Abstract. Group II introns are a class of retroelements capable of carrying out both self-splicing and retromobility reactions. In recent years, the number of known group II introns has increased dramatically, particularly in bacteria, and the new information is altering our understanding of these intriguing elements. Here we review the basic properties of group II introns, and summarize the differences between the organellar and bacterial introns with regard to structures, insertion patterns and inferred behaviors. We also discuss the evolution of group II introns, as they are the putative ancestors of spliceosomal introns and possibly non-LTR retroelements, and may have played an important role in the development of eukaryote genomes.

Group II introns are a unique type of genetic element with two remarkable properties. They are catalytic RNAs, or ribozymes, that can excise themselves from pre-mRNA without the aid of proteins. They are also retroelements that encode reverse transcriptases (RTs) and can insert themselves into new locations. Group II introns are present in mitochondria and chloroplasts of plants and fungi, where they are relatively abundant (Michel et al., 1989), and they are also found in about 25% of eubacterial genomes, in either full-length or truncated forms (Ferat and Michel, 1993; Dai and Zimmerly, 2002). Recently, they were found even in some archaeabacterial species (Dai and Zimmerly, 2003; Rest and Mindell, 2003; Toro, 2003). Due to the burgeoning information from the genome projects, our knowledge about the introns has increased significantly in the past few years, and this review emphasizes this new genomic and comparative perspective.

The typical group II intron consists of a conserved intron RNA structure and an RT ORF. The RNA structure is comprised of six domains, of which domain V is believed to be the catalytic core of the ribozyme (Fig. 1A) (Michel and Ferat, 1995; Qin and Pyle, 1998). Domain I is the largest domain and is also important for catalysis, while domain VI contains the bulged A that forms the branch site in the spliced product. RNA structures are divided into two major classes, IIA and IIB, with class IIC being a minor class of bacterial introns.

The ORF is encoded within the loop of domain IV (Fig. 1A) and has four functionally defined domains: RT (reverse transcriptase activity, with sub-domains 0–7), X (maturase activity associated with intron splicing), D (non-conserved DNA-binding domain) and En (endonuclease activity) (Zimmerly et al., 2001; Belfort et al., 2002; San Filippo and Lambowitz, 2002). The protein serves two functions for the intron. It assists intron splicing in vivo through an activity associated with the X domain (maturase activity), and it also allows the intron to act as a mobile element and invade intronless sites, a process which involves all four of the ORF domains, as well as the catalytic RNA activity.

Self-splicing reaction

Ribozyme activity was the first property assigned to group II introns (van der Veen et al., 1986; Peebles et al., 1987). Self-splicing is accomplished by two transesterifications (Fig. 2A). In the first step, the 5’ exon is defined through two base-pairing interactions, in which the IBS 1 and 2 pair to their corresponding exon binding sequences (EBS)
Fig. 1. Basic group II intron structure. (A) Simplified secondary structure of IIB introns, composed of six structural domains, with the ORF located in domain IV. The intron is shown by black lines, exons by thick gray lines, and the ORF by a dotted black line. Black dots indicate individual nucleotides involved in base pairing tertiary interactions. The most important tertiary interactions are IBS1,2,3/EBS1,2,3, $\gamma-\gamma'$ and $\delta-\delta'$, which are indicated by gray lines and arrows. For IIA introns, $\delta'$ is located at the first nucleotide of the 3’ exon rather than in domain I, and there is no IBS3/EBS3 pairing. See Toor et al. (2001) for detailed information about RNA structural subclasses. (B) Intron-encoded ORF domains. The ORF consists of an RT domain (outlined gray boxes; reverse transcriptase with sub-domains 0–7), domain X (solid box; maturase domain), domain D (hatched box; DNA binding domain, not conserved in sequence among different ORFs), and the En domain (double-hatched box; endonuclease domain). Exons are dark gray, and intron RNA structure light gray.

Fig. 2. Reactions performed by group II introns. (A) Self-splicing reaction. Self-splicing is catalyzed by the intron RNA structure through two transesterification reactions, yielding spliced exons and intron lariat. The bulged A branch site that attacks the 5’-splice site is shown in domain VI of Fig. 1A. (B) Basic mechanism of group II intron mobility. The intron lariat in an RNP particle (lariat+RT) reverse splices into the top strand of the DNA target site. The bottom strand is cleaved by the En domain, and the intron is reverse transcribed. Recombination and repair activities complete the insertion.

1 and 2 (Fig. 1A). The two interactions position the 5’ splice site in proximity to the bulged adenosine in DVI (Jacquier and Michel, 1987). The 2’-OH of the adenosine acts as a nucleophile, attacking the phosphodiester bond of the 5’ junction to break the bond and form a lariat intermediate with a 2’-5’ linkage (Fig. 2A).

In the second splicing step, the 3’ exon is defined by two, single-base-pair interactions, which vary for the RNA structural classes. For IIA introns, the base pairs are $\gamma-\gamma'$ and $\delta-\delta'$, while for IIB and IIC introns, they are $\gamma-\gamma'$ and IBS3-EBS3 (Fig. 1A) (Jacquier and Michel, 1990; Costa et al., 2000). The second transesterification is similar to the first, but with the now free 3’-OH of the 5’ exon attacking the phosphodiester bond of the 3’ junction, resulting in exon ligation and release of the intron lariat (Fig. 2A). Because the number of phosphate bonds broken and created during the reaction is equal, splicing is energetically neutral and reversible, and does not require energy sources such as ATP.

The self-splicing reaction in vitro requires relatively extreme reaction conditions of salt, magnesium and sometimes temperature (e.g., 500 mM KCl, 100 mM MgCl2, 45°C) (Jarrell et al., 1988), and it is known that protein factors are required for splicing of introns in vivo. For many introns, the most important splicing factor is the intron-encoded protein. Upon transcription, the protein is translated from unspliced mRNA, and binds to a high affinity binding site in domain IVA in un-
spliced intron (Fig. 1A) (Wank et al., 1999). Combined with contacts with other regions of the intron RNA, the protein induces structural changes that stimulate self-splicing (Matsuurra et al., 2001; Noah and Lambowitz, 2003), to yield spliced exons and a ribonucleoprotein (RNP) particle consisting of a lariat with the protein still attached. This RNP is the entry molecule to the mobility pathway.

**The basic intron mobility reaction**

Mobility of group II introns occurs through a simple yet specialized mechanism. In the predominant event, retrohoming, the introns invade double-stranded DNA targets with high site specificity (Moran et al., 1995). Non-site-specific mobility, or retrotransposition, also occurs (see below) (Mueller et al., 1993; Sellem et al., 1993; Cousineau et al., 2000; Martínez-Abarca and Toro, 2000), but the majority of characterized introns home to consistent and predictable locations.

Homing is initiated by recognition of the DNA target site by the RNP, with the homing site typically extending from ~25 bp upstream to ~10 bp downstream of the insertion site. In recognizing the correct site, the protein binds to the upstream target site (~25 to ~13 for the well-studied *Lactococcus lactis* intron LL.ltrB), which stimulates local unwinding of the DNA. The remaining upstream target site (~12 to ~1 for LL.ltrB) is recognized by base pairing to the intron RNA through the same IBS1/EBS1 and IBS2/EBS2 interactions that occur during splicing (Guo et al., 1997, 2000; Mohr et al., 2000; Singh and Lambowitz, 2001). Concomitantly, the 3' exon is defined by the IBS3-EBS3 (or δ–δ') base pair(s) (Guo et al., 1997; Jiménez-Zurdo et al., 2003), which also are analogous to the pairing interactions during splicing. The intron then reverse-splices into the top strand (Zimmerly et al., 1995b) in a reaction that is mechanistically the reverse of splicing, and therefore intrinsically RNA-catalyzed. The protein recognizes sequences in the 3' exon and uses the En domain to cleave the bottom strand of the DNA target (Fig. 2B) at position +9 for LL.ltrB, creating a free 3'-OH from which the RT can prime reverse transcription (Zimmerly et al., 1995a; Matsuura et al., 1997). Once primed, the RT converts the inserted intron into cDNA to produce the bottom strand of the inserted intron. Formation of the top (second) strand of intron DNA has not been experimentally demonstrated, but is believed to be through recombination/repair pathways (below) (Fig. 2B). Overall, this process is called target-primed reverse transcription (TPRT), and is similar to reactions performed by non-LTR retroelements. For more detailed mechanistic discussion see other reviews (Lambowitz et al., 1999; Belfort et al., 2002; Lambowitz and Zimmerly, 2004).

**Variations in the mobility reaction**

Because the final steps of intron insertion are carried out by host repair activities, which can vary among organisms, multiple pathways occur. In yeast mitochondria, three pathways have been identified based on coconversion data for flanking markers (Eskes et al., 2000). The pathways differ with regard to the role of double-strand break repair (DSBR) activities that complete the insertion event. In bacteria, group II introns have not been observed to coconvert exon markers, and mobility occurs in RecA strains, indicating independence from a general DSBR system during retrohoming (Mills et al., 1997; Cousineau et al., 1998; Martínez-Abarca and Toro, 2000).

Interestingly, many bacterial introns do not encode En domains (Fig. 3B), yet appear to be mobile. The best studied of these is the *Sinorhizobium* intron RmInt1 (Martínez-Abarca et al., 2000; Muñoz-Adelantado et al., 2003), which is efficiently mobile in vivo even though it does not have an En domain. A similar process was characterized for an LL.ltrB intron with a mutated En domain (Zhong and Lambowitz, 2003). The data suggested a mechanism in which the intron reverse splices into the double-stranded DNA target, and cDNA synthesis is primed by the 3' end of a DNA at the replication fork, rather than priming by cleaved target DNA.

A final variation of mobility is insertion into noncognate sites, known as retrotransposition. In general, retrotransposition is thought to occur by reverse splicing directly into DNA sites, similar to retrohoming (Yang et al., 1998; Dickson et al., 2001), although most retrotransposition sites lack 3' exon sequences required for second strand cleavage (Mueller et al., 1993; Sellem et al., 1993; Ichiyanagi et al., 2002). In bacteria, retrotransposition occurs independently of the En domain (Cousineau et al., 2000), suggesting that the mechanism may be similar to homing without an En domain, where a nascent DNA at a replication fork is proposed to prime reverse transcription of the inserted intron RNA (Ichiyanagi et al., 2002; Zhong and Lambowitz, 2003).

**Group II introns as tools in biotechnology**

Because of their novel properties and wide species distribution, group II introns may provide a powerful new strategy for site-specific DNA manipulation. The basic idea is based on: 1) their long target site recognition (~30 bp); 2) the ability to insert foreign sequence into domain IV for coinserterion with the intron; and 3) the ability to alter the DNA target site by manipulating the EBS sequences that pair with the DNA exons during intron insertion. The region of protein recognition (~25 to ~13, and ~4 to ~10 for LL.ltrB), only requires a loose consensus to the cognate site, while the RNA-recognized region (~12 to +3 for LL.ltrB) can in principle be mutated to any new pairing interaction (Mohr et al., 2000). Together, these requirements allow predictable manipulation of new target site specificities, and such modified group II introns are known as targetrons.

There are a number of variations for the uses of targetrons. Importantly, a method has been developed to generate an intron to insert site-specifically into any target, through randomization of the EBS sequences and selection (Guo et al., 2000; Zhong et al., 2003). The RT ORF can be expressed in *trans* (Guo et al., 2000; Frazier et al., 2003), and intron insertions can ‘knock out’ selected genes (Karberg et al., 2001). Site-specific mutations can be targeted into bacterial chromosomes using a specialized intron that stimulates homologous recomb
Fig. 3. Variety of group II introns. Intron variations found in nature are diagrammed. Exons are shown as dark gray boxes, intron RNAs as black lines and intron-encoded ORFs as light gray boxes. RNA secondary structure domains I–VI are denoted by stem loops. Panels A–J represent double stranded DNA structures, with boxes and stems positioned up or down to indicate the strand orientation. Panels K–M represent RNA secondary structures. (A) Typical group II intron with all motifs. (Examples: *Lactococcus lactis* Ll.ltrB, yeast mitochondrial introns aI1, aI2) (Belfort et al., 2002). (B) Intron lacking the En domain, common in bacteria. (Example: *Sinorhizobium meliloti* RmInt1) (Martínez-Abarca et al., 2000). (C) ORF-less intron. (Examples: yeast mitochondrial introns aI5, cob1I1) (Michel et al., 1989). (D) Bacterial intron truncation. This example intron is truncated by an IS5 insertion in the opposite orientation. (Example: *E. coli* I4) (Dai and Zimmerly, 2002). (E) ORF-encoding intron with degenerate ORF, likely immobile. (Example shown is *trnK* I1 (matK) in plant chloroplasts; different degenerations are seen for *nadH*I4 (matR) in higher plant mitochondria and for *psbC*14 in *Euglena* chloroplasts (Mohr et al., 1993; Zimmerly et al., 2001). (F) Free-standing ORF without an intron RNA structure. (Examples: *nMat-1a, nMat-1b, nMat-2a, nMat-2b* in plant nuclear genomes; *matK* in mt genome of *Epiphyllus virginianum*) (Mohr and Lambowitz, 2003). (G) Insertion after rho-independent terminators. (Examples: all bacterial class C introns) (Dai and Zimmerly, 2002). (H) Insertion in the reverse orientation. (Examples: *X. f. I1* in *Xylella fastidiosa* DNA methyltransferase gene; *Th. e. I12* in *Thermosynechococcus elongatus* glycosyltransferase gene) (Dai and Zimmerly, 2002; L. Dai, S. Zimmerly, unpublished). (I) Twintron in *Euglena* chloroplasts. Twintrons can consist of many combinations of group II and III introns. (Example: *psbI3* gene twintron) (Copertino et al., 1992). (J) Twintron in the Archaeabacteria. The example shown consists of four nested introns, each inserted into a conserved sequence in another intron’s ORF. Two of the intron copies are incomplete, so that sequential splicing cannot remove all of the introns, and there is no ORF flanking the introns in any case. (Example: *M. a. F1/M. a. F1/M. a. F1/M. a. 15-1* in *Methanosarcina acetivorans*) (Dai and Zimmerly, 2003). (K) Two-piece trans-splicing intron, which is common in plant mitochondria. The intron is encoded in two different genomic locations, and splicing in trans produces a functional mRNA from two separately encoded exons. (Examples: *nadI1, nadI2* in higher plants) (Bonen, 1993). (L) Three-piece trans-splicing intron. This is similar to two-piece trans-splicing, but the intron is encoded in three pieces. (Example: *Chlamydomonas psI1*) (Choquet et al., 1988; Goldschmidt-Clermont et al., 1991). (M) Group III intron. Only found in *Euglena chloroplasts*, these introns are short, degenerate forms consisting of portions of domain I and VI. (Examples: *psbK*) (Doetsch et al., 2001).
nation with introduced mutated DNA sequence (Karberg et al., 2001). Finally, targetrons have been shown to insert into plasmid targets in human cells, pointing to potential applications in functional genomics and gene therapy in higher organisms (Guo et al., 2000). Clearly, the development of group II introns as vectors has very broad potential, and further work will develop their uses and applications.

Diversity of group II introns in nature: bacterial versus organellar introns and unusual variations

Organellar introns

Since the discovery of group II introns roughly twenty years ago, the models for study have been organellar introns. In the past ten years, however, our knowledge of bacterial introns has grown immensely, and combined with the probability that the bacterial introns are the most ancient, bacterial group II introns will soon be the new prototype if they are not already. In this section, we compare group II introns in organelles and in bacteria, and describe the impressive range of variations in the introns known to date (Fig. 3).

In mitochondria and chloroplasts, group II introns are primarily splicing entities, as most of them do not encode mobility ORFs (Fig. 3C) (Michel et al., 1989; Michel and Ferat, 1995; Zimmerly et al., 2001). Organellar introns are typically located in conserved housekeeping genes such as cytochrome oxidase and rubisco subunits, with sometimes up to ten introns (both group I and II) in a given gene. Therefore, splicing must be efficient enough to maintain function of the host gene. The introns are considered silent phenotypically because there is no apparent difference in fitness between strains that contain or lack the introns. A minority of organellar introns have mobility properties, but there is a bias in their distribution. In lower eukaryotes, about half of group II introns encode RT ORFs and would be predicted to be mobile, while in higher plants, nearly all introns are ORF-less, and there is only one ORF-containing intron out of about twenty introns in each organelle (matK in chloroplasts, matR in mitochondria). This difference between lower eukaryotes and plants suggests a progressive ORF degeneration and loss during plant evolution (below).

Degeneration is common for organellar group II introns, both in RNA structures and ORF motifs. In many cases, degenerate introns have probably lost mobility competence, but they must retain splicing competence because they are located in conserved genes. Similarly, ORFs with degenerate RT motifs probably no longer promote mobility but may retain splicing (maturase) function. Many host-encoded genes are known to facilitate splicing of group II introns in organelles (Lehmann and Schmidt, 2003). These include the helicase-related gene MSSI16 (Séraphin et al., 1989) and the magnesium ion transport gene MRS2 (Gregan et al., 2001) in yeast mitochondria. Maize chloroplast splicing proteins include the novel gene crs1 and the tRNA peptidyl hydrolase-related gene crs2, with its associated factors caf1 and caf2 (Jenkins and Barkan, 2001; Till et al., 2001; Ostheimer et al., 2003). Because some of the proteins are related to proteins with known activities, it is thought that pre-existing organellar proteins were adapted to function in intron splicing (Lambowitz et al., 1999). These factors would be predicted to be especially critical in splicing of degenerate introns.

The simplest type of degeneration is deviation of the RNA structure, most obviously through mispairings in domains 5 or 6, major deviations within conserved motifs, or large insertions or localized deletions. Such deviations are common for higher plant organellar introns and occur to different extents (Michel et al., 1989; Toor et al., 2001; N. Toor, S. Zimmerly, unpublished). While it is not possible to know for certain that individual deviations compromise catalytic activity, this is suggested cumulatively by the fact that no higher plant intron has yet self-spliced in vitro (Michel and Ferat, 1995). In contrast, group II introns in lower eukaryotes tend to have more “standard” secondary structures that are easier to model, and many of them self-splice in vitro (Michel and Ferat, 1995; N. Toor, L. Dai, S. Zimmerly, unpublished).

ORF degeneration is also a common theme. Many organellar ORFs have modest defects in mobility-related domains, such as mutations in the YADD motif or truncations of the endonuclease, and these losses appear to have occurred independently many times because they are distributed spottily among distantly related introns (Zimmerly et al., 2001). More dramatic degenerations of ORFs are also found. For example, matK, the ORF encoded by tRNA^{I3} introns in plants, contains no sequence identity in the region corresponding to RT domains 0–4, but has detectable remnants of domains 5–7 and very strong domain X conservation (Fig. 3E) (Mohr et al., 1993). From this it is inferred that the protein retains its maturase activity associated with domain X, but has lost mobility functions. In higher plant chloroplasts, matR is truncated for RT domains 0–1, and has two large insertions (>150 aa each) within the RT domain, raising doubts as to its mobility functions (Zimmerly et al., 2001). Also in plants, several nuclear-encoded proteins, nMat-1a, nMat-1b, nMat-2a and nMat-2b, are closely related to mitochondrial maturase proteins but are not found within an intron (Mohr and Lambowitz, 2003). These proteins are presumably derived from mitochondrial maturases, and are imported into mitochondria to function in intron splicing and/or mobility (Fig. 3F). Finally, Euglena chloroplasts have two intron-encoded ORFs, mat1 and mat2, which are severely degenerate throughout the RT and X motifs, but which likely have a function in splicing since they are conserved among Euglena species and their host intron is located in a photosystem gene (Zhang et al., 1995; Doetsch et al., 1998).

Trans-splicing is a type of “genomic degeneration” in which an intron RNA is encoded in two or more pieces, which are encoded in different locations in the genome (Bonen, 1993). The intron pieces along with their flanking exons are transcribed separately, and the intron segments associate to form an RNA structure that approximates a standard cis-splicing intron. The resulting splicing event ligates two exons encoded in distant parts of the genome (Fig. 3K). There are many examples of trans-splicing introns in higher plants (e.g., nadI11, nadI13 nadI212, nadI512, nadI513) (Pereira de Souza et al., 1991; Wissinger et al., 1991). Trans-splicing introns are believed to be formed by genomic rearrangements that split an intron, a scen-
ario that is supported by the presence of cis-splicing introns in early branching plants that are homologous to the trans-splicing homologs in higher plants (Malek and Knoop, 1998). Trans-splicing in three parts has also been found, for example, in the green alga *Chlamydomonas reinhardtii*, in which the *psaA* I1 intron is split into three pieces, which together fold into a structure resembling a degenerate group II intron structure. For this intron and its downstream trans-splicing intron in the *psaA* gene, a dozen proteins have been implicated genetically as affecting the trans-splicing process (Fig. 3L) (Choquet et al., 1988; Goldschmidt-Clermont et al., 1991).

Another type of dramatic degeneration is found in *Euglena* chloroplasts (Copertino et al., 1992). The *Euglena gracilis* chloroplast genome contains about 150 introns in total (~ 39% of its sequence), with two general types, group II and group III. The group II introns resemble other group II introns but have structural irregularities, while the group III introns are drastically degenerate, and are minimal introns of roughly 100 bp, which retain only the 5’ and 3’ motifs of the introns (Fig. 3M). Because of the range of intron degeneration in *Euglena* and the small size of many introns, it is clear that the introns need additional help from proteins or other RNAs acting in trans in order to splice, but these factors have not yet been identified.

*Euglena* chloroplasts also contain twintrons, an unusual organization in which one intron is nested inside another (either copy can be group II or III) (Fig. 3I). Sequential splicing of the introns produces a functional mRNA (Copertino et al., 1992). An interesting question, given the prevalence of the degenerate introns in the genome, is whether the introns are mobile in the degenerate form in *Euglena*, or whether only “standard” introns are mobile, with the current introns in the genome being immobile remnants.

In summary, there are a multitude of examples of degenerate introns in organelles. In nearly all cases, splicing of degenerate introns must be retained, because the introns are in conserved genes, and this may be promoted by splicing factors contributed by the host cell. In cases of ORF degeneration, it is likely that splicing function is usually retained while mobility function is lost.

**Bacteria**

In bacteria, in contrast to organellar introns, group II introns appear to behave as retroelements more than introns. Bacterial group II introns are associated with mobile DNAs and are rarely found in housekeeping genes (Dai and Zimmerly, 2002). The introns are mainly found in IS elements, plasmids, or pathogenicity islands, and are sometimes inserted between genes. Together, this suggests that the introns are excluded from conserved genes, either because they interfere with gene function, or because there is a replicative advantage by residing on other mobile DNAs. Unlike organellar introns, they almost always encode RT ORFs and are either retroelements or truncated retroelements (Fig. 3A). In general, they are not degenerate in RNA structure, but instead conform to the IIA or IIB consensus, and have conserved sequences and structural variations specific to each phylogenetic subclass (Toor et al., 2001). One subclass (bacterial class C) cannot be classified as IIA or IIB, and is designated IIC (Ferat et al., 2003; Toro, 2003).

The main form of degeneration for bacterial group II introns is fragmentation, and there are roughly equal numbers of full-length and truncated introns in sequenced bacterial genomes (Dai and Zimmerly, 2002). Truncations can be caused in a number of ways, such as incomplete reverse transcription during TPRT, an attractive idea because it would cause 5’ truncations analogous to those found in non-LTR elements. However, 5’ truncations account for a minority of fragmented introns. Figure 3D shows an example of 3’ truncation that was caused by the insertion of an IS element in the opposite orientation (Dai and Zimmerly, 2002). Truncations in bacteria are fundamentally different from the degenerations in organellar introns because the truncations completely inactivate the intron, whereas organellar introns must remain splicing-competent since they are in conserved genes. It should be noted that there are some examples of RNA structural degeneration in bacterial introns such as mispairs and indels, but since the introns are not located in conserved genes and do not need to retain splicing function, it is likely that many of these are “dying” copies of retroelements (Dai and Zimmerly, 2002).

The most unusual types of bacterial introns are those of the ORF phylogenetic class “bacterial class C.” The RNA is classified IIC because it does not conform to either the IIA or IIB consensus, as it lacks EBS2, and has a shortened domain 5 helix, and other deviations (Toor et al., 2001; Toro, 2003). Interestingly, the introns are not located in ORFs, but instead are located directly after rho-independent terminator motifs (Fig. 3G) (Dai and Zimmerly, 2002). This insertion strategy ensures that the intron does not impair expression of a host gene, and that the intron is transcribed at a low level, by termination readthrough. Often identical bacterial class C introns insert into multiple targets having very little sequence identity, but still having a terminator motif (inverted repeat followed by T’s). This pattern appears to defy the principle of homing into defined sequences, and it is not yet clear how the introns recognize the terminator motifs in the DNA.

Although there is a bias for bacterial introns being located outside of conserved genes, some notable exceptions have been found recently. The genome of the cyanobacterium *Trichodesmium sp.* contains 21 group II introns, twelve of which are located in conserved genes, such as allophycocyanin-B, IMP dehydrogenase, a ribonucleotide reductase, a DNA polymerase and RNase H (L. Dai, S. Zimmerly, unpublished). The ribonucleotide reductase in fact contains many genic interruptions, including three group II introns and, curiously, four inteins (Liu et al., 2003). The sheer number of examples in *Trichodesmium* suggests that the introns do not adversely affect host gene function in this cyanobacterium. Other examples of introns in conserved genes tend to be exceptions that prove the rule. In *Xylella fastidiosa* and *Thermosynechococcus elongatus*, two introns have inserted into conserved genes in the reverse orientation (DNA methyltransferase and glycosyltransferase genes, respectively) (Fig. 3H) (Dai and Zimmerly, 2002; L. Dai, S. Zimmerly, unpublished). In both cases, the intron cannot be spliced out from the mRNA transcript, and intron insertion has in effect knocked out the gene, an event more characteristic of retroelements than introns. In *Azotobacter vinelandii*, an intron has inserted into the stop codon of the GroEL gene, but because
the stop codon is changed to a different stop codon, it is not clear that intron insertion affects the host gene expression (Adamidi et al., 2003; Ferat et al., 2003).

Another interesting subset of introns in bacteria is ORF-less introns, which are now known in archaeabacteria and in cyanobacteria (four examples out of 130 known full-length introns) (Dai et al., 2003; L. Dai, S. Zimmerly, unpublished). Interestingly, the ORF-less introns appear to be mobile despite their lack of an ORF, which is inferred from their presence in multiple polynorphic targets within a genome. Their targets are generally conserved for IBS1 and 2 pairings, but not other protein recognition regions, suggesting that the target site is determined by the reverse splicing step rather than protein recognition. Intriguingly, for each ORF-less intron, there is a closely related intron within the genome that encodes an ORF, suggesting that the ORF-less introns require an RT protein acting in trans, making them “satellites” of the ORF-containing introns.

Twintrons have also been found in bacterial genomes, but they are different from twintrons described in Euglena chloroplasts. In the cyanobacterium Trichodesmium there are six twintrons, one of which is in a conserved gene (RNase H) (L. Dai, S. Zimmerly, unpublished). However, the majority of twintrons are not located in ORFs, and sequential splicing would not produce a functional mRNA. The most bizarre organization is found in the archaeabacterium Methanosarcina acetivorans, which contains one twintron with four levels of nesting (Fig. 3J) (Dai and Zimmerly, 2003). Some of the intron copies are incomplete, such that all introns cannot be spliced out, and no possible ORF surrounds the intron cluster that might encode a functional product. Instead, the introns appear to be homing into conserved sequence motifs of the RTs of other introns, presumably a selfish strategy that finds an innocuous target, but at the expense of other group II intron retroelements.

Together, these properties suggest that group II introns behave as retroelements in bacteria. In most cases, it is of little consequence to the cell whether the introns splice, although splicing does serve a purpose to the intron in providing RNPs for further mobility reactions. Bacterial introns appear to have adapted a survival strategy typical of selfish DNAs, based on constant movement to new genomic locations, balanced with deletions and loss. In contrast, organellar group II introns are adapted primarily to splicing, with many host enzymes having been recruited to promote and control the splicing process. In many cases in organelles, particularly for ORF-less introns, the mobility properties have been discarded in exchange for being an indispensable step in gene expression.

**Group II intron evolution**

Evolution of group II introns is an important topic because the introns are believed to be the ancestors of spliceosomal introns. According to this idea, the introns began as self-splicing entities, and after invasion of the nucleus the intron RNAs fragmented into multiple pieces (the snRNAs), which together carry out the splicing reaction within the spliceosome (Sharp, 1991). The parallel between group II introns and spliceosomal introns is based on similarities in the splicing reactions and RNA structures (Padgett et al., 1994; Villa et al., 2002), and recently, a functional similarity was observed in that a group II intron domain 5 stem can be swapped for the analogous U2/U6atac pairing the U12-dependent spliceosome (Shukla and Padgett, 2002).

The expanding microcosm of group II introns in bacteria has brought new insights to the history of the introns, but many questions remain unanswered. The diversity of introns in bacteria strengthens the idea that the introns evolved in bacteria and subsequently spread to the organelles (Ferat and Michel, 1993; Dai and Zimmerly, 2002). However, an interesting twist comes from the emerging picture that bacterial introns are retroelement-like, because it suggests that spliceosomal introns are ultimately derived from retroelements, a scenario which might help account for their successful spread in eukaryotic genomes.

Phylogenetic analysis of the group II intron ORFs divides them into eight groups: the mitochondrial class, the chloroplast-like classes 1 and 2, and bacterial classes A–E (Zimmerly et al., 2001; Toro et al., 2002). Each class is represented by mixed host organisms (for example, the mitochondrial and chloroplast-like classes each contain bacterial introns), and a single species can contain introns in multiple classes (for example, E. coli introns are in four different phylogenetic classes, and one E. coli intron is closely related to an archaeabacterial intron). Therefore, as perhaps expected for a mobile element, there appears to have been much horizontal transfer of the introns, particularly among bacteria.

A detailed analysis of RNA secondary structures and comparison with the phylogenetic classes indicated coevolution between intron RNA and ORFs, with each ORF subclass being associated with a different intron RNA structural class, which cumulatively represent all previously characterized RNA structural classes plus new varieties (Fontaine et al., 1997; Toor et al., 2001). Based on this general trend, as well as other evidence (Toor et al., 2001), we proposed the “retroelement ancestor hypothesis” to describe an overall evolutionary history for group II introns (Fig. 4). This hypothesis predicts that the ancestor of all known group II introns was a retroelement in bacteria having an RNA structure that was not of the “standard” IIA or IIB type as originally defined for organellar introns, and that the varieties of ribozyme structures differentiated as components of these retroelements. Some of the intron classes (mitochondrial and chloroplast-like classes) migrated to organelles, where a large percentage of introns lost their ORFs and in plants, nearly all introns lost ORFs and degenerated in RNA structure, becoming dependent on protein splicing factors supplied by the host organism. The concept of ORF loss is validated by examples of “ORF-less” introns in lower plant mitochondria that contain remnants of RT ORFs (Toor et al., 2001).

Group II introns are also evolutionarily related to non-LTR elements. RTs of group II introns and non-LTR elements are related phylogenetically, and structurally, they share a conserved N-terminal extension (domain 0) and domain 2a that is absent from retroviral and other RTs (Malik et al., 1999). Mechanistically, the two types of elements are both mobile through TPRT mechanisms, in which cleaved target DNA is
used to prime cDNA synthesis (Luan et al., 1993; Zimmerly et al., 1995b). If group II introns were present in primitive eukaryotes, as expected if they were precursors to spliceosomal introns, then it is also possible that they could have been precursors to non-LTR elements. Accordingly, one simple scenario is that mobile group II introns invaded the nucleus of a primitive eukaryote (either from an organellar or bacterial source) and that some introns lost RT ORFs and became spliceosomal introns, while others lost the intron RNA structure and became non-LTR elements. At this time one cannot say with certainty that group II introns were the ancestors, either direct or indirect, of retroelements or introns in higher eukaryotic genomes. Still, it is tantalizing to consider that group II intron invasion of a primitive eukaryotic nucleus was a pivotal occurrence in the development of eukaryotic genomes.

Fig. 4. The retroelement ancestor hypothesis for the evolution of group II introns. The model predicts that the ancestor of all known group II introns was a mobile bacterial intron with RNA structural features not conforming to canonical IIA or IIB structures. The different subclasses of ribosome structures are predicted to have differentiated as components of retroelements in bacteria, and two of these lineages (canonical IIA and IIB) migrated to organelles where many lost their ORFs and mobility properties. Spliceosomal introns could be derived from group II introns at any point in this history.

References


Jenkins BD, Barkan A: Recruitment of a peptidyl-
RNA hydrolase as a facilitator of group II intron
splicing in chloroplasts. EMBO J 20:872–879

Jimenez-Zurdo JI, Garcia-Rodriguez FM, Barrientos-
Durán A, Toro N: DNA target site requirements for
a nick at the chromosomal target site: a mechanism
for homing in vivo of a bacterial group II intron
encoding a protein lacking the DNA endonuclease

Karberg M, Guo H, Zhong J, Coon R, Perutka J, Lamb-
owitz AM: Group II introns as controllable gene
targeting vectors for genetic manipulation of bacte-

Lambowitz AM, Zimmerly S: Mobile group II introns.

Lambrecht PW, van den Engh G, Perler FB, Perlman PS:
Group I and group II ribozymes as RNPs; clues to
the past and guides to the future, in Gesteland
RF, Cech TR, Atkins JF (eds): The RNA World, pp

Lehmann K, Schmidt U: Group II introns: structural
and catalytic versatility of large natural ribozymes.

Liu QX, Yang J, Meng Q: Four inteins and three group
II introns encoded in a bacterial ribonucleotide
(2003).

Luoh AA, Viteri ME, Tschopp AE, Kleinman ME, Eckbich
TH: Reverse transcription of R2BM RNA is primed by
a nick in the chromosomal target site: a mechanism
(1993).

Malek O, Knoup V: Trans-splicing group II introns in
plant mitochondria: the complete set of car-
ranged homologs in ferns, fern allies, and a horn-

Malik HS, Burke WD, Eckbich TH: The age and evo-
lution of non-LTR retrotransposable elements.

Martinez-Abarca F, Toro N: RecA-independent ectop-
ical transposition in vivo of a bacterial group II

Martinez-Abarca F, Garcia-Rodriguez FM, Toro N:
Homинг of a bacterial group II intron with an
intron-encoded protein lacking a recognizable en-
donuclease domain. Mol Microbiol 35:1405–1412
(2000).

Matsura M, Saldanha R, Ma H, Wank H, Yang J, Mohr
G, Cavanagh DS, Dunn GM, Belfort M, Lambowitz
AM: A bacterial group II intron encoding
reverse transcriptase, maturase, and DNA endo-
nuclease activities: biochemical demonstration of
maturase activity and insertion of new genetic
information within the intron. Genes Dev 11:

Matsura M, Noah JW, Lambowitz AM: Mechanism of
maturase-promoted group II intron splicing.

Michel F, Fariel JL: Structure and activities of group II

Michel F, Umesono K, Ozeki H: Comparative and
functional anatomy of group II catalytic introns – a

Mills DA, Manias DA, McKay LL, Dunn GM: Hom-
ing of a group II intron from Lactococcus lactis
(1997).

Mohr G, Lambowitz AM: Putative proteins related to
group II intron reverse transcriptase/maturases are
encoded by nuclear genes in higher plants. Nucleic

Mohr G, Perlman PS, Lambowitz AM: Evolutionary
relationships among group II intron-encoded pro-
teins and identification of a conserved domain that
can be related to maturase function. Nucleic Acids

Mohr G, Smith D, Belfort M, Lambowitz AM: Rules
for DNA target-site recognition by a lacticoccal
group II intron enable retargeting of the intron to
specific DNA sequences. Genes Dev 14:559–573
(2000).

Moran JV, Zimmerly S, Eskes R, Kennell JC, Lambo-
owitz AM: The cellular environment of group II
introns of yeast mitochondrial DNA are novel site-
specific retroelements. Mol Cell Biol 15:2828–

Mueller MW, Allmaier M, Eskes R, Schweny RJ:
Homing of a bacterial group II intron with an
intron-encoded protein lacking the DNA endonuclease
activities: biochemical demonstration of reverse
splicing into DNA, but requires an unknown reverse transcrip-
(2003).

Noah JW, Lambowitz AM: Effects of maturase binding
and Mg2+ concentration of group II intron RNA
folding investigated by UV cross-linking. Biochemistry

Osthmeier GJ, Williams-Carrier R, Belcher S, Osborne
EG, Gierke J, Barkan A: Group II intron splicing fac-
tors derived by diversification of an ancient RNA-

Padgett RA, Poda M, Boulanger SC, Perlman PS: The
structure of a group II intron-encoded mat2 locus

Peebles CL, Benatan EJ, Jarrell KA, Perlman PS:
Group II intron self-splicing: development of alter-
native reaction conditions and identification of a
predicted intermediate. Cold Spring Harb Symp

Pereira de Souza A, Jubier MF, Delcher E, Lancelin D,
Etoumbal, R, Cech TR, Atkins JF (eds): The RNA World, pp

Pinto CR, Viteri ME, Tschopp AE, Kleinman ME, Eckbich
TH: The age and evolution of non-LTR retrotransposable
elements. Proc Natl Acad Sci USA 106:13053–13058
(2009).

van den Engh G, Perler FB, Perlman PS: The
mature group II intron RNA structures with their intron-encoded
maturase activity and insertion of new genetic
information within the intron. J Mol Biol 282:505–523

Zhang L, Jenkins KP, Stutz E, Hallick RB: The Euglena
gracilis intron-encoded mat2 locus is interrupted by
due to the past and guides to the future, in Gesteland
RF, Cech TR, Atkins JF (eds): The RNA World, pp

Zimmermer S, Guo H, Eskes R, Yang J, Perlman PS,
Lambowitz AM: Group II intron RNA structures with their intron-encoded

Zhou J, Garcia-Rodriguez FM, Toro N: Mobile group II
introns: additional mobile genetic elements in the environment.

Zhu J, Molina-Sánchez MD, Fernández-Lázaro M:
Identification and characterization of bacterial
DNA-primed reverse transcription activity of aII introns. Mol Cell 4:239–
250 (1999).

Wissinger B, Schuster W, Bremnicke A: Trans-splicing
in Oenothera mitochondria: nad1 mRNAs are
edited in exon and trans-splicing group II intron

Yang J, Pfeifer G, Perlman PS: Group II intron mobility in yeast mitochondria: target-
DNA primed reverse transcription activity of aII and reverse splicing into DNA transposition sites

Zhong J, Jenkins KP, Stutz E, Hallick RB: The Euglena
gracilis intron-encoded mat2 locus is interrupted by
due to the past and guides to the future, in Gesteland
RF, Cech TR, Atkins JF (eds): The RNA World, pp

Zhou J, Karberg M, Lambowitz AM: Targeted and
random bacterial gene disruption using a group II
intron (targetron) vector containing a retrotranspos-
tion-activated selectable marker. Nucleic Acids

Zimmermer S, Guo H, Eskes R, Yang J, Perlman PS,
Lambowitz AM: Group II intron RNA is a cata-
lytic component of a DNA endonuclease involved

Zimmermer S, Guo H, Perlman PS, Lambowitz AM:
Group II intron mobility occurs by target DNA-
primed reverse transcription. Cell 82:545–554

Zimmermer S, Hausner G, Wu X: Phylogenetic relation-
ships among group II intron ORFs. Nucleic Acids