationships of these introns to retrotransposons and spliceosomes, and how their common features inform our thinking about bacterial group II introns at the crux of eukaryotic evolution. We rely on previous reviews for more detailed coverage of history, structure, mechanism and biotechnological applications of group II introns (1–6).

Background

Group II introns are found predominantly in bacteria and in the mitochondrial (mt) and chloroplast (cp) genomes of some eukaryotes, particularly fungi and plants, but are rare in archaea and absent from eukaryotic nuclear genomes (4). Mobile group II introns consist of a catalytically active intron RNA (a ribozyme) and an intron-encoded protein (IEP), which is a multifunctional RT. The IEP functions in intron mobility by synthesizing a cDNA copy of the intron RNA and as a “maturase” that promotes folding of the intron RNA into a catalytically active ribozyme structure required for both RNA splicing and mobility reactions. Some IEPs also have a DNA endonuclease (En) activity that plays a role in intron mobility.

Group II intron splicing

The splicing pathway, which is assisted by the IEP, involves two reversible transesterifications catalyzed by the intron RNA (7). In the first transesterification, the 2'-OH of a “branch-point” adenosine near the 3' end of the intron attacks the 5'-splice site (Fig. 1A). This reaction releases the 5' exon and produces a branched intermediate in which the attacking adenosine is linked to the 5' intron residue by a 2'-5' phosphodiester bond. In the second transesterification, the newly released 3'-OH of the 5'-exon attacks the 3' splice site, resulting in ligation of the 5' and 3' exons and excision of the intron lariat. A linear intron can result from hydrolysis rather than transesterification at the 5'-splice site, or by a lariat reopening reaction (8, 9). Circular introns can also form (10). The reversibility of the transesterifications (Fig. 1A) enables “reverse splicing” of the excised intron into RNA or DNA containing the ligated-exon sequence, and may also provide a proof-reading mechanism for 5'-splice site selection (11). Reverse splicing into DNA plays a key role in intron mobility.

Intron architecture

Group II intron RNAs have conserved 5'- and 3'-end sequences (GUGYC and AY, respectively), which resemble those of spliceosomal introns (GU and AG, respectively), and fold into a conserved three-dimensional structure consisting of six interacting secondary structure domains (DI to DVI) (Fig. 1B and C). This folded RNA forms the ribozyme active site, which uses specifically bound Mg^{2+} ions to catalyze RNA splicing and reverse splicing reactions.

Biochemical studies and X-ray crystal structures of a group II intron from the halophile *Oceanobacillus iheyensis* have provided insight into the function of group II introns domains

(Fig. 1C) (5, 10, 12, 13). DI, the largest domain, is a scaffold, which contains sequence motifs that base pair with exon sequences to align them at the active site. These exon-recognition motifs differ for group II intron subgroups (IIA, IIB and IIC; see below) and are denoted exon-binding sites (EBSs) and δ with the complementary exon motifs denoted intron-binding sites (IBSs) and δ' (Fig. 1D). DV is the active site helix that binds metal ions at the catalytic center of the intron. DVI contains the branch-point A residue and undergoes a conformational change to reposition the branch-point A between the two steps of splicing (14–16). DII and DIII engage in stabilizing interactions with other domains, with DIII functioning as an effector to increase the rate of catalysis. The ORF encoding the RT protrudes from DIV, which projects away from the catalytic core.

Intron-encoded reverse transcriptase

Figure 2 compares two well-studied group II intron RTs (the LtrA protein encoded by the *Lactococcus lactis* LI.LtrB intron and the *Sinorhizobium meliloti* RmInt1 RT) with two non-LTR-retrotransposon RTs (R2Bm and LINE-1), telomerase, and retroviral (HIV-1) RT. The LI.LtrB RT contains four conserved domains, RT, X/thumb, DNA binding (D), and DNA endonuclease (En), whereas the RmInt1 RT belongs to a subset of group II intron RTs that lacks the En domain. The RT domain contains conserved amino acid sequence blocks 1–7 that are present in the fingers and palm regions of retroviral RTs, while the X/thumb domain has a predicted structure similar to the thumb domain of retroviral RTs (17). Although homologous to retroviral RTs, group II intron RTs have an N-terminal extension containing an additional conserved sequence block (RT-0), as well as insertions between other conserved sequence blocks (RT-2a, -3a, -4a, and -7a), some of which are conserved in non-LTR-retrotransposons and telomerase RTs (17, 18). A three-dimensional model of a group II intron RT predicts a right hand-like structure similar to that of retroviral RTs with the N-terminal extension comprising part of a larger fingers region and the other insertions lying outside the RT active site (Fig. 2B) (17).

In addition to reverse transcription, the RT and X/thumb domains of group II intron RTs bind the intron RNA for splicing. Mutational and high-throughput unigenic evolution analyses suggest an extended RNA-binding surface that includes distal parts of the fingers, regions in and around the template-primer binding tract, and patches on the back of the hand (Fig. 2B, red left and dark blue right highlight regions that potentially interact with different parts of the intron RNA; see legend and below) (17, 19). The RT active site is not required for splicing, with mutations in the conserved YADD metal binding motif having little effect on splicing activity (19, 20). The overlap between regions of the IEP required for splicing with those that bind RNA templates for reverse transcription suggests how an RT that became associated with a catalytic RNA could evolve a secondary function in RNA splicing (19, 21).

The D domain of group II IEPs functions in DNA target site binding during intron mobility. Studies with the LI.LtrB IEP identified two regions of this domain that are functionally important and conserved in other group II IEPs, an upstream cluster of basic amino acids and a downstream predicted α -helix (22). Mutations in these regions affect both the efficiency and target specificity of retrohoming, consistent with their involvement in DNA

target site recognition (22, 23). Distinctive variations of the D domain have been described for RmInt1 and related group II intron lineages whose IEPs lack an En domain (22, 24, 25).

The En domain belongs to the H-N-H family of DNA endonucleases, which includes bacterial colicins, phage T4 endonuclease VII, and some group I intron homing endonucleases (22, 26–29). In group II intron RTs, residues of the H-N-H motif coordinate a catalytic Mg^{2+} ion and are interspersed with two pairs of conserved cysteine residues, which may coordinate another metal ion to stabilize the protein fold (22). These cysteine residues are important for the endonuclease activity of the L1.LtrB RT, but have diverged in some group II intron IEP lineages (22).

Group II intron RNPs

Group II intron RNAs and IEPs function together in a ribonucleoprotein (RNP) complex, which forms when the IEP binds to the intron in unspliced precursor RNA to promote RNA-catalyzed splicing (30, 31). Group II intron IEPs typically function as intron-specific splicing factors, discriminating even against closely related group II introns (32–34). L1.LtrB RNPs contain two molecules of IEP per intron RNA, suggesting that the IEP functions as a dimer, similar to HIV-1 RT (30, 35, 36).

The IEP has a high-affinity binding site in intron subdomain DIVa, a variable stem loop structure that lies outside the intron's catalytic core near the beginning of DIV and is a feature that contributes to intron specificity (37–39). The IEP also makes weaker secondary contacts with conserved core regions, including in DI, DII, and DVI, that stabilize the active ribozyme structure for RNA splicing and reverse splicing (31, 40). This mode of interaction in which the IEP is anchored by binding to DIVa and binds more weakly to the core may enable the IEP to accommodate conformational changes within the intron RNA during RNA splicing and different steps in intron mobility (31, 36, 41–43).

When mapped onto a three-dimensional model of the L1.LtrB intron, putative IEP contact sites identified by RNA footprinting, site-specific cross-linking and fluorescence quenching revealed a broad IEP binding surface that extends from DIVa across the interface of DI, DII, and DVI (31, 40). The location of the putative IEP contact sites and analysis of conformational changes that occur upon IEP binding suggest that the IEP stabilizes interactions between DI, II and VI, as well as the folded structure of DI (31, 41). Although much group II intron RNA tertiary structure can form at physiological Mg^{2+} concentrations in the absence of the IEP (31, 41), continued binding of the L1.LtrB IEP is required to stabilize the active ribozyme structure even for lariat RNA, which rapidly reverts to an inactive structure if the IEP is removed (44, 45).

Group II intron classification

Group II intron IEPs have a strong *cis*-preference for splicing and mobilizing the intron RNA in which they are encoded, likely reflecting that the nascent IEP binds to the intron during or just after translation and then remains bound to the excised intron during RNA splicing and intron mobility (19, 46). As a result, group II intron RNAs and IEPs have co-evolved, leading to divergence of mobile group II introns into distinct evolutionary lineages (47, 48). Group II intron RNAs are classified into three major subgroups (IIA, IIB, and IIC)

distinguished by their size, secondary structure and mode of exon recognition by DI (Fig. 1D) (discussed in detail in (4)). Whereas EBS1-IBS1 interactions are common to all group II introns, EBS2-IBS2 base pairings between the intron and 5' exon are confined to IIA and IIB introns. The 3' exon is recognized by the δ - δ' interaction in IIA introns and by the EBS3-IBS3 interaction involving another region of the intron RNA in IIB and IIC introns (Fig. 1D). The IEPs have diverged into at least nine subclasses (A, B, C, D, E, F, chloroplast-like 1 and 2 (CL1 and CL2, respectively) and mitochondria-like (ML)), each associated with a particular RNA structure (25, 49–52). The En domain, which enables an efficient mode of intron mobility (see below), is present only in IEPs of subclasses B, ML, and CL, suggesting that it may have been acquired by a common ancestor of these subclasses (25).

Group II intron retrohoming mechanisms

Group II intron retromobility occurs by a target DNA-primed reverse transcription (TPRT) mechanism in which the excised intron RNA reverse splices into one strand of a DNA target site and is then reverse transcribed by the IEP to produce an intron cDNA that is integrated into the genome (53–57). This mechanism is used by group II introns both for retrohoming to specific DNA target sites and for retrotransposition (also referred to as ectopic retrohoming) to sites that resemble the normal homing site (Fig. 3A–C). In contrast to retrohoming, which occurs at frequencies that can approach 100% of recipients acquiring the intron, retrotransposition occurs at lower frequency, typically $<10^{-4}$ /recipient (58, 59). The ability of the intron to insert into different genes and then remove itself by RNA splicing minimizes damage to the host. Variations of the retrohoming mechanism discussed below illustrate how group II introns can evolve different modes of action and adapt to different hosts.

IEP expression

Building upon earlier genetic studies (reviewed in (1)), the major features of group II intron retromobility mechanisms were elucidated for *Saccharomyces cerevisiae* mtDNA introns aI1 and aI2, in the *COX1* gene encoding cytochrome oxidase subunit I (53–56), and in bacteria, for the L1.LtrB intron, in a relaxase gene in a conjugative element (57) and the *Sinorhizobium meliloti* RmInt1 intron in an insertion sequence (IS) element (60–62). The aI1, aI2 and L1.LtrB introns are group IIA, whose IEPs contain an En domain, whereas RmInt1 is group IIB and does not. For all these introns, the IEP is translated from the intron-containing precursor RNA, binds to the unspliced intron, and promotes formation of the active ribozyme structure for RNA splicing. Cryo-EM and small-angle X-ray scattering show that precursor RNP forms a loosely packed structure that undergoes a dramatic conformational change upon splicing, resulting in a compact excised intron RNP particle (36, 43).

The nascent IEP is thought to bind first to its primary high-affinity binding site in DIVa, using regions near its N-terminus, including the N-terminal extension upstream of RT-0 (46) (Fig. 2B left, regions highlighted in red). In the L1.LtrB intron as in many other bacterial group II introns, DIVa contains the Shine-Dalgarno sequence and initiation codon of the intron ORF and the binding of the IEP to DIVa down-regulates its own translation (38, 63).

This binding to DIVa halts ribosome entry into the intron, enabling it to fold into the active ribozyme structure for RNA splicing, and also prevents accumulation of excess free IEP, which would likely be deleterious to the host cell. In the yeast *aI1* and *aI2* introns and some other fungal mtDNA introns, the intron ORF is continuous with that of the upstream exon and translated as part of a pre-protein that is proteolytically processed to yield the active IEP (38, 64). DIVa nevertheless remains a high-affinity binding site for the IEP, reflecting a critical role of this interaction in nucleating RNP assembly and positioning the IEP for initiation of TPRT (65). After splicing, the IEP remains tightly bound to the intron lariat in a stable RNP that promotes retrohoming (30, 42).

Endonuclease-dependent retrohoming

Group II intron RNPs initiate retrohoming by recognizing a DNA target sequence using both the IEP and motifs within the intron RNA that base pair with the DNA target. The intron RNA then reverse splices into the DNA strand to which it is paired (Fig. 3A and B, step 1). In the En-dependent retrohoming pathway, the IEP uses its En domain to nick the opposite strand (Fig. 3A, step 1') and then uses the 3' DNA end generated at the nick as a primer for TPRT of the inserted intron RNA (Fig. 3A, step 2). The resulting intron cDNA is integrated into the genome by host enzymes, and after intron degradation and second-strand cDNA synthesis (Fig. 3A, step 3), the nicks are ligated to complete the reaction (Fig. 3A, step 4). Unlike retroviruses, whose RTs have low fidelity and processivity that introduce and propagate mutations for evasion of host defenses (66), group II intron retrohoming requires an RT with high processivity and fidelity (error rate of $\sim 10^{-5}/\text{nt}$ for the L1.LtrB RT) in order to preserve the functional integrity of the intron (57, 67, 68).

Group II introns can use different mechanisms for cDNA integration, depending upon both the intron and DNA repair pathways that are available in different hosts. For the *S. cerevisiae* mtDNA introns, most retrohoming events are completed via a recombination mechanism in which a nascent intron cDNA strand invades an intron-containing allele for the completion of intron DNA synthesis and a second cross-over occurs in the homologous upstream exon (56, 69). This process results in co-conversion of the intron with sequence polymorphisms markers in the upstream exon (20, 70). A smaller proportion of retrohoming events occurs without co-conversion of exon sequences, possibly via synthesis of a full-length intron cDNA that is integrated by DNA repair (69).

By contrast to the *S. cerevisiae* introns, which rely heavily on homologous recombination, the L1.LtrB intron and other bacterial group II introns use RecA-independent DNA repair (57, 60, 71). The preference of bacterial introns for this mechanism may reflect less proficient homologous recombination than in fungal mitochondria.

Genetic screens and biochemical assays in which purified group II intron RNPs were combined with *E. coli* extracts to reconstitute the complete retrohoming reaction *in vitro* suggested a model for host contributions to L1.LtrB retrohoming shown in Fig. 3A (72, 73). According to this model, enzymes including nucleases directed against RNA and DNA, replicative and repair DNA polymerases and components of the replisome act in their various roles to resect DNA, degrade the template RNA, traverse RNA-DNA junctions, carry out second-strand DNA synthesis and ligate DNA (Fig. 3A and Table 1). A key feature is

that after TPRT, the intron cDNA is extended into the upstream exon, either by the group II intron RT or by a switch to a host DNA repair polymerase, resulting in a branched intermediate that resembles a stalled DNA replication fork (Fig. 3A, step 2). The latter is recognized by the replication-restart proteins PriA or PriC, which act on branched structures having different length gaps between the branch and the nascent DNA strand. PriA and PriC are thought to function in retrohoming much as they would during replication restart at a stalled or collapsed replication fork by initiating a replisome loading cascade leading to recruitment of DNA polymerase PolII and the host replicative polymerase, PolIII (Fig. 3A, step 3).

Because group II introns encode RTs lacking an RNase H domain (17), they must rely on a host RNase H to degrade the intron RNA template. For the Ll.LtrB intron in *E. coli*, this is achieved by RNase H1, with a likely contribution from the 5' to 3' exonuclease activity of PolII, which removes RNA primers during DNA replication (Fig. 3A, step 3) (72, 73). DNA overhangs are trimmed by host 5'-3' exonucleases, such as RecJ, and after completion of second-strand synthesis by PolIII, nicks are sealed by host DNA ligase (Fig. 3A, step 4). The requirement for host DNA polymerases for second-strand synthesis is again dictated by the properties of group II intron RTs, which do not initiate efficiently from an annealed primer on DNA templates (72).

Whereas hosts provide housekeeping functions, such as the single-strand binding protein Ssb and nucleoid components H-NS and StpA, that promote group II intron retromobility (73, 74), they also mount counterattacks, to inhibit intron proliferation (75). Both RNases (RNase E, RNase I and RNase LS) and DNases (ExoIII) are degradative enzymes that keep retromobility in check (45, 72, 73, 76). The RNases act by degrading the intron RNA, whereas ExoIII is thought to resect newly-synthesized cDNA.

Endonuclease-independent retrohoming

Group II introns whose IEPs lack DNA endonuclease activity retrohome by a mechanism in which a nascent leading or lagging strand at a DNA replication fork rather than a cleaved DNA strand is used to prime reverse transcription. The RmInt1 group IIB intron and group IIC introns, which encode IEPs lacking an En domain, preferentially use lagging DNA strand primers (Okazaki fragments), likely reflecting that these introns reverse splice into transiently ssDNA at DNA replication forks (62) (Fig. 3B). The relatively high efficiency with which RmInt1 retrohomes by this mechanism (20–45% per recipient target site) raises the possibility of a yet-to-be-demonstrated interaction between the intron RNP and DNA replication machinery (77). In the case of Ll.LtrB, which reverse splices efficiently into dsDNA, En-deficient mutants show the opposite bias for using the nascent leading strand as a primer for reverse transcription of the intron RNA (78). This bias likely reflects that after reverse splicing into dsDNA, an intron RNP integrated into the leading template strand is positioned to directly use a nascent leading strand from an approaching replication fork to prime reverse transcription; by contrast, an intron RNP integrated into the lagging template strand could not be reverse transcribed until after the potentially disruptive passage of the replication fork (78).

Retrohoming of linear group II intron RNAs

Experiments in which Ll.LtrB intron RNPs were microinjected into *Xenopus laevis* oocyte nuclei or *Drosophila melanogaster* embryos showed that linear group II intron RNAs can retrohome by a mechanism in which group II intron RNPs catalyze the first step of reverse splicing, resulting in attachment of the linear intron RNA to the downstream exon of the DNA target site. TPRT then yields an intron cDNA whose 3' end is ligated to the upstream exon DNA by non-homologous end-joining (NHEJ), yielding a mixture of precise and aberrant 5' junctions (79). In *D. melanogaster*, linear intron RNA retrohoming occurs by both major Lig4-dependent and minor Lig4-independent mechanisms, which appear to be related to classical and alternate NHEJ, respectively (80). The DNA repair polymerase θ plays a crucial role in both pathways, presumably by adding extra nucleotides to the 3' end of the intron cDNA to generate microhomologies that enable annealing of the cDNA end to the upstream exon. However, linear intron RNA retrohoming can occur independently of Ku70, which functions in capping chromosome ends during NHEJ.

RT-independent homing

A subset of filamentous fungal mt group IIB introns and a recently identified giant sulfur bacterial group IIC intron encode a LAGLIDADG DNA homing endonuclease instead of an RT (81–84). The LAGLIDADG ORF in the bacterial intron is located in DIV, whereas those in the fungal introns are located in DIIL, indicating two separate insertions. Biochemical assays show that both the ribozyme and endonuclease of the fungal introns are active (82). *S. cerevisiae* mtDNA introns with mutations in the RT active site have been shown to home efficiently by a conventional double-strand break repair mechanism, similar to group I introns (56, 69), and group II introns encoding LAGLIDADG proteins are assumed to use the same mechanism (85). Phylogenetic analysis indicates horizontal transfer of LAGLIDADG introns, suggesting that they are actively mobile in fungal populations (83).

Mechanism and regulation of group II intron retrotransposition

Whereas retrohoming is important for maintaining introns at a fixed site in genomes, retrotransposition to ectopic sites plays a major role in intron dissemination to novel sites, thereby increasing the diversity of intron-containing host genes on an evolutionary time-frame (58, 86–89). Retrotransposition is distinguished by relaxed sequence requirements for intron integration. Retrotransposition mechanisms of the Ll.LtrB intron have been analyzed in both *L. lactis* and *E. coli* by using powerful selection systems based on inserting a retrotransposition indicator gene (RIG) into intron DIV (58, 59). RIGs, which were first developed for other retroelements (90, 91), allow direct selection of a drug resistance marker after insertion into DNA via an RNA intermediate. It was thereby shown that retrotransposition of the Ll.LtrB intron in its native *L. lactis* host is biased toward the template for lagging-strand DNA synthesis, suggesting the replication fork as a source of ssDNA and Okazaki fragments as primers for cDNA synthesis (Fig. 3C, step 1). Many bacterial group II introns reside on the lagging-strand template, implicating a role for DNA replication in intron spread in nature. Also consistent with a role of the replication fork is the preponderance of events that occurs into plasmid targets, where plasmids have a higher number of forks per unit length of DNA than does the chromosome (92).

The bacterial host in which retrotransposition occurs not only influences the mobility pathway, but also the chromosomal location of intron-insertion sites. In contrast to *L. lactis*, retrotransposition of the LI.LtrB intron in *E. coli* is predominantly En-dependent occurring frequently into dsDNA. Furthermore, whereas events are scattered in *L. lactis*, with some preference for the replication terminus (*ter* domain), their localization in *E. coli* is bipolar, to the replication origin (*ori*) and *ter* domains (59, 74, 93). A polar pattern in *E. coli*, with preference for the *ori* domain, was also found for retrohoming of an LI.LtrB intron with randomized EBS and δ sequences (94) and attributed to localization of the IEP to the cellular poles (95), an explanation that may also apply to the bipolar pattern of retrotransposition events (74). These studies also gave the first hint that culture conditions can regulate retrotransposition, as slow growth eliminated all events except those around *ori* and *ter* and confined them to an endonuclease-dependent pathway into dsDNA (59). Additionally, a mutagenesis screen showed that LI.LtrB IEP localization is related to its interaction with negatively charged intracellular polyphosphates, which are normally pole-localized but delocalize away from the poles in response to cellular stress, leading to a more uniform distribution of intron-insertion sites (96). These studies portended growth phase and cellular stress as regulatory forces in retrotransposition.

Importantly, nutritional stress stimulates retrotransposition. Both amino acid and glucose starvation promote retrotransposition, via the small molecule alarmones ppGpp and cAMP, respectively (76, 97). Their stimulatory effects are indirect. Rather than promote RNP production, they appear to act on the target DNA, facilitating intron access. Precisely how the stress is transduced to the retroelement is an unanswered question. The same mutant screen that indicated a role for these alarmones and for RNase E showed that retromobility is stimulated by mutations in cytochrome oxidase, raising the possibility that oxidative stress promotes intron movement (76).

Both retrohoming and retrotransposition are suppressed by RNase E, which degrades the intron RNA (45, 72, 76). Because RNase E is part of the degradosome, which is tuned to the metabolic status of the cell, there is a potential for regulation of retromobility in response to changes in the physiological state of the cell. In *L. lactis*, the LI.LtrB intron associates with the ribosome, particularly the 30S subunit, which affords protection to the degradative silencing imposed by RNase E (45).

Finally, retrotransposition of the group II LI.LtrB intron, which resides within the relaxase gene of a conjugative plasmid, pRS01, is boosted in the presence of the plasmid (93, 98). Relaxase initiates conjugation by nicking DNA at the transfer origin and is a stimulatory factor for retrotransposition. By developing a genomic retrotransposition detection system called RIG-Seq, in which events are profiled by high-throughput sequencing, it was shown that relaxase, expressed either from pRS01 or from a plasmid vector in the absence of pRS01, increases both the frequency of retrotransposition and the diversity of DNA target sites. Specific point mutations and relaxase inhibitors support the hypothesis that relaxase elevates retrotransposition by inducing spurious nicks in the host DNA (93). Thus, the plasmid and retrotransposon act synergistically, with intron splicing required for relaxase expression and conjugation, and relaxase stimulating retrotransposition and intron spread via horizontal transfer.

DNA target site recognition

For retrohoming, group II intron RNPs recognize DNA target sequences by using both the IEP and base pairing of motifs within the intron RNA (56, 99–101). However, there are differences in detail for different intron subgroups and lineages. DNA target site interactions for representative group IIA, IIB, and IIC introns are diagrammed in Figure 4. In each case, the intron RNA base pairs to a central region of the DNA target site that encompasses the intron-insertion site, while the IEP recognizes exon sequences flanking the region recognized by intron. For each subgroup, the intron RNA motifs that base pair to the DNA target site are the same as those that base pair to flanking 5'- and 3'-exon sequences for RNA splicing (Fig. 1D), favoring intron insertion at sites that allow RNA splicing with minimal impairment to gene expression.

In group IIA and IIB introns, the intron RNA base-pairing interactions typically span 12–16 nts, while the IEP recognizes a small number of nucleotide residues in the distal 5' and 3' exons, which differ even for closely related introns (99, 102–104). The stringency of IEP recognition varies for different nucleotides and typically none is absolutely required for retrohoming. For the *S. cerevisiae* *ai2* and Ll.LtrB group IIA introns, IEP recognition of the sequences in the distal 5'-exon region likely involves the D domain and is needed for efficient reverse splicing into dsDNA but not ssDNA, indicating a contribution to local DNA melting (78, 99, 105). IEP recognition of the 3' exon is required only for bottom-strand cleavage (99, 101). Group II introns that reverse splice preferentially into transiently ssDNA, such as RmInt1, are less dependent upon IEP interactions with the distal 5' exon for DNA melting, and correspondingly their IEPs recognize fewer distal 5'-exon nucleotide residues (77, 106). In group IIC introns, intron RNA base pairing is limited to short EBS1/IBS1 and EBS3/IBS3 interactions, and the region upstream of IBS1 contains a hairpin structure corresponding to a transcription terminator or integron attachment (*attC*) site that is recognized by an as-yet unknown mechanism (107). The formation of this hairpin structure is presumably favored in transiently ssDNA at replication forks, thereby directing group II introns to this location where they can use nascent DNA strands to prime reverse transcription (see above).

For the Ll.LtrB intron, DNA target site interactions of RNPs have been investigated biochemically suggesting a model in which the RNPs bind DNA non-specifically and scan for target sites by facilitated diffusion along the DNA (105, 108). The IEP is thought to first recognize a small number of bases, including T-23, G-21, and A-20, in the distal 5' exon via major-groove interactions. These base contacts bolstered by phosphate-backbone interactions along one face of the helix promote local DNA melting, enabling the intron RNA to base pair to the target DNA (105). The small number of 5'-exon nucleotides recognized initially by the IEP suggests that the intron RNA may contribute to initial target site recognition, either by triplex formation or by processively base pairing to the IBS and δ' sequences. Bottom-strand cleavage between positions +9 and +10 of the DNA target site occurs after a lag and requires additional interactions of the IEP with the 3' exon, most critically recognition of T+5. Atomic force microscopy showed that the binding of RNPs bends the target DNA into two progressively sharper bend angles, the first correlated with 3'-exon interactions that position the scissile phosphate at the En active site for bottom-strand

cleavage, and the second with repositioning of the 3' end of the cleaved strand from the En to the RT active site for initiation of cDNA synthesis (42). Notably, even with this DNA bending, the distance between DNA target site residues T-23 and T+5 recognized by the IEP is longer than can be spanned readily by a single IEP molecule, suggesting either that T-23 and T+5 may be recognized sequentially by a single IEP or simultaneously by different subunits of an RT homodimer (17).

Retrotransposition to ectopic sites has relaxed sequence requirements. For all three classes of group II introns, the relatively limited contribution of IEP recognition facilitates dispersal to new target sites compatible with base-pairing interactions of the intron RNA. The Ll.LtrB intron in its native *L. lactis* host retrotransposes into ssDNA, and thus the residues recognized by the IEP for either DNA unwinding in the upstream exon (T-23, G-21, A-20) or for second-strand DNA cleavage in the downstream exon (T+5) are not required (58, 59, 92, 93). Furthermore, although there is a preference for native IBS1 and δ' sequences that base pair with sequences in the intron, there appears to be minimal dependence on an IBS2-EBS2 interaction. There is, however a strong bias toward -6C in IBS1, a residue also required for reverse splicing into dsDNA (59, 92, 93, 101).

Group II intron proliferation to high copy number

The distribution of group II introns in bacteria suggests waves of intron invasion, proliferation, and extinction, with the latter resulting from a combination of mutational inactivation, limiting intron spread, and removal of genomes harboring deleterious group II introns by purifying selection (109). As a result, group II intron distribution in bacteria is highly variable, even among different strains of the same species, with most bacteria containing only one or a small number of group II introns (109–112). In several organisms, however, group II introns have escaped host defenses and proliferated to high copy number in the genome. These systems are of interest because the evolution of spliceosomal introns is hypothesized to have involved proliferation of invasive group II introns to high copy number in the genomes of early eukaryotes.

Group II introns have been successful mobile elements not only because of their self-contained mobility apparatus, but also because they have evolved mechanisms for minimizing host damage, including high DNA target specificity, ability to remove themselves by RNA splicing and the sequestration of potentially deleterious IEPs by tight binding to the intron RNA. Factors that act to prevent group II intron proliferation include saturation of potential target sites (113), host defense mechanisms, including bacterial ribonucleases (see above), and negative selection due to deleterious effects on the host. Such deleterious effects might include inefficient splicing from some genomic sites, negative effects of high expression of group II intron RNPs, such as promiscuous breaks or integrations in chromosomal DNA, recombination between multiple dispersed copies of the introns, leading to genomic rearrangements (86, 87, 114, 115), and replicative disadvantage due to larger genome size.

The balance of factors that favor mobility and act against proliferation varies for different introns and different hosts. For example, group IIC introns have lower DNA target

specificity than do IIA or IIB introns and preferentially insert downstream of transcription terminators, enabling them to proliferate in some bacteria to moderately high copy number (116). Some strains of *Wolbachia*, a bacterial endosymbiont found in arthropods and insects, resemble organelles in containing relatively high numbers of group II introns of different families (115). This accumulation has been attributed to multiple intron invasions combined with inefficient purifying selection due to the limited population size of intracellular symbionts (115).

A dramatic example of group II intron proliferation to high copy number is provided by *Euglena gracilis* cpDNA, which contains more than 150 group II introns that interrupt almost every protein-coding gene (117, 118). Most of the introns are degenerate lacking elements of the conserved group II intron RNA structure, including entire RNA domains, and only two contain intact ORFs that could potentially encode functional proteins (49, 118). The smallest of these introns, referred to as group III introns, consist only of DI and DVI and lack DV, which is essential for ribozyme activity (117, 119). Comparisons of cpDNA sequences of different euglenoid species show that early branching species contain fewer introns and suggest at least two bursts of intron proliferation (120, 121). These proliferations may have been enabled by an efficient common splicing apparatus that provides key proteins and possibly RNA domains in *trans* (117). Other contributing factors include inefficient purifying selection against organellar genomes harboring group II introns and the ability of some group II introns to avoid host genes by inserting into other group II introns, forming “twintrons” (117). The euglenoid cpDNA introns most likely proliferated by a promiscuous reverse splicing-based retrotransposition mechanism that enables dispersal to large numbers of new sites, with bursts potentially reflecting acquisition of new actively mobile introns by horizontal transfer (122).

The thermophilic cyanobacterium *Thermosynechococcus elongatus* contains 28 closely related copies of a group IIB intron constituting ~1.3% of the genome (123). A combination of bioinformatics and mobility assays at different temperatures identified four mechanisms that contributed to the proliferation of *T. elongatus* group II introns (114). First, the *T. elongatus* introns have diverged into six families with different EBS sequences that target the introns to different sites. This divergence suggests waves of retrohoming following intron mutations that enable insertion into different target sites, thereby circumventing the limitation of target-site saturation. Second, some of the *T. elongatus* IEPs have evolved relaxed intron specificity and can efficiently splice and mobilize other introns that have lost their own ORFs, forming a rudimentary common splicing apparatus for multiple dispersed introns. Further, deletion of the intron ORF favors proliferation because the smaller, more compact ORF-less introns splice more efficiently and are less susceptible to nuclease degradation. More efficient splicing in turn enables intron insertion into housekeeping genes at lower fitness cost to the host. Third, some *T. elongatus* introns have evolved to insert at silent sites, such as a conserved sequence in another group II intron, leading to twintrons, or into an IS element. Finally, unlike mesophilic group II introns, the thermophilic *T. elongatus* introns rely on elevated temperatures to promote DNA strand separation, enabling access to a larger number of DNA target sites by base pairing of the intron RNA, with minimal requirement for recognition by the IEP to promote DNA melting. Thus, higher temperatures,

which are thought to have prevailed on Earth during the emergence of eukaryotes, favor intron proliferation by increasing the accessibility of DNA target sites.

Group II intron behavior in eukaryotes

Although mobile group II introns are thought to have proliferated in the nuclear genomes of early eukaryotes before evolving into spliceosomal introns (124, 125) and integration of organellar DNA fragments containing group II introns into the nuclear genome is an ongoing process (126, 127), functional group II introns have not been found in eukaryotic nuclear genomes. In order to survive a potentially overwhelming group II intron invasion early in their evolution, eukaryotes had to evolve mechanisms that restrain group II intron proliferation and expression. One such restraint may have been lower free Mg^{2+} concentrations, which limit group II intron ribozyme activity (128–130). Intracellular free Mg^{2+} concentrations are lower in eukaryotes (0.2–1 mM) than in bacteria (1–4 mM) and may be particularly low in eukaryotic nuclei, where Mg^{2+} is chelated to chromosomal DNA (130, 131). Efficient group II intron retrohoming in *X. laevis* oocyte nuclei or *D. melanogaster* or zebrafish embryos occurs only after injection of sufficient $MgCl_2$ (129). Spliceosomal introns evolved to function at the lower Mg^{2+} concentrations in eukaryotes, possibly by their disintegration into snRNAs, and by their increased reliance on protein cofactors, which can substitute for Mg^{2+} to promote RNA folding (130).

Studies in which the L1.LtrB group IIA intron and its IEP were inserted within the *S. cerevisiae* nuclear genes *CUP1* or *URA3* showed that the pre-mRNAs harboring the intron are not spliced in the nucleus but are spliced accurately in the cytoplasm (132), possibly reflecting higher free Mg^{2+} concentrations in that compartment. Interestingly, the mRNA from which L1.LtrB is spliced in the cytoplasm is subjected to nonsense-mediated decay, translational repression of mRNA, and targeting of the RNA to cytoplasmic foci (132, 133). By contrast, a spliceosomal intron assayed in parallel evaded these surveillance mechanisms (132). Insertion of the brown algal P1.LSU/2 group IIB intron into a yeast nuclear gene results in similar silencing of spliced mRNAs (134). Tenacious intermolecular interactions between the mRNA and the pre-mRNA and/or excised intron, based on IBS-EBS base pairings, contribute to translational silencing of the spliced mRNAs (133). Whereas disruption of these pairings by mutation promoted gene expression, compensatory mutations to restore the pairings again resulted in silencing. Thus, impaired expression of host genes containing inserted group II introns could have selected against ancestral eukaryotes carrying intact group II introns in their nuclear genomes and favored the evolution of spliceosomal introns.

Mobile group II intron-eukaryotic retrotransposon relationships

Mobile group II intron RTs are closely related to non-LTR-retrotransposons and telomerase RTs of eukaryotes (Fig. 2A). All three types of RT use analogous reverse transcription mechanisms, in which the site of initiation of cDNA synthesis is dictated by specific binding of the RT to the RNA template and the 3' end of a target DNA is used as the primer for reverse transcription (17, 135–138). Recent studies extend these similarities by showing that all three RTs employ regions upstream of the RT1 motif that are not present in retroviral RTs

for the specific binding of RNA templates (46, 139, 140) (Fig. 2A). In the case of group II introns and non-LTR-retrotransposons, the primer for TPRT is typically generated by cleavage of the target DNA by an En-domain appended to the RT (H-N-H for group II introns, restriction enzyme-like endonuclease (REL) for R2, and apurinic endonuclease (APE) for LINEs; Fig. 2A). The identity of the En is variable, with some nuclear non-LTR-retrotransposons encoding both APE and REL domains, and others, the *Penelope* retrotransposons, encoding a GIY-YIG endonuclease domain related to group I intron homing endonucleases (141–143). Both group II introns and non-LTR-retrotransposons can also use random nicks in DNA as primers for reverse transcription (93, 144, 145). In contrast, telomerase uses the 3' end of a chromosomal DNA as the primer, an ability shared by non-LTR-retrotransposons that lack En activity (142, 146, 147), but not yet demonstrated for mobile group II introns.

Like group II intron RTs, LINE-1 elements and other non-LTR-retrotransposon RTs show a *cis*-preference for mobilizing the element in which they are encoded, presumably reflecting that their RTs bind rapidly and preferentially to the RNA from which they are translated (148–150). The assembled RNP is then transported into the nucleus or nucleolus (for elements that insert into rDNA), which are sites of TPRT. Like group II intron RTs, mammalian LINE-1 elements and most other non-LTR-retrotransposon RTs lack an RNase H domain and may rely at least in part on a host RNase H for degradation of the RNA template strand in addition to strand displacement by the RT (136, 151). Additionally, the TPRT mechanism used by non-LTR-retrotransposons results in a cDNA whose 3' end may be ligated to upstream sequences by NHEJ enzymes (152), analogous to the retrohoming mechanism used by linear group II introns (see above), or linked to upstream sequences via template switching, a proficient activity of both group II intron and non-LTR-retrotransposon RTs (68, 136, 153, 154). Finally, the mechanism used for second-strand DNA synthesis by non-LTR-retrotransposons is not known, but given their nuclear localization could involve a host DNA polymerase, similar to mobile group II introns (73), rather than the DNA-dependent polymerase activity of the RT.

Structural and mechanistic similarities between group II and spliceosomal introns

Since the discovery of self-splicing RNAs and realization that group II introns splice via a lariat pathway, group II introns have been suspected of being the ancestors of spliceosomal introns (7, 155–157). In the ensuing years, evidence for this relationship has strengthened, based on shared splicing pathways and chemistries, together with sequence and structural similarities.

Structural similarities between group II and spliceosomal introns

The spliceosome comprises a plethora of proteins and five small nuclear RNAs (snRNAs), U1, U2, U4, U5 and U6 (158–160). During spliceosome assembly, there is a flux of both proteins and snRNAs with U2, U5 and U6 being the only snRNAs remaining in the catalytic particle (Fig. 5A). All three of these snRNAs have counterparts in the group II intron ribozyme, with U6 base-paired to U2 similar to the active site helix DV, U2 in complex with

the branch-point region of the intron resembling DVI, and U5 base pairing to 5' and 3' exons paralleling the DI-exon interactions of group II intron RNAs (Fig. 5A) (4, 161–163). Indeed, several group II intron domains (DIc, DIII, DI-DIV, DV) have been shown to be modular and promote splicing of group II introns lacking them when provided in *trans* (164–166) or to be interchangeable with analogous spliceosomal snRNAs: DV for U6 and U5 for DIId stem-loop containing EBS1 (162, 167). Additionally, there are many examples of naturally occurring fragmented group II introns that consist of two or three unlinked segments that reassemble for splicing *in vivo* (reviewed in (4, 157, 168)). Systematic studies with the L1.LtrB intron identified numerous functional breakpoints for bi- and tri-partite group II introns (169–171).

Besides these structural similarities, group II introns like spliceosomal introns can utilize alternative splicing with potential to generate different protein isoforms (172). Additionally, both group II and spliceosomal introns can splice via 5'-splice site hydrolysis when regions required for branching (DVI and U2 snRNA) are absent or compromised (173, 174).

Mechanistic similarities between group II and spliceosomal intron RNAs

The parallels between the group II autocatalytic pathway and spliceosome-mediated splicing (Fig. 1A), and the high degree of snRNA conservation over evolutionary time resulted in the proposal that the spliceosome is a ribozyme (175, 176). Details of the reaction in spliceosomes and group II introns are identical down to stereochemistry of the oxygen atoms that bind the two metals and the role of these divalent metals in stabilizing the leaving group in both steps of splicing (Fig. 5B & C) (12, 13, 175, 177). U6 base-paired with U2 catalyzes both steps of splicing by positioning divalent metals in precisely the same way as DV, with an internal stem-loop (ISL) of U6 resembling the DV helix in both structure and sequence (Fig. 5C) (178). In group II introns, the catalytic Mg^{2+} ions bind to DV via a CGC triad and bulge, whereas U6 contains an AGC triad and bulge that position the catalytic Mg^{2+} (Fig. 5B & C) (12, 163, 179). Finally, U6, like DV, uses base-triple interactions to present the catalytic metals to the scissile phosphates for the two steps of splicing (Fig. 5C–E) (163, 180). The Mg^{2+} -positioning triple helix in group II introns is formed between the DV triad and two nucleotides in the J2/3 linker (between DII and DIII), and one nucleotide in the DV bulge. In precise analogy, the U6 AGC triad, which base-pairs with U2, forms two base triples with the highly conserved 5'-ACAGAGA-3' box in U6, and one nucleotide in the U6 ISL bulge (Fig. 5C–E). These parallels leave little doubt as to an ancestral relationship between these two intron types.

Prp8 interacts with snRNAs at the active site of the spliceosome and has similarities to group II intron RTs

Beyond the structural and mechanistic similarities of group II intron and spliceosomal RNAs, recent studies have shown a relationship between group II intron RTs and Prp8, a highly conserved, 280-kDa protein at the heart of the spliceosome (181–183). Prp8 interacts with the three snRNAs, U2, U5 and U6, that form the catalytic core of the spliceosome (159, 184). The structure of a large fragment of *S. cerevisiae* Prp8 in complex with the U5 snRNP assembly factor Aar2 showed that the central conserved region contains RT and thumb/X domains related to those of group II intron RTs, separated by a relatively long linker from a

type II restriction enzyme-like (REL) En domain (Fig. 6). These domains are followed by an RNase H-like domain and a domain with a Jab1/MPN fold found in de-ubiquitinating enzymes. None of these domains appears to be enzymatically active, with the active site of the RT-like domain containing only one of three aspartates involved in Mg^{2+} coordination in active RTs.

The crystal structure of Prp8 shows that the RT, thumb/X, linker, En and RNase H-like domains form a large internal cavity that is the site of mutations that suppress splicing defects caused by mutations in the 5'- and 3'-splice sites and branch-point nucleotide, as well as the site of cross-links with the 5'- and 3'-splice sites, the branch point, and U5 and U6 snRNA (185, 186). Moreover, this internal cavity can accommodate the catalytic core of a group II intron RNA, which is thought to have evolved into that of the spliceosome (182). Based on these findings, it was suggested that Prp8 may have been derived from an ancestral group II intron RT that evolved to function as an assembly platform for snRNAs and pre-mRNA (182). The similarities of the RT and X/thumb domain suggest an evolutionary relationship to group II IEPs, with the lack of conservation of conserved aspartates at the RT active site consistent with the finding that mutations in the group II intron RT active site have minimal effect on splicing activity (19).

Plant nMat proteins provide previous examples of group II intron RTs that were subsumed into a nuclear genome without their associated intron RNAs and continue to function as splicing factors, albeit for group II introns within organelles (187–190). nMat2 proteins have a conventional H-N-H En domain, whereas in nMat1 proteins the En domain has been replaced by or diverged into a novel C-terminal domain that is conserved in different plant species and may contribute to the RNA splicing function (Fig. 6).

Despite these tantalizing similarities, a more complex situation is suggested by the finding that the En domain of Prp8 is a REL domain rather than an H-N-H endonuclease as in group II intron RTs and that group II intron RTs lack an associated RNase H domain (Fig. 6). As both REL and RNase H domains are found in some subgroups of nuclear non-LTR-retrotransposons (135, 191), these findings raise the possibility that Prp8 may be derived from another non-LTR-retroelement that is evolutionarily related to and possibly a missing link between group II introns and the spliceosome (192).

Group II intron evolution

Group II intron origins and other bacterial RT-containing elements

Although it is often conjectured that group II introns and other ribozymes arose in a primordial “RNA world”, wherein catalysis was performed by RNA, the relationship of group II introns to the RNA world remains uncertain (193). The order of events for evolving mobile group II introns is also obscure, and appears to have occurred independently in at least two ways: through acquisition of an RT that imparted RNA-based mobility needed for intron dispersal in bacteria, and through acquisition of a LAGLIDADG DNA endonuclease that likely conferred the ability to mobilize the intron via DNA-based recombination in both bacteria and fungi. In both cases, the DNA encoding the catalytic RNA splicing apparatus is assumed to have pre-existed an invasion event by different genes coding for mobilizing

activities. But for RT-driven mobility, a co-evolutionary scenario is also possible, where self-splicing ability developed from a retroelement under selective pressure to minimize transposon damage to the genome (194).

In addition to group II introns, bacteria contain a variety of other retroelements and RTs, which may have evolved from or into group II intron RTs (195, 196). These include retrons (197), abortive phage infection (Abi) retroelements (196), diversity generating retroelement RTs (DGRs) (198, 199) and RTs associated with some CRISPR-Cas systems (195, 196). Abi retroelements are host-encoded elements that defend against phage infection (200), while DGRs are bacterial and phage-encoded elements that diversify protein-encoding DNA sequences, such as those involved in cell surface display or phage tropism switching, via an error prone retrohoming mechanism (198, 201, 202). The RTs associated with CRISPR-Cas systems may function in cellular immunity by synthesizing cDNAs of RNA phages or transcripts of invading DNA elements that are then incorporated into the CRISPR repeats (195, 196). Thus bacterial RTs are engaged in the phage-host arms race, to evade host defenses on one hand and to develop immunity to invasive DNA on the other.

Role of group II introns in the origin of eukaryotes

A favored scenario for the origin of a eukaryotic cell is endosymbiosis, based on an archaeon hosting a bacterial invader that carried group II introns and evolved into mitochondria (124, 125, 203–205). Although this scenario remains a matter of debate particularly among those who favor bacterial-archaeal fusion scenarios for the origin of eukaryotes (206–208), the bacterial species thought to have evolved into mitochondria and chloroplasts, α -proteobacteria and cyanobacteria, respectively, both harbor copious numbers of group II introns (209). Bacteria contain all group II intron lineages whereas organelles contain only the IIA/ML and IIB/CL lineages, possibly reflecting that group II introns of these lineages were present in the bacteria that gave rise to mitochondria and chloroplasts (49). The early occurrence of spliceosomal introns in eukaryotes is supported by studies inferring the existence of at least a primitive spliceosome in the last common ancestor of extant eukaryotes (210).

Martin and Koonin (211) proposed that intron proliferation in a primitive eukaryote stimulated the evolution of the nuclear envelope in order to separate splicing from translation and prevent ribosomes from traversing the intron, resulting in mistranslated proteins, and to counter further intron invasions. At the same time, the separation of transcription and splicing in the nucleus from translation of spliced mRNAs in the cytosol necessitated the evolution of machinery that could function in *trans* to splice multiple dispersed introns. The presence of large numbers of introns in protein-coding genes also necessitated mechanisms like nonsense-mediated decay, which helps ensure that only accurately spliced mRNAs are translated.

Next, group II introns are hypothesized to have fragmented and reassembled into a spliceosome, which has been augmented by continued evolution into a megamachine. Consistent with this hypothesis is the absence of functional group II introns from extant nuclear genes, where they appear unable to function, and where silencing of gene expression

and nuclear compartmentalization may have added impetus to the evolution of the spliceosome through fragmentation and reassembly (130, 132, 133).

Evolution of intron-related retroelements and their role in sculpting our genomes

The discovery that group II intron retrohoming occurs by TPRT raised the possibility that group II introns are the ancestors of eukaryotic non-LTR retrotransposons and telomerase (54). This hypothesis is supported by the phylogenetic distribution of group II introns, which suggests that they evolved in eubacteria and entered eukaryotes with bacterial endosymbionts that evolved into mitochondria and chloroplasts. However, bacteria contain diverse retroelements and RTs (see above), and as well as single-copy *rvt* genes. RVTs are of unknown function and found sporadically in all eukaryotic kingdoms (195, 196, 212) (Fig. 6). Consistent with their single-copy nature, they are not associated with known endonucleases or integrases (212). These considerations leave one pondering whether cellular RT genes originated from retrotransposons or vice versa. Regardless, retroelements related to group II introns have sculpted our genomes, together with the spliceosomal introns to which these RNA catalysts almost certainly gave rise.

ACKNOWLEDGEMENTS

We thank Anne Pyle and Marco Marcia for structure representations in Figures 1 and 5, Matt Stanger and Georg Mohr for rendering the figures and Rebecca McCarthy and Susan Snyder for help with the manuscript. We thank Georg Mohr (UT Austin), John Moran (U. Michigan), Anna Pyle and Marco Marcia (Yale), Steven Zimmerly (U. Calgary) and Carol Lyn Piazza and Dorie Smith (U. Albany) for comments on the manuscript. Work in our labs is supported by NIH grants GM37949 and GM37951 and Welch Foundation Grant F-1607 to AML and NIH grants GM39422 and GM44844 to MB.

REFERENCES

1. Lambowitz AM, Zimmerly S. Mobile group II introns. *Annu Rev Genet.* 2004; 38:1–35. [PubMed: 15568970]
2. Pyle, AM, Lambowitz, AM. *The RNA World*. 3rd ed.. Cold Spring Harbor Laboratory Press; 2006. Group II introns: ribozymes that splice RNA and invade DNA; 469–505.
3. Toro N, Jiménez-Zurdo JJ, García-Rodríguez FM. Bacterial group II introns: not just splicing. *FEMS Microbiol Rev.* 2007; 31:342–358. [PubMed: 17374133]
4. Lambowitz AM, Zimmerly S. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harbor Perspectives in Biology.* 2011; 3:a003616. [PubMed: 20463000]
5. Marcia M, Somarowthu S, Pyle AM. Now on display: a gallery of group II intron structures at different stages of catalysis. *Mobile DNA.* 2013; 4:14. [PubMed: 23634971]
6. Enyeart PJ, Mohr G, Ellington AD, Lambowitz AM. Biotechnological applications of mobile group II introns and their reverse transcriptases: gene targeting, RNA-seq, and non-coding RNA analysis. *Mobile DNA.* 2014; 5:2. [PubMed: 24410776]
7. Peebles CL, Perlman PS, Mecklenburg KL, Petrillo ML, Tabor JH, Jarrell KA, Cheng H-L. A self-splicing RNA excises an intron lariat. *Cell.* 1986; 44:213–223. [PubMed: 3510741]
8. Jarrell KA, Peebles CL, Dietrich RC, Romiti SL, Perlman PS. Group II intron self-splicing. Alternative reaction conditions yield novel products. *J Biol Chem.* 1988; 263:3432–3439. [PubMed: 2830285]
9. Podar M, Chu VT, Pyle AM, Perlman PS. Group II intron splicing *in vivo* by first-step hydrolysis. *Nature.* 1998; 391:915–918. [PubMed: 9495347]
10. Pyle AM. The tertiary structure of group II introns: implications for biological function and evolution. *Crit Rev Biochem Mol Biol.* 2010; 45:215–232. [PubMed: 20446804]

11. Chin K, Pyle AM. Branch-point attack in group II introns is a highly reversible transesterification, providing a potential proofreading mechanism for 5'-splice site selection. *RNA*. 1995; 1:391–406. [PubMed: 7493317]
12. Toor N, Keating KS, Taylor SD, Pyle AM. Crystal structure of a self-spliced group II intron. *Science*. 2008; 320:77–82. [PubMed: 18388288]
13. Marcia M, Pyle AM. Visualizing group II intron catalysis through the stages of splicing. *Cell*. 2012; 151:497–507. [PubMed: 23101623]
14. Li CF, Costa M, Michel F. Linking the branchpoint helix to a newly found receptor allows lariat formation by a group II intron. *EMBO J*. 2011; 30:3040–3051. [PubMed: 21712813]
15. Somarowthu S, Legiewicz M, Keating KS, Pyle AM. Visualizing the ai5 γ group IIB intron. *Nucl Acids Res*. 2014; 42:1947–1958. [PubMed: 24203709]
16. Robart AR. Crystal structure of a eukaryotic group II intron lariat. *Nature*: Manuscript under review. 2014
17. Blocker FH, 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. [PubMed: 15574519]
18. Malik HS, Burke WD, Eickbush TH. The age and evolution of non-LTR retrotransposable elements. *Molecular Biology and Evolution*. 1999; 16:793–805. [PubMed: 10368957]
19. 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; 340:211–231. [PubMed: 15201048]
20. Moran JV, Zimmerly S, Eskes R, Kennell JC, Lambowitz AM, Butow RA, Perlman PS. Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. *Mol Cell Biol*. 1995; 15:2828–2838. [PubMed: 7537853]
21. Kennell JC, Moran JV, Perlman PS, Butow RA, Lambowitz AM. Reverse transcriptase activity associated with maturase-encoding group II introns in yeast mitochondria. *Cell*. 1993; 73:133–146. [PubMed: 7681727]
22. 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; 324:933–951. [PubMed: 12470950]
23. San Filippo, J. The DNA-binding and DNA endonuclease domains of a group II intron-encoded protein: characterization and application to the engineering of gene targeting vectors. University of Texas at Austin; 2003. Ph.D. Thesis
24. Molina-Sánchez M, Martínez-Abarca F, Toro N. Structural features in the C-terminal region of the *Sinorhizobium meliloti* RmInt1 group II intron-encoded protein contribute to its maturase and intron DNA-insertion function. *FEBS Journal*. 2010; 277:244–254. [PubMed: 19951359]
25. Toro N, Martínez-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. [PubMed: 23355907]
26. Gorbalenya AE. Self-splicing group I and group II introns encode homologous (putative) DNA endonucleases of a new family. *Protein Sci*. 1994; 3:1117–1120. [PubMed: 7920259]
27. Belfort M, Roberts RJ. Homing endonucleases: keeping the house in order. *Nucl Acids Res*. 1997; 25:3379–3388. [PubMed: 9254693]
28. Belfort M, Bonocora RP. Homing endonucleases: from genetic anomalies to programmable genomic clippers. :1–26. *Methods in Molecular Biology*. 2014; 1123 [PubMed: 24510256]
29. Stoddard BL. Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mobile DNA*. 2014; 5:7. [PubMed: 24589358]
30. 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; 38:9069–9083. [PubMed: 10413481]
31. Matsuura M, Noah JW, Lambowitz AM. Mechanism of maturase-promoted group II intron splicing. *EMBO J*. 2001; 20:7259–7270. [PubMed: 11743002]
32. Anziano PQ, Hanson DK, Mahler HR, Perlman PS. Functional domains in introns: trans-acting and cis-acting regions of intron 4 of the *cob* gene. *Cell*. 1982; 30:925–932. [PubMed: 6754094]

33. Carignani G, Groudinsky O, Frezza D, Schiavon E, Bergantino E, Slonimski PP. An mRNA maturase is encoded by the first intron of the mitochondrial gene for the subunit I of cytochrome oxidase in *S. cerevisiae*. *Cell*. 1983; 35:733–742. [PubMed: 6317200]
34. Moran JV, Mecklenburg KL, Sass P, Belcher SM, Mahnke D, Lewin A, Perlman P. Splicing defective mutants of the *COX1* gene of yeast mitochondrial DNA: initial definition of the maturase domain of the group II intron aI2. *Nucl Acids Res*. 1994; 22:2057–2064. [PubMed: 8029012]
35. Rambo RP, Doudna JA. Assembly of an active group II intron-maturase complex by protein dimerization. *Biochemistry*. 2004; 43:6486–6497. [PubMed: 15157082]
36. Gupta K, Contreras LM, Smith D, Qu G, Huang T, Spruce LA, Seeholzer SH, Belfort M, Van Duyne GD. Quaternary arrangement of an active, native group II intron ribonucleoprotein complex revealed by small-angle X-ray scattering. *Nucl Acids Res*. 2014; 42:5347–5360. [PubMed: 24567547]
37. 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; 4:239–250. [PubMed: 10488339]
38. Singh RN, Saldanha RJ, D'Souza LM, Lambowitz AM. Binding of a group II intron-encoded reverse transcriptase/maturase to its high affinity intron RNA binding site involves sequence-specific recognition and autoregulates translation. *J Mol Biol*. 2002; 318:287–303. [PubMed: 12051838]
39. Watanabe K, Lambowitz AM. High-affinity binding site for a group II intron-encoded reverse transcriptase/maturase within a stem-loop structure in the intron RNA. *RNA*. 2004; 10:1433–1443. [PubMed: 15273321]
40. Dai L, Chai D, Gu SQ, Gabel J, Noskov SY, Blocker FJ, Lambowitz AM, Zimmerly S. A three-dimensional model of a group II intron RNA and its interaction with the intron-encoded reverse transcriptase. *Mol Cell*. 2008; 30:472–485. [PubMed: 18424209]
41. Noah JW, Lambowitz AM. Effects of maturase binding and Mg^{2+} concentration on group II intron RNA folding investigated by UV cross-linking. *Biochemistry*. 2003; 42:12466–12480. [PubMed: 14580192]
42. Noah JW, Park S, Whitt JT, Perutka J, Frey W, Lambowitz AM. Atomic force microscopy reveals DNA bending during group II intron ribonucleoprotein particle integration into double-stranded DNA. *Biochemistry*. 2006; 45:12424–12435. [PubMed: 17029398]
43. Huang T, Shaikh TR, Gupta K, Contreras-Martin LM, Grassucci RA, Van Duyne GD, Frank J, Belfort M. The group II intron ribonucleoprotein precursor is a large, loosely packed structure. *Nucl Acids Res*. 2011; 39:2845–2854. [PubMed: 21131279]
44. Mohr S, Matsuura M, Perlman PS, Lambowitz AM. A DEAD-box protein alone promotes group II intron splicing and reverse splicing by acting as an RNA chaperone. *Proc Natl Acad Sci U S A*. 2006; 103:3569–3574. [PubMed: 16505350]
45. Contreras LM, Huang T, Piazza CL, Smith D, Qu G, Gelderman G, Potratz J, Russell R, Belfort M. Group II intron-ribosome association protects intron RNA from degradation. *RNA*. 2013; 19:1497–1509. [PubMed: 24046482]
46. Gu SQ, Cui X, Mou S, Mohr S, Yao J, Lambowitz AM. Genetic identification of potential RNA-binding regions in a group II intron-encoded reverse transcriptase. *RNA*. 2010; 16:732–747. [PubMed: 20179150]
47. Fontaine JM, Goux D, Kloareg B, Loiseaux-de Goër S. 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. [PubMed: 9010134]
48. Toor N, Hausner G, Zimmerly S. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptase. *RNA*. 2001; 7:1142–1152. [PubMed: 11497432]
49. Zimmerly S, Hausner G, Wu X. Phylogenetic relationships among group II intron ORFs. *Nucl Acids Res*. 2001; 29:1238–1250. [PubMed: 11222775]
50. Simon DM, Clarke NA, McNeil BA, Johnson I, Pantuso D, Dai L, Chai D, Zimmerly S. Group II introns in eubacteria and archaea: ORF-less introns and new varieties. *RNA*. 2008; 14:1704–1713. [PubMed: 18676618]

51. Simon DM, Kelchner SA, Zimmerly S. A broadscale phylogenetic analysis of group II intron RNAs and intron-encoded reverse transcriptases. *Molecular Biology and Evolution*. 2009; 26:2795–2808. [PubMed: 19713327]
52. Nagy V, Pirakitikulr N, Zhou KI, Chillón I, Luo J, Pyle AM. Predicted group II intron lineages E and F comprise catalytically active ribozymes. *RNA*. 2013; 19:1266–1278. [PubMed: 23882113]
53. Zimmerly S, Guo H, Eskes R, Yang J, Perlman PS, Lambowitz AM. A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell*. 1995; 83:529–538. [PubMed: 7585955]
54. Zimmerly S, Guo H, Perlman PS, Lambowitz AM. Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell*. 1995; 82:545–554. [PubMed: 7664334]
55. Yang J, Zimmerly S, Perlman PS, Lambowitz AM. Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature*. 1996; 381:332–335. [PubMed: 8692273]
56. 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; 88:865–874. [PubMed: 9118229]
57. Cousineau B, Smith D, Lawrence-Cavanagh S, Mueller JE, Yang J, Mills D, Manias D, Dunny G, Lambowitz AM, Belfort M. Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell*. 1998; 94:451–462. [PubMed: 9727488]
58. Ichianagi K, Beauregard A, Lawrence S, Smith D, Cousineau B, Belfort M. Retrotransposition of the L1.LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. *Mol Microbiol*. 2002; 46:1259–1272. [PubMed: 12453213]
59. Coros CJ, Landthaler M, Piazza CL, Beauregard A, Esposito D, Perutka J, Lambowitz AM, Belfort M. Retrotransposition strategies of the *Lactococcus lactis* L1.LtrB group II intron are dictated by host identity and cellular environment. *Mol Microbiol*. 2005; 56:509–524. [PubMed: 15813740]
60. Martínez-Abarca F, García-Rodríguez FM, Toro N. Homing of a bacterial group II intron with an intron-encoded protein lacking a recognizable endonuclease domain. *Mol Microbiol*. 2000; 35:1405–1412. [PubMed: 10760141]
61. Muñoz-Adelantado E, San Filippo J, Martínez-Abarca F, García-Rodríguez 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; 327:931–943. [PubMed: 12662921]
62. Martínez-Abarca F, Barrientos-Durán A, Fernández-López M, Toro N. The RmInt1 group II intron has two different retrohoming pathways for mobility using predominantly the nascent lagging strand at DNA replication forks for priming. *Nucl Acids Res*. 2004; 32:2880–2888. [PubMed: 15155857]
63. Candales MA, Duong A, Hood KS, Li T, Neufeld RA, Sun R, McNeil BA, Wu L, Jarding AM, Zimmerly S. Database for bacterial group II introns. *Nucl Acids Res*. 2012; 40:D187–D190. [PubMed: 22080509]
64. Zimmerly S, Moran JV, Perlman PS, Lambowitz AM. Group II intron reverse transcriptase in yeast mitochondria. Stabilization and regulation of reverse transcriptase activity by the intron RNA. *J Mol Biol*. 1999; 289:473–490. [PubMed: 10356323]
65. Huang HR, Chao MY, Armstrong B, Wang Y, Lambowitz AM, Perlman PS. The DIVa maturase binding site in the yeast group II intron aI2 is essential for intron homing but not for in vivo splicing. *Mol Cell Biol*. 2003; 23:8809–8819. [PubMed: 14612420]
66. Hu WS, Hughes SH. HIV-1 reverse transcription. *Cold Spring Harbor Perspectives in Medicine*. 2012; 2:a006882. [PubMed: 23028129]
67. Conlan LH, Stanger MJ, Ichianagi K, Belfort M. Localization, mobility and fidelity of retrotransposed group II introns in rRNA genes. *Nucl Acids Res*. 2005; 33:5262–5270. [PubMed: 16170154]
68. Mohr S, Ghanem E, Smith W, Sheeter D, Qin Y, King O, Polioudakis D, Iyer VR, Hunicke-Smith S, Swamy S, Kuersten S, Lambowitz AM. Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA*. 2013; 19:958–970. [PubMed: 23697550]

69. Eskes R, Liu L, Ma H, Chao MY, Dickson L, Lambowitz AM, Perlman PS. Multiple homing pathways used by yeast mitochondrial group II introns. *Mol Cell Biol.* 2000; 20:8432–8446. [PubMed: 11046140]
70. Lazowska J, Meunier B, Macadre C. Homing of a group II intron in yeast mitochondrial DNA is accompanied by unidirectional co-conversion of upstream-located markers. *EMBO J.* 1994; 13:4963–4972. [PubMed: 7525273]
71. Mills DA, Manias DA, McKay LL, Dunny GM. Homing of a group II intron from *Lactococcus lactis* subsp. *lactis* ML3. *J Bacteriol.* 1997; 179:6107–6111. [PubMed: 9324259]
72. 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; 19:2477–2487. [PubMed: 16230535]
73. Yao J, Truong DM, Lambowitz AM. Genetic and biochemical assays reveal a key role for replication restart proteins in group II intron retrohoming. *PLoS Genet.* 2013; 9:e1003469. [PubMed: 23637634]
74. Beauregard A, Chalamcharla VR, Piazza CL, Belfort M, Coros CJ. Bipolar localization of the group II intron Ll.LtrB is maintained in *Escherichia coli* deficient in nucleoid condensation, chromosome partitioning and DNA replication. *Mol Microbiol.* 2006; 62:709–722. [PubMed: 17005014]
75. Beauregard A, Curcio MJ, Belfort M. The take and give between retrotransposable elements and their hosts. *Annu Rev Genet.* 2008; 42:587–617. [PubMed: 18680436]
76. Coros CJ, Piazza CL, Chalamcharla VR, Belfort M. A mutant screen reveals RNase E as a silencer of group II intron retromobility in *Escherichia coli*. *RNA.* 2008; 14:2634–2644. [PubMed: 18945808]
77. Jiménez-Zurdo JI, García-Rodríguez FM, Barrientos-Durán 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; 326:413–423. [PubMed: 12559910]
78. Zhong J, Lambowitz AM. Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription. *EMBO J.* 2003; 22:4555–4565. [PubMed: 12941706]
79. Zhuang F, Mastroianni M, White TB, Lambowitz AM. Linear group II intron RNAs can retrohome in eukaryotes and may use nonhomologous end-joining for cDNA ligation. *Proc Natl Acad Sci U S A.* 2009; 106:18189–18194. [PubMed: 19833873]
80. White TB, Lambowitz AM. The retrohoming of linear group II intron RNAs in *Drosophila melanogaster* occurs by both DNA ligase 4-dependent and -independent mechanisms. *PLoS Genet.* 2012; 8:e1002534. [PubMed: 22359518]
81. Toor N, Zimmerly S. Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. *RNA.* 2002; 8:1373–1377. [PubMed: 12458791]
82. 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; 16:1818–1831. [PubMed: 20656798]
83. 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; 72:433–449. [PubMed: 21479820]
84. 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; 109:4203–4208. [PubMed: 22371583]
85. Lambowitz AM, Belfort M. Introns as mobile genetic elements. *Annu Rev Biochem.* 1993; 62:587–622. [PubMed: 8352597]
86. Mueller MW, Allmaier M, Eskes R, Schweyen RJ. Transposition of group II intron *aII* in yeast and invasion of mitochondrial genes at new locations. *Nature.* 1993; 366:174–176. [PubMed: 8232557]
87. Sellem CH, Lecellier G, Belcour L. Transposition of a group II intron. *Nature.* 1993; 366:176–178. [PubMed: 8232558]
88. Martínez-Abarca F, Toro N. RecA-independent ectopic transposition *in vivo* of a bacterial group II intron. *Nucl Acids Res.* 2000; 28:4397–4402. [PubMed: 11058141]

89. Dickson L, Huang HR, Liu L, Matsuura M, Lambowitz AM, Perlman PS. Retrotransposition of a yeast group II intron occurs by reverse splicing directly into ectopic DNA sites. *Proc Natl Acad Sci U S A*. 2001; 98:13207–13212. [PubMed: 11687644]
90. Curcio MJ, Garfinkel DJ. Single-step selection for Ty1 element retrotransposition. *Proc Natl Acad Sci U S A*. 1991; 88:936–940. [PubMed: 1846969]
91. Heidmann T, Heidmann O, Nicolas JF. An indicator gene to demonstrate intracellular transposition of defective retroviruses. *Proc Natl Acad Sci U S A*. 1988; 85:2219–2223. [PubMed: 2832848]
92. Ichiyanagi K, Beauregard A, Belfort M. A bacterial group II intron favors retrotransposition into plasmid targets. *Proc Natl Acad Sci U S A*. 2003; 100:15742–15747. [PubMed: 14673083]
93. Novikova O, Smith D, Hahn I, Beauregard A, Belfort M. Synergy between conjugative and retrotransposable elements in horizontal gene transfer. Submitted. 2014
94. Zhong J, Karberg M, Lambowitz AM. Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucl Acids Res*. 2003; 31:1656–1664. [PubMed: 12626707]
95. Zhao J, Lambowitz AM. A bacterial group II intron-encoded reverse transcriptase localizes to cellular poles. *Proc Natl Acad Sci U S A*. 2005; 102:16133–16140. [PubMed: 16186487]
96. Zhao J, Niu W, Yao J, Mohr S, Marcotte EM, Lambowitz AM. Group II intron protein localization and insertion sites are affected by polyphosphate. *PLoS Biology*. 2008; 6:e150. [PubMed: 18593213]
97. Coros CJ, Piazza CL, Chalamcharla VR, Smith D, Belfort M. Global regulators orchestrate group II intron retromobility. *Mol Cell*. 2009; 34:250–256. [PubMed: 19394301]
98. Belhocine K, Plante I, Cousineau B. Conjugation mediates transfer of the Ll.LtrB group II intron between different bacterial species. *Mol Microbiol*. 2004; 51:1459–1469. [PubMed: 14982638]
99. 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; 16:6835–6848. [PubMed: 9362497]
100. Guo H, Karberg M, Long M, Jones JP III, Sullenger B, Lambowitz AM. Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science*. 2000; 289:452–457. [PubMed: 10903206]
101. 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; 14:559–573. [PubMed: 10716944]
102. Yang J, Mohr G, Perlman PS, Lambowitz AM. Group II intron mobility in yeast mitochondria: target DNA-primed reverse transcription activity in aI1 and reverse splicing into DNA transposition sites *in vitro*. *J Mol Biol*. 1998; 282:505–523. [PubMed: 9737919]
103. Perutka J, Wang W, Goerlitz D, Lambowitz AM. Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. *J Mol Biol*. 2004; 336:421–439. [PubMed: 14757055]
104. Zhuang F, Karberg M, Perutka J, Lambowitz AM. EcI5, a group IIB intron with high retrohoming frequency: DNA target site recognition and use in gene targeting. *RNA*. 2009; 15:432–449. [PubMed: 19155322]
105. 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; 309:361–386. [PubMed: 11371159]
106. Barrientos-Durán A, Chillón I, Martínez-Abarca F, Toro N. Exon sequence requirements for excision *in vivo* of the bacterial group II intron RmInt1. *BMC Mol Biol*. 2011; 12:12–24. [PubMed: 21453489]
107. Robart AR, Seo W, Zimmerly S. Insertion of group II intron retroelements after intrinsic transcriptional terminators. *Proc Natl Acad Sci U S A*. 2007; 104:6620–6625. [PubMed: 17420455]
108. Aizawa Y, Ziang Q, Lambowitz AM, Pyle AM. The pathway for DNA recognition and RNA integration by a group II intron retrotransposon. *Mol. Cell*. 2003; 11:795–805. [PubMed: 12667460]

109. Leclercq S, Cordaux R. Selection-driven extinction dynamics for group II introns in *Enterobacteriales*. PLoS One. 2012; 7:e52268. [PubMed: 23251705]
110. Dai L, Zimmerly S. Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. Nucl Acids Res. 2002; 30:1091–1102. [PubMed: 11861899]
111. Fernández-López M, Muñoz-Adelantado E, Gillis M, Willems A, Toro N. Dispersal and evolution of the *Sinorhizobium meliloti* group II RmInt1 intron in bacteria that interact with plants. Molecular Biology and Evolution. 2005; 22:1518–1528. [PubMed: 15814827]
112. Tourasse NJ, Kolst AB. Survey of group I and group II introns in 29 sequenced genomes of the *Bacillus cereus* group: insights into their spread and evolution. Nucl Acids Res. 2008; 36:4529–4548. [PubMed: 18587153]
113. Goddard MR, Burt A. Recurrent invasion and extinction of a selfish gene. Proc Natl Acad Sci U S A. 1999; 96:13880–13885. [PubMed: 10570167]
114. Mohr G, Ghanem E, Lambowitz AM. Mechanisms used for genomic proliferation by thermophilic group II introns. PLoS Biology. 2010; 8:e1000391. [PubMed: 20543989]
115. Leclercq S, Giraud I, Cordaux R. Remarkable abundance and evolution of mobile group II introns in *Wolbachia* bacterial endosymbionts. Molecular Biology and Evolution. 2011; 28:685–697. [PubMed: 20819906]
116. Ueda K, Yamashita A, Ishikawa J, Shimada M, Watsuji TO, Morimura K, Ikeda H, Hattori M, Beppu T. Genome sequence of *Symbiobacterium thermophilum*, an uncultivable bacterium that depends on microbial commensalism. Nucl Acids Res. 2004; 32:4937–4944. [PubMed: 15383646]
117. Copertino DW, Hallick RB. Group II and group III introns of twintrons: potential relationships to nuclear pre-mRNA introns. Trends Biochem Sci. 1993; 18:467–471. [PubMed: 8108859]
118. Hallick RB, Hong L, Drager RG, Favreau MR, Monfort A, Orsat B, Spielmann A, Stutz E. Complete sequence of *Euglena gracilis* chloroplast DNA. Nucl Acids Res. 1993; 21:3537–3544. [PubMed: 8346031]
119. Christopher DA, Hallick RB. *Euglena gracilis* chloroplast ribosomal protein operon: a new chloroplast gene for ribosomal protein L5 and description of a novel organelle intron category designated group III. Nucl Acids Res. 1989; 17:7591–7608. [PubMed: 2477800]
120. Pombert JF, James ER, Janoukovec J, Keeling PJ. Evidence for transitional stages in the evolution of euglenid group II introns and twintrons in the *Monomorpha aenigmatica* plastid genome. PLoS One. 2012; 7:e53433. [PubMed: 23300929]
121. Wiegert KE, Bennett MS, Triemer RE. Tracing patterns of chloroplast evolution in euglenoids: contributions from *Colacium vesiculosum* and *Strombomonas acuminata* (Euglenophyta). J Euk Micro. 2013; 60:214–221.
122. Sheveleva EV, Hallick RB. Recent horizontal intron transfer to a chloroplast genome. Nucl Acids Res. 2004; 32:803–810. [PubMed: 14762207]
123. Nakamura Y, Kaneko T, Sato S, Ikeuchi M, Katoh H, Sasamoto S, Watanabe A, Iriguchi M, Kawashima K, Kimura T, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Nakazaki N, Shimpō S, Sugimoto M, Takeuchi C, Yamada M, Tabata S. Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. DNA Res. 2002; 9:123–130. [PubMed: 12240834]
124. Cavalier-Smith T. Intron phylogeny: a new hypothesis. Trends Genet. 1991; 7:145–148. [PubMed: 2068786]
125. Palmer JD, Logsdon JM Jr. The recent origins of introns. Curr Opin Genet Devel. 1991; 1:470–477. [PubMed: 1822279]
126. Knoop V, Brennicke A. Promiscuous mitochondrial group II intron sequences in plant nuclear genomes. J Mol Evol. 1994; 39:144–150. [PubMed: 7932778]
127. Lin X, Kaul S, Rounsley S, Shea TP, Benito MI, Town CD, Fujii CY, Mason T, Bowman CL, Barnstead M, Feldblyum TV, Buell CR, Ketchum KA, Lee J, Ronning CM, Koo HL, Moffat KS, Cronin LA, Shen M, Pai G, Van Aken S, Umayam L, Tallon LJ, Gill JE, Adams MD, Carrera AJ, Creasy TH, Goodman HM, Somerville CR, Copenhaver GP, Preuss D, Niernan WC, White O, Eisen JA, Salzberg SL, Fraser CM, Venter JC. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature. 1999; 402:761–768. [PubMed: 10617197]

128. Gregan J, Kolisek M, Schweyen RJ. Mitochondrial Mg^{2+} homeostasis is critical for group II intron splicing in vivo. *Genes Dev.* 2001; 15:2229–2237. [PubMed: 11544180]
129. Mastroianni M, Watanabe K, White TB, Zhuang F, Vernon J, Matsuura M, Wallingford J, Lambowitz AM. Group II intron-based gene targeting reactions in eukaryotes. *PLoS One.* 2008; 3:e3121. [PubMed: 18769669]
130. Truong DM, Sidote DJ, Russell R, Lambowitz AM. Enhanced group II intron retrohoming in magnesium-deficient *Escherichia coli* via selection of mutations in the ribozyme core. *Proc Natl Acad Sci U S A.* 2013; 110:E3800–E3809. [PubMed: 24043808]
131. Günther T. AtConcentration, compartmentation and metabolic function of intracellular free Mg^{2+} Magnesium Research. 2006; 19:225–236. [PubMed: 17402290]
132. Chalamcharla VR, Curcio MJ, Belfort M. Nuclear expression of a group II intron is consistent with spliceosomal intron ancestry. *Genes Dev.* 2010; 24:827–836. [PubMed: 20351053]
133. Qu G, Dong X, Piazza CL, Chalamcharla V, Lutz S, Curcio MJ, Belfort M. RNA-RNA interactions and pre-RNA mislocalization as the drivers of group II intron loss from nuclear genomes. *Proc Natl Acad Sci U S A.* 2014; 111:6612–6617. [PubMed: 24722636]
134. Zerbato M, Holic N, Moniot-Frin S, Ingrao D, Galy A, Perea J. The brown algae *PLSU/2* group II intron-encoded protein has functional reverse transcriptase and maturase activities. *PLoS One.* 2013; 8:e58263. [PubMed: 23505475]
135. Eickbush, T, Malik, H. Origins and Evolution of Retrotransposons. In: Craig, N, Craigie, R, Gellert, M, Lambowitz, A, editors. *Mobile DNA II*. ASM Press; 2002. 1111–1146.
136. Eickbush TH, Eickbush DG. Integration, regulation, and long-term stability of R2 retrotransposons. *Mobile DNA III*: In Press. 2014
137. Fujiwara HS. Site-specific non-LTR elements. *Mobile DNA III*: In Press. 2014
138. Moran JV. LINE retrotransposons. *Mobile DNA III*: In press. 2014
139. Mitchell M, Gillis A, Futahashi M, Fujiwara H, Skordalakes E. Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. *Nat Struct Mol Biol.* 2010; 17:513–518. [PubMed: 20357774]
140. Jamburuthugoda VK, Eickbush TH. Identification of RNA binding motifs in the R2 retrotransposon-encoded reverse transcriptase. *Nucl Acids Res.* 2014; 42:8405–8415. [PubMed: 24957604]
141. Kojima KK, Fujiwara H. An extraordinary retrotransposon family encoding dual endonucleases. *Genome Res.* 2005; 15:1106–1117. [PubMed: 16077010]
142. Gladyshev EA, Arkhipova IR. Telomere-associated endonuclease-deficient *Penelope*-like retroelements in diverse eukaryotes. *Proc Natl Acad Sci U S A.* 2007; 104:9352–9357. [PubMed: 17483479]
143. Arkhipova IR, Yushenova IA, Rodriguez F. Endonuclease-containing *Penelope* retrotransposons in the bdelloid rotifer *Adineta vaga* exhibit unusual structural features and play a role in expansion of host gene families. *Mobile DNA.* 2013; 4:19. [PubMed: 23981484]
144. Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat Genet.* 2002; 31:159–165. [PubMed: 12006980]
145. Onozawa M, Zhang Z, Kim YJ, Goldberg L, Varga T, Bergsagel PL, Kuehl WM, Aplan PD. Repair of DNA double-strand breaks by templated nucleotide sequence insertions derived from distant regions of the genome. *Proc Natl Acad Sci U S A.* 2014; 111:7729–7734. [PubMed: 24821809]
146. Morrish TA, Garcia-Perez JL, Stamato TD, Taccioli GE, Sekiguchi J, Moran JV. Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature.* 2007; 446:208–212. [PubMed: 17344853]
147. Curcio MJ, Belfort M. The beginning of the end: Links between ancient retroelements and modern telomerases. *Proc Nat Acad Sci U S A.* 2007; 104:9107–9108.
148. Esnault C, Maestre J, Heidmann T. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet.* 2000; 24:363–367. [PubMed: 10742098]

149. Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH, Boeke JD, Moran JV. Human L1 retrotransposition: *cis* preference versus *trans* complementation. *Mol Cell Biol*. 2001; 21:1429–1439. [PubMed: 11158327]
150. Kulpa DA, Moran JV. *Cis*-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat Struct Mol Biol*. 2006; 13:655–660. [PubMed: 16783376]
151. Richardson SR, Narvaiza I, Planegger RA, Weitzman MD, Moran JV. APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. *eLife*. 2014; 3:e02008. [PubMed: 24843014]
152. Suzuki J, Yamaguchi K, Kajikawa M, Ichianagi K, Adachi N, Koyama H, Takeda S, Okada N. Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. *PLoS Genet*. 2009; 5:e1000461. [PubMed: 19390601]
153. Bibillo A, Eickbush TH. The reverse transcriptase of the R2 non-LTR retrotransposon: continuous synthesis of cDNA on non-continuous RNA templates. *J Mol Biol*. 2002; 316:459–473. [PubMed: 11866511]
154. Bibillo A, Eickbush TH. End-to-end template jumping by the reverse transcriptase encoded by the R2 retrotransposon. *J Biol Chem*. 2004; 279:14945–14953. [PubMed: 14752111]
155. Sharp PA. On the origin of RNA splicing and introns. *Cell*. 1985; 42:397–400. [PubMed: 2411416]
156. Cech TR. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell*. 1986; 44:207–210. [PubMed: 2417724]
157. Sharp PA. Five easy pieces. *Science*. 1991; 254:663. [PubMed: 1948046]
158. Valadkhan S. Role of the snRNAs in spliceosomal active site. *RNA Biology*. 2010; 7:345–353. [PubMed: 20458185]
159. Will CL, Lührmann R. Spliceosome structure and function. *Cold Spring Harbor Perspectives in Biology*. 2011; 3:a003707. [PubMed: 21441581]
160. Valadkhan S. The role of snRNAs in spliceosomal catalysis. *Progress in Molecular Biology and Translational Science*. 2013; 120:195–228. [PubMed: 24156945]
161. 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; 71:803–817. [PubMed: 1423631]
162. Shukla GC, Padgett RA. A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome. *Mol Cell*. 2002; 9:1145–1150. [PubMed: 12049749]
163. Keating KS, Toor N, Perlman PS, Pyle AM. A structural analysis of the group II intron active site and implications for the spliceosome. *RNA*. 2010; 16:1–9. [PubMed: 19948765]
164. Jarrell KA, Dietrich RC, Perlman PS. Group II intron domain 5 facilitates a *trans*-splicing reaction. *Mol Cell Biol*. 1988; 8:2361–2366. [PubMed: 3405208]
165. Goldschmidt-Clermont M, Choquet Y, Girard-Bascou J, Michel F, Schirmer-Rahire M, Rochaix J-D. A small chloroplast RNA may be required for *trans*-splicing in *Chlamydomonas reinhardtii*. *Cell*. 1991; 65:135–143. [PubMed: 1707343]
166. Suchy M, Schmelzer C. Restoration of the self-splicing activity of a defective group II intron by a small *trans*-acting RNA. *J Mol Biol*. 1991; 222:179–187. [PubMed: 1960721]
167. Hetzer M, Wurzer G, Schweyen RJ, Mueller MW. *Trans*-activation of group II intron splicing by nuclear U5 snRNA. *Nature*. 1997; 386:417–420. [PubMed: 9121561]
168. Glanz S, Kück U. *Trans*-splicing of organelle introns—a detour to continuous RNAs. *Bioessays*. 2009; 31:921–934. [PubMed: 19644921]
169. Belhocine K, Mak AB, Cousineau B. *Trans*-splicing versatility of the Ll.LtrB group II intron. *RNA*. 2008; 14:1782–1790. [PubMed: 18648072]
170. Quiroga C, Kronstad L, Ritlop C, Filion A, Cousineau B. Contribution of base-pairing interactions between group II intron fragments during *trans*-splicing in vivo. *RNA*. 2011; 17:2212–2221. [PubMed: 22033330]
171. 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. [PubMed: 22876289]

172. McNeil BA, Simon DM, Zimmerly S. Alternative splicing of a group II intron in a surface layer protein gene in *Clostridium tetani*. Nucl Acids Res. 2014; 42:1959–1969. [PubMed: 24214997]
173. Chu VT, Liu Q, Podar M, Perlman PS, Pyle AM. More than one way to splice an RNA: branching without a bulge and splicing without branching in group II introns. RNA. 1998; 4:1186–1202. [PubMed: 9769094]
174. Tseng CK, Cheng SC. Both catalytic steps of nuclear pre-mRNA splicing are reversible. Science. 2008; 320:1782–1784. [PubMed: 18583613]
175. Gordon PM, Sontheimer EJ, Piccirilli JA. Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: extending the parallels between the spliceosome and group II introns. RNA. 2000; 6:199–205. [PubMed: 10688359]
176. Collins CA, Guthrie C. The question remains: is the spliceosome a ribozyme? Nat Struct Biol. 2000; 7:850–854. [PubMed: 11017191]
177. Sontheimer EJ, Sun S, Piccirilli JA. Metal ion catalysis during splicing of premessenger RNA. Nature. 1997; 388:801–805. [PubMed: 9285595]
178. Fica SM, Tuttle N, Novak T, Li NS, Lu J, Koodathingal P, Dai Q, Staley JP, Piccirilli JA. RNA catalyses nuclear pre-mRNA splicing. Nature. 2013; 503:229–234. [PubMed: 24196718]
179. Gordon PM, Piccirilli JA. Metal ion coordination by the AGC triad in domain 5 contributes to group II intron catalysis. Nat Struct Biol. 2001; 8:893–898. [PubMed: 11573097]
180. 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; 21:464–471. [PubMed: 24747940]
181. 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. [PubMed: 21441348]
182. Galej WP, Oubridge C, Newman AJ, Nagai K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. Nature. 2013; 493:638–643. [PubMed: 23354046]
183. Galej WP, Nguyen TH, Newman AJ, Nagai K. Structural studies of the spliceosome: zooming into the heart of the machine. Curr Opin Struct Biol. 2014; 25C:57–66.
184. Anokhina M, Bessonov S, Miao Z, Westhof E, Hartmuth K, Lührmann R. RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. EMBO J. 2013; 32:2804–2818. [PubMed: 24002212]
185. Reyes JL, Gustafson EH, Luo HR, Moore MJ, Konarska MM. The C-terminal region of hPrp8 interacts with the conserved GU dinucleotide at the 5' splice site. RNA. 1999; 5:167–179. [PubMed: 10024169]
186. Turner IA, Norman CM, Churcher MJ, Newman AJ. Dissection of Prp8 protein defines multiple interactions with crucial RNA sequences in the catalytic core of the spliceosome. RNA. 2006; 12:375–386. [PubMed: 16431982]
187. Mohr G, Lambowitz AM. Putative proteins related to group II intron reverse transcriptase/maturases are encoded by nuclear genes in higher plants. Nucl Acids Res. 2003; 31:647–652. [PubMed: 12527773]
188. 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; 47:772–783. [PubMed: 16621844]
189. Keren I, Bezawork-Geleta A, Kolton M, Maayan I, Belausov E, Levy M, Mett A, Gidoni D, Shaya F, Ostersetter-Biran O. AtnMat2, a nuclear-encoded maturase required for splicing of group-II introns in *Arabidopsis* mitochondria. RNA. 2009; 15:2299–2311. [PubMed: 19946041]
190. Brown GG, Colas des Francs-Small C, Ostersetter-Biran O. Group II intron splicing factors in plant mitochondria. Frontiers in Plant Science. 2014; 5:35. [PubMed: 24600456]
191. Smyshlyaev G, Voigt F, Blinov A, Barabas O, Novikova O. Acquisition of an Archaea-like ribonuclease H domain by plant L1 retrotransposons supports modular evolution. Proc Nat Acad Sci U S A. 2013; 110:20140–20145.
192. Query CC, Konarska MM. Spliceosome's core exposed. Nature. 2013; 493:615–616. [PubMed: 23354053]

193. Doolittle WF. The spliceosomal catalytic core arose in the RNA world. or did it? *Genome Biol.* 2013; 14:141. [PubMed: 24330866]
194. Curcio MJ, Belfort M. Retrohoming: cDNA-mediated mobility of group II introns requires a catalytic RNA *Cell.* 1996; 84:9–12. [PubMed: 8548830]
195. Kojima KK, Kanehisa M. Systematic survey for novel types of prokaryotic retroelements based on gene neighborhood and protein architecture. *Mol Biol Evol.* 2008; 25:1395–1404. [PubMed: 18391066]
196. Simon DM, Zimmerly S. A diversity of uncharacterized reverse transcriptases in bacteria. *Nucl Acids Res.* 2008; 36:7219–7229. [PubMed: 19004871]
197. Lampson BC, Inouye M, Inouye S. Retrons, msDNA, and the bacterial genome. *Cytogenet Genome Res.* 2005; 110:491–499. [PubMed: 16093702]
198. Doulatov S, Hodes A, Dai L, Mandhana N, Liu M, Deora R, Simons RW, Zimmerly S, Miller JF. Tropism switching in *Bordetella* bacteriophage defines a family of diversity-generating retroelements. *Nature.* 2004; 431:476–481. [PubMed: 15386016]
199. Guo H, Arambula D, Ghosh P, Miller JF. Diversity-generating retroelements in phage and bacterial genomes. *Mobile DNA III*:In press. 2014
200. Fortier LC, Bouchard JD, Moineau S. Expression and site-directed mutagenesis of the lactococcal abortive phage infection protein AbiK. *J Bacteriol.* 2005; 187:3721–3730. [PubMed: 15901696]
201. Medhekar B, Miller JF. Diversity-generating retroelements. *Curr Opin Microbiol.* 2007; 10:388–395. [PubMed: 17703991]
202. Minot S, Grunberg S, Wu GD, Lewis JD, Bushman FD. Hypervariable loci in the human gut virome. *Proc Natl Acad Sci U S A.* 2012; 109:3962–3966. [PubMed: 22355105]
203. Martin W, Koonin EV. Introns and the origin of nucleus-cytosol compartmentalization. *Nature.* 2006; 440:41–45. [PubMed: 16511485]
204. Williams TA, Foster PG, Cox CJ, Embley TM. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature.* 2013; 504:231–236. [PubMed: 24336283]
205. Irimia M, Roy SW. Origin of spliceosomal introns and alternative splicing. *Cold Spring Harbor Perspectives in Biology.* 2014; 6:a01607.
206. Poole AM. Did group II intron proliferation in an endosymbiont-bearing archaeon create eukaryotes? *Biology Direct.* 2006; 1:36. [PubMed: 17156426]
207. Koonin EV. Intron-dominated genomes of early ancestors of eukaryotes. *The Journal of Heredity.* 2009; 100:618–623. [PubMed: 19617525]
208. Doolittle WF. The trouble with (group II) introns. *Proc Natl Acad Sci U S A.* 2014; 111:6536–6537. [PubMed: 24757059]
209. Robart AR, Zimmerly S. Group II intron retroelements: function and diversity. *Cytogenet Genome Res.* 2005; 110:589–597. [PubMed: 16093712]
210. Collins L, Penny D. Complex spliceosomal organization ancestral to extant eukaryotes. *Molecular Biology and Evolution.* 2005; 22:1053–1066. [PubMed: 15659557]
211. Martin W, Koonin EV. A positive definition of prokaryotes. *Nature.* 2006; 442:868.
212. Gladyshev EA, Arkipova IR. A widespread class of reverse transcriptase-related cellular genes. *Proc Natl Acad Sci U S A.* 2011; 108:20311–20316. [PubMed: 21876125]
213. Costa M, Michel F, Westhof E. A three-dimensional perspective on exon binding by a group II self-splicing intron. *EMBO J.* 2000; 19:5007–5018. [PubMed: 10990464]
214. Toor N, Robart AR, Christianson J, Zimmerly S. Self-splicing of a group IIC intron: 5' exon recognition and alternative 5' splicing events implicate the stem-loop motif of a transcriptional terminator. *Nucl Acids Res.* 2006; 34:6461–6471. [PubMed: 17130159]
215. Michel F, Costa M, Doucet AJ, Ferat JL. Specialized lineages of bacterial group II introns. *Biochimie.* 2007; 89:542–553. [PubMed: 17391829]
216. Piskareva O, Ernst C, Higgins N, Schmatchenko V. The carboxy-terminal segment of the human LINE-1 ORF2 protein is involved in RNA binding. *FEBS Open Bio.* 2013; 3:433–437.
217. Rouda S, Skordalakes E. Structure of the RNA-binding domain of telomerase: implications for RNA recognition and binding. *Structure.* 2007; 15:1403–1412. [PubMed: 17997966]

218. Marcia M, Pyle AM. Principles of ion recognition in RNA: insights from the group II intron structures. *RNA*. 2014; 20:516–527. [PubMed: 24570483]
219. Strobel SA. Biochemistry: metal ghosts in the splicing machine. *Nature*. 2013; 503:201–202. [PubMed: 24196709]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Group II intron RNA splicing mechanism and structure

A. Splicing and reverse splicing. Step 1. The 2'-OH of the branch-point adenosine acts as nucleophile to attack the 5' splice site. Step 2. The 3'-OH of the upstream exon is the nucleophile that attacks the 3' splice site to generate ligated exons and an excised intron lariat. Both reactions are reversible. B. Group II intron secondary structure. The *L. lactis* L1.LtrB group IIA intron is shown, with the six domains DI-DVI. Exons are represented by thicker lines with the EBS-IBS pairings shown by dashed black lines. The IEP ORF is looped out of DIV (not drawn to scale). C. Group II intron crystal structure. The

representation is of DI-DV of the *O. iheyensis* group IIC intron (PDB:4E8K) bound to ligated exons, before the spliced exon reopening reaction, provided by Marco Marcia and Anna Pyle (13). Colors are coded to the domain labels in panel B, although these are different introns that belong to different structural subgroups. The 5' exon is black, and the 3' exon is dark blue. D. Base-pairing interactions of group IIA, IIB and IIC introns with flanking exons. Group IIA and IIB recognize 5' exons by similar IBS1-EBS1 and IBS2-EBS2 interactions, but use different interactions to recognize 3' exons (δ - δ' in IIA introns and EBS3/IBS3 in IIB introns (213). Group IIC ribozymes are only ~400 nt long, considerably smaller than IIA and IIB introns, and they are located downstream of inverted repeats, such as transcription terminators, which contribute to exon recognition along with short EBS1/IBS1 and EBS3/IBS3 interactions similar to those of IIB introns (214, 215). Panel D is adapted from reference (4), with permission of the publisher (copyright Cold Spring Harbor Laboratory Press).

Figure 2. Group II intron and related RTs

A. Schematics of RTs. Two group II intron RTs, *L. lactis* L1.LtrB (denoted LtrA protein; GenBank: AAB06503) and *Sinorhizobium meliloti* RmInt (NCBI Reference Sequence: NP_438012) are compared with two non-LTR-retrotransposon RTs, *Bombyx mori* R2Bm (GenBank: AAB59214) and human LINE-1 (UniProtKB/Swiss-Prot: O00370) RTs; yeast telomerase RT (GenBank: AAB64520); and retrovirus HIV-1 RT (PDB:2HMI). Conserved sequence blocks in the RT domain are numbered, and the sequence motif containing two of the conserved aspartic acid residues at the RT active site is shown below for each protein. B.

Three-dimensional model of the L1.LtrB RT. Regions identified by unigenic evolution analysis as being required for binding the high-affinity binding site DIVa and catalytic core regions of group II intron RNAs are highlighted in red and dark blue in the left and right panels, respectively, with pink in the left panel indicating a region that may enhance DIVa binding by stabilizing the structure of neighboring regions (46). The model was constructed by threading the amino acid sequence of the L1.LtrB RT onto a HIV-1 RT crystal structure, with one subunit (denoted α ; gray) modeled based on the catalytic p66 subunit of HIV-1 RT and the other subunit (denoted β ; cyan) modeled based on the p51 subunit of HIV-1 RT (17). The N-terminal 36 amino acid residues of the L1.LtrB RT could not be modeled based on the HIV-1 RT and are represented as spheres. Abbreviations: APE, apurinic endonuclease domain; CTS, conserved carboxy-terminal segment found to bind RNA non-specifically in LINE-1 RTs (216); Cys, cysteine-rich sequence conserved in LINE-1 RTs; DB, DNA-binding domain in R2Bm RT; REL, restriction endonuclease-like domain; TEN, telomerase N-terminal domain; TRBD, telomerase RNA-binding domain including motifs CP and T, which contact telomerase RNA (139, 217).

Figure 3. Representative retrohoming and retrotransposition pathways

A and B. Retrohoming into cognate sites is represented by the two pathways on the left. C and D. Retrotransposition to ectopic sites is represented by the two pathways on the right. For all panels, the intron is red (DNA solid lines, RNA hatched lines); the intron RNP is represented by a grey rectangle with a red intron lariat and an IEP with RT and maturase (X) domains and either containing or lacking an En domain. Exons of the donor (encircled D) are white, and those of the recipient (encircled R) are either white (retrohoming) or grey (retrotransposition). Each pathway ends with a product (encircled P). Green exon fragment

represents primer for reverse transcription. The retrohoming pathways (A and B) differ by the presence of the En domain, and whether dsDNA (A) or ssDNA, such as at a replication fork (B), is the target. Black dot in the recipient strand represents the intron-insertion site. In pathway A, the IEP contains an En domain and after reverse splicing into the top strand, cleavage of the bottom strand occurs 9 or 10 nt downstream (step 1'), as for the Ll.LtrB and yeast *aI2* introns, respectively (54, 101). In pathway B, the IEP lacks an En domain and integrates into DNA at a replication fork (step 1) as demonstrated for RmInt1 intron (62) and En-deficient mutants of the Ll.LtrB intron (78). Reverse splicing into the leading strand is not shown (see (78)). In both pathways, cDNA synthesis proceeds with the intron as template, using the 3'-OH of either the cleaved bottom strand (A) or an Okazaki fragment (B) to prime reverse transcription (step 2). Intron degradation and second-strand cDNA synthesis (step 3) is followed by DNA repair to generate the retrohoming products for both pathways (step 4). Host factors that participate in the process are indicated on pathway A, as established for the Ll.LtrB intron in *E. coli*, with those that silence the pathway shown in red, and those that promote retrohoming indicated in green (Table 1) (72, 73). EPP, error-prone polymerase. Retrotransposition (C and D) occurs when the intron integrates into ectopic sites with reduced specificity. This occurs for the Ll.LtrB intron into the lagging strand template for DNA synthesis, as pathway B, where Okazaki fragments prime cDNA synthesis (C) (58). Stimulatory and repressive host factors are again represented in green and red, respectively (Table 1) (45, 74, 97). Alternatively primers can be provided by nicks introduced into dsDNA by relaxase, the product of the *ltrB* gene that hosts the intron (pathway D) (93). Steps 1–4 in pathway D are as for retrohoming.

Figure 4. DNA target site recognition by group IIA, IIB and IIC introns

Target site interactions are shown for retrohomology (100, 101, 103) and retrotransposition (59, 92, 93) of the LI.LtrB group IIA intron, and for retrohomology of group IIB introns EcI5 (104) and RmInt1 (77), and group IIC intron *B.h.*I1-B (107). Intron RNA regions involved in EBS1-IBS1, EBS2-IBS2, δ - δ' or EBS3-IBS3 base-pairing interactions with the DNA target site are shown in red. A representative ectopic site is shown for the LI.LtrB retrotransposition pathway. Base-pairs in the 5' and 3' exons that are recognized by the IEP are highlighted in mauve and blue, respectively. CS, bottom-strand cleavage site; IS, intron-

insertion site; RF, replication fork; RH, retrohoming; RTP, retrotransposition; SL, stem-loop.
CS for Ec15 is not known.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 5. Similar RNA active sites of group II introns and the spliceosome

A. Schematic of group II intron (left) and spliceosome (right). Group II intron elements are colored as in Figure 1B and C, with corresponding spliceosomal elements in the same color (D1 and U5 blue; DV and U6 green; DVI and U2, purple). Conserved residues at the splice sites are in red, and the branch-point adenosine (A) is circled. Elements are not drawn to scale. B. Crystal structure of *O. iheyensis* group II intron in the precatalytic state (from PDB file 4FFAQ). The 5' exon (black) is shown before 5' splice site hydrolysis. Color-coding is as in Figure 1B & C and Figure 5A, with the two catalytic Mg^{2+} shown as yellow spheres

bound to DV (green) (13, 218). The putative water nucleophile is a cyan sphere. Images in panel B and D provided by Marco Marcia and Anna Pyle. C. Secondary structures and Mg^{2+} interactions in *O. iheyensis* intron DV and spliceosomal U6 snRNA. DV (left, shown also in panel B) corresponds to the internal stem loop (ISL) of U6 (right). M1 and M2 are catalytic Mg^{2+} ions coordinated by phosphate oxygens of the nucleotides shown in red. The circled A in the intron is the adenosine nucleophile with the 2'-OH corresponding to the water molecule in B, which attacks the 5' splice site (arrow). The introns are depicted in red hatch marks, with a small segment in green representing DV of the group II intron detailed to the right. U2 3' to the U2-U6 pairing extends to interact with the branch-point region of the intron (arrow to circled A). The three boxed nucleotides in each case comprise the conserved catalytic triad (163, 178, 183, 219). Dotted brown lines join residues involved in base triples, which are formed by two pairings between the catalytic triad and either J2/3 or the 5'-ACAGAGA-3' box and the third pairing to the bulge in each structure (163, 180). D. Base triples in DV of the *O. iheyensis* intron. The J2/3 nucleotides (orange ribbon) form a triple helix with the major groove at the base of DV (from (10)). Two metal ions (yellow spheres) are bound near the twisted bulge loop. E. Example of base triples in DV and U2-U6. The color-coded base triples are shown for the lower-most base-pair in the catalytic triad with a nucleotide in J2/3 in the group II intron and in the 5'-ACAGAGA-3' box in U6 of the spliceosome, as diagramed in panel C.

Figure 6. Comparison of Prp8 to group II intron and related RTs

Schematic comparing *S. cerevisiae* Prp8 (PDB:4I43), with the Ll.LtrB group II intron RT (LtrA protein; GenBank:AAB06503), *Arabidopsis thaliana* nMat1 and nMat2 proteins (NCBI Reference Sequence: NP_174294 and NP_177575, respectively), non-LTR-retrotransposon R2Bm RT (GenBank:AAB59214) and *Neurospora crassa* RVT RT (GenBank:CAE76174). The plant nMat1 and nMat2 proteins are previous examples of group II intron RTs that were subsumed into nuclear genomes and retained RNA splicing function, with the nMat1 proteins acquiring a novel conserved domain (green) in place of the En domain. The Prp8 configuration of the thumb, long linker, and REL domain is similar to that in the R2Bm RT. Although group II intron RTs and the R2Bm RT lack an RNase H domain, RNase H domains are known to be acquired sporadically in different non-LTR-retrotransposon lineages (135, 191). RVT RTs, another potential candidate for an ancestor of Prp8 (181), lack En and integrase domains and are further distinguished from group II intron and non-LTR-retrotransposon RTs by a large acidic insertion within RT-2a and conserved N- and C-terminal domains that are not found in other proteins (212). Conserved sequence blocks in the RT domain are numbered, and the sequence motif containing two of the conserved aspartic acid residues at the RT active site is shown below for each protein. The locations of conserved RT sequence blocks in Prp8 are from structure-based sequence alignments by Georg Mohr. Abbreviations: DB, DNA binding domain; REL, restriction endonuclease-like domain.

Table 1Cellular factors that affect group II intron retromobility^a

Factor ^b	E ^c	Identified Function	Putative effect on group II intron	Ref.
cAMP	S*	Global small-molecular regulator	Promotes retromobility	(97)
DnaB	S	Helicase	Primosome assembly	(73)
DnaC	S	Loading factor for DnaB	Primosome assembly	(73)
DnaG	S	Primase	Second-strand DNA synthesis; possible role in initiation of TPRT	(73)
DnaT	S	Loading factor with DnaC-DnaB	Primosome assembly	(73)
Exo III	I	3'-5' exonuclease	Degrade nascent cDNA	(72)
H-NS	S*	Nucleoid component, transcription regulator	Promotes retromobility – global	(74)
Ligase	S	DNA ligase	Sealing of DNA nicks	(72, 73)
LtrB	S	Relaxase from conjugative plasmid	Introduces spurious nicks into DNA	(93)
MutD	S	3'-5' exonuclease subunit of Pol III (<i>dnaQ</i>)	Repair second-strand DNA synthesis	(72)
Pol I	S	5'-3' exonuclease; removal of RNA primer from Okazaki fragments	Remove intron RNA template	(72, 73)
Pol II	S	Repair polymerase (<i>polB</i>)	Repair polymerization across DNA-RNA junctions	(72)
Pol III	S	Replicative polymerase	Second-strand DNA synthesis	(72, 73)
Pol IV	S	Repair polymerase (<i>dinB</i>)	Repair polymerization	(72)
Pol V	S	Repair polymerase (<i>umuDC</i>)	Repair polymerization	(72)
poly(P)	S*	Global regulator	Alters IEP localization and intron integration bias	(96)
PriA	S	Replication restart	Recognizes branched DNA with short gap and initiates replisome loading	(73)
PriC	S	Replication restart	Recognizes branched DNA with long gap and initiates replisome loading	(73)
ppGpp	S*	Global small-molecule regulator	Can promote retromobility	(97)
RbfA	S	Ribosome 30S and 50S association	Protects intron from RNase E degradation	(45)
RecJ	S	5'-3' exonuclease	5'-3' resection of DNA	(72, 73)
RNase E	I	Ribonuclease; part of RNA degradosome	Reduces half-life of intron RNA	(45, 72, 76)
RNase H1	S	Ribonuclease; cleaves RNA strand in RNA/DNA hybrid	Removes intron RNA template	(72, 73)
RNase I	I	Ribonuclease	Reduces half-life of intron RNA	(72)
RNase LS	I	Ribonuclease	Reduces half-life of intron RNA	(73)
Ssb	S	Single-stranded binding protein	Stabilizes ssDNA and may interact with PriA to recruit DnaB	(73)
StpA	S*	Nucleoid component, RNA chaperone	Promotes retromobility – global	(74)

^aRevised and updated from Beauregard et al. (75).^bAll factors shown to function in *E. coli* except the LtrB relaxase, which was shown to function in *L. lactis*.^cE, Effect; S, stimulates retromobility; I, inhibits retromobility; S*, stimulates retromobility into the chromosome only.