

Circularization pathway of a bacterial group II intron

Caroline Monat and Benoit Cousineau*

Department of Microbiology and Immunology, Microbiome and Disease Tolerance Centre (MDTC), McGill University, Montréal, Québec, Canada H3A 2B4

Received July 31, 2015; Revised November 25, 2015; Accepted November 26, 2015

ABSTRACT

Group II introns are large RNA enzymes that can excise as lariats, circles or in a linear form through branching, circularization or hydrolysis, respectively. Branching is by far the main and most studied splicing pathway while circularization was mostly overlooked. We previously showed that removal of the branch point A residue from Ll.LtrB, the group II intron from *Lactococcus lactis*, exclusively leads to circularization. However, the majority of the released intron circles harbored an additional C residue of unknown origin at the splice junction. Here, we exploited the Ll.LtrB- Δ A mutant to study the circularization pathway of bacterial group II introns *in vivo*. We demonstrated that the non-encoded C residue, present at the intron circle splice junction, corresponds to the first nt of exon 2. Intron circularization intermediates, harboring the first 2 or 3 nts of exon 2, were found to accumulate showing that branch point removal leads to 3' splice site misrecognition. Traces of properly ligated exons were also detected functionally confirming that a small proportion of Ll.LtrB- Δ A circularizes accurately. Overall, our data provide the first detailed molecular analysis of the group II intron circularization pathway and suggests that circularization is a conserved splicing pathway in bacteria.

INTRODUCTION

Group II introns are retromobile genetic elements found in bacteria and bacterial-derived organelles, such as plant and fungal mitochondria, and plant chloroplasts. However, no functional copies were discovered in the nuclear genome of eukaryotes, whereas only few examples are present in archaea (1,2). Bacterial group II introns have a wide phylogenetic distribution although only one quarter of all sequenced bacterial genomes harbor one to a few copies (2). They are mostly found in non-coding sequences and within other mobile genetic elements while organellar group II introns generally interrupt housekeeping genes (3).

On an evolutionary standpoint, group II introns are considered as important retromobile elements that significantly shaped the origin and evolution of contemporary eukaryotic genomes. They are the proposed ancestors of the telomerase enzyme, the spliceosome machinery as well as the highly abundant spliceosomal introns and non-LTR retroelements (3,4). Group II introns can also be disseminated within and between bacterial species by conjugation. They can invade the chromosome or resident plasmids of their new hosts following their conjugative transfer using either the retrohoming or retrotransposition pathway (5–8).

Group II introns are also large RNA enzymes that self-splice from their primary mRNA transcripts both *in vivo* and *in vitro*. Even though, they absolutely require the assistance of maturases to adopt their catalytically active tridimensional conformation *in vivo*, they can self-splice without protein assistance under high salt conditions *in vitro* (2). Group II introns can excise while concurrently ligating their flanking exons through three different splicing pathways: branching (Figure 1A), hydrolysis (Figure 1B) and circularization (Figure 1C). These splicing pathways involve two consecutive transesterification reactions (Figure 1, steps 1 and 2) and release the intron as either branched lasso-like structures called lariats, in linear forms or as closed circles, respectively (9). Branching is definitively the most studied pathway and considered as the main group II intron splicing pathway. Hydrolysis and circularization are regarded as secondary splicing pathways and are not as well characterized as branching (9,10). Although a circularization pathway for group II introns was first proposed in 2001 (Figure 1C) (11) and that the first group II intron circles were detected by electron microscopy 35 years ago (12,13) this splicing pathway was not subjected to detailed molecular analyses. Following the initial description of the branching (14–16) and the hydrolysis pathways (17,18), circularization of group II introns was mostly overlooked. Nevertheless, since that time, group II intron RNA circles from various systems were described both *in vivo* (19–23) and *in vitro* (11,24). We recently demonstrated that a significant proportion of the model group II intron, Ll.LtrB, from the Gram-positive bacterium *Lactococcus lactis*, excises as circles *in vivo* (19). Interestingly, we detected intron circles for all the Ll.LtrB mutants studied and demonstrated that both branching and circularization take place concurrently *in vivo*. In addition,

*To whom correspondence should be addressed. Tel: +1 514 398 8929; Fax: +1 514 398 7052; Email: benoit.cousineau@mcgill.ca

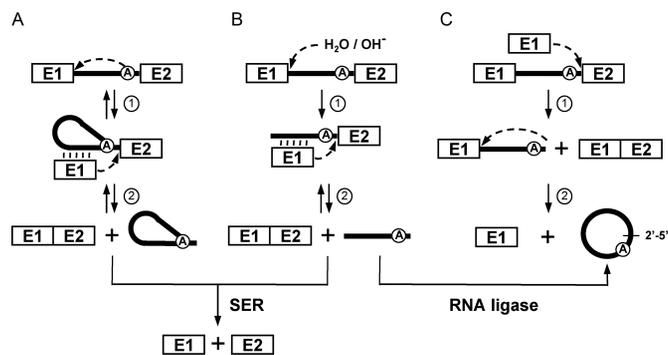


Figure 1. Group II intron splicing pathways. (A) Branching pathway. Following transcription of the interrupted gene, the 2'-OH residue of the branch-point nucleotide (A) initiates the first nucleophilic attack at the exon 1-intron junction (step 1). This transesterification reaction connects the 5' end of the intron to the branch point and releases exon 1 that remains associated to the intron through base pairing interactions (EBS-IBS interactions) (vertical lines). The liberated 3'-OH at the end of exon 1 then initiates a second nucleophilic attack at the intron-exon 2 junction (step 2), ligating the two exons and releasing the intron as a lariat. (B) Hydrolytic pathway. A hydroxyl ion or a water molecule initiates the first nucleophilic attack at the exon 1-intron junction (step 1). The second nucleophilic attack at the intron-exon 2 junction is initiated by the liberated 3'-OH at the end of exon 1 (step 2) which ligates the two exons and releases a linear intron. (C) Circularization pathway. The first nucleophilic attack takes place at the intron-exon 2 junction and is initiated by the 3'-OH of a free exon 1 (step 1) generating ligated exons and a circularization intermediate where the linear intron is still attached to exon 1. Next, the 2'-OH of the last intron residue is thought to initiate the second nucleophilic reaction at the exon 1-intron junction (step 2) resulting in intron circularization and the release of free exon 1. An additional potential source of free exon 1 is the spliced exon reopening (SER) reaction, where both excised lariats and linear introns can recognize and hydrolyze ligated exons at the splice junction. An alternative pathway for intron circle formation is the potential circularization of an excised linear intron by a host-encoded RNA ligase.

we found that the Ll.LtrB- Δ A mutant, lacking the branch point A residue, does not produce lariats and exclusively excises as circles. However, upon circularization, a majority of the Ll.LtrB- Δ A circles were found to harbor an additional non-encoded C residue of unknown origin at the splice junction (19).

Here, we took advantage of the Ll.LtrB- Δ A mutant, that exclusively splices as circles, to address at the molecular level the circularization pathway of bacterial group II introns *in vivo*. Using various mutants of Ll.LtrB- Δ A we demonstrated that the non-encoded C residue, present at the splice junction of released intron circles, corresponds to the first nt of exon 2. We also discovered that removal of the branch point A residue from Ll.LtrB leads to accumulation of precursor mRNAs and intron circularization intermediates in *L. lactis* total RNA extracts. The Ll.LtrB- Δ A circularization intermediates were also found to harbor short CA and CAU tails at the 3' end of the intron, corresponding to the first nts of exon 2, demonstrating that the 3' splice site is mostly misrecognized during the first transesterification reaction of the circularization pathway. On the other hand, a functional splicing assay confirmed the presence of accurately ligated exons consistent with the detection of small amounts of perfect intron circles not harboring an extra C at the splice junction. Taken together, our data indicate that Ll.LtrB- Δ A and Ll.LtrB-WT share the same circulariza-

tion pathway *in vivo*. Removal of the branch point A residue from Ll.LtrB prevents branching and significantly tilts the balance of the first transesterification reaction of the circularization pathway towards misrecognition of the 3' splice site, ultimately leading to the addition of the first nt of exon 2 (C) at the splice junction of released intron circles. Overall, this work provides the first detailed molecular analysis of the group II intron circularization pathway.

MATERIALS AND METHODS

Bacterial strains and plasmids

The *Escherichia coli* DH10 β strain was used for both cloning and plasmid amplification, and was grown in LB broth at 37°C with shaking. The *Lactococcus lactis* strains NZ9800 Δ ltrB::tet (NZ9800 Δ ltrB) (Tet^R) (25) and LMO231 (Fus^R) were grown in M17 broth supplemented with 0.5% glucose (GM17) at 30°C without shaking. The following antibiotic concentrations were used: chloramphenicol (Cam), 10 μ g/ml; spectinomycin (Spc), 300 μ g/ml; tetracycline (Tet), 3 μ g/ml; fusidic acid (Fus), 25 μ g/ml.

Some plasmids were previously described: pDL-P₂₃²-Ll.LtrB-WT (WT), pDL-P₂₃²-Ll.LtrB- Δ LtrA (Δ LtrA) (amino acids 40 to 572 in LtrA were replaced by RT residues (26)), pDL-P₂₃²-Ll.LtrB-LtrAMat⁻ (LtrAMat⁻) (amino acids SC463 were mutated to LA (26)), pLE-P₂₃²-LtrA (7), pDL-P₂₃²-Ll.LtrB- Δ A (Δ A) (19). The following kits were used for site directed mutagenesis: QuickChange Multi Site-Directed Mutagenesis, Stratagene (pDL-P₂₃²-Ll.LtrB-WT-C/U-E1, pDL-P₂₃²-Ll.LtrB- Δ A-C/U-E1, Ll.LtrB-LtrAMat⁻-C/U-E1) and Q5 Site-Directed Mutagenesis, New England Biolabs (pDL-P₂₃²-Ll.LtrB-WT-C/U-E2, pDL-P₂₃²-Ll.LtrB- Δ A-C/U-E2, Ll.LtrB-LtrAMat⁻-C/U-E2, pDL-P₂₃²-LtrB- Δ Ll.LtrB, pDL-P₂₃²-LtrB- Δ Ll.LtrB- Δ E2, pDL-P₂₃²-LtrB- Δ Ll.LtrB- Δ His) (primers in Supplementary Table S1).

Nucleic acid preparations, RT-PCR, poisoned primer extension, Northern blots and conjugation assays

Total RNA and nucleic acids were isolated from NZ9800 Δ ltrB cells harboring different intron variants as previously described (7). RT-PCR (7), conjugation (7), poisoned primer extensions (27) and Northern blots (27) were performed as previously described (primers in Supplementary Table S1).

Amplification of 3' and 5' intron ends

Total RNA extracts (1 μ g) of *L. lactis* expressing various Ll.LtrB constructs were analyzed. To identify intron 3' ends (Figure 5A), 5U of *E. coli* Poly(A) Polymerase (New England Biolabs) was used to extend a polyA tail (10 min at 37°C). Using an oligo dT and the SSIIRT (Invitrogen) a cDNA was synthesized at 42°C for 50 min. Following removal of the RNA strand of the RNA/DNA duplex by the RNase H (New England Biolabs) (20 min at 37°C) the single strand cDNA was amplified by PCR with the Phusion High-Fidelity DNA polymerase (New England Biolabs) using an intron specific primer (Supplementary Ta-

ble S1) and the oligo dT. For L1.LtrB- Δ A a second amplification of the 3' end was performed where a polyU instead of a polyA tail was added using 2U of Poly(U) Polymerase (New England Biolabs) (10 min at 37°C). In that case, an oligo dA was used for cDNA synthesis. The isolated PCR bands (Figure 5C–E) were sequenced toward the 5' end using intron specific primers (Supplementary Table S1). Amplification of free 5' intron ends (Supplementary Figure S1A) started by cDNA synthesis using the SSIIRT (Invitrogen) and an intron specific primer (Supplementary Table S1) at 42°C for 50 min. Following removal of the RNA strand of the RNA/DNA duplex by the RNase H (New England Biolabs) (20 min at 37°C) the single strand cDNA was extended with a polyA tail with 10U of Terminal Transferase (New England Biolabs) (30 min at 37°C). The single strand cDNA was amplified by PCR with the Phusion High-Fidelity DNA polymerase (New England Biolabs) using an intron specific primer (Supplementary Table S1) and an oligo dT. The series of G residues present between the intron 5' end and the polyA tail was most probably added by the intrinsic terminal transferase activity of the RT enzyme when it reached the 5' end of the RNA during cDNA synthesis (28). The isolated PCR bands (Supplementary Figure S1C–E) were sequenced toward the 3' end using intron specific primers (Supplementary Table S1).

RESULTS

The first nucleotide of exon 2 is incorporated at the splice junction of L1.LtrB- Δ A circles

We previously demonstrated that the great majority of the released RNA circles of L1.LtrB- Δ A and L1.LtrB-LtrAMat⁻ *in vivo* harbor an additional non-encoded C residue at the spliced junction (19). The origin of this additional residue was unclear since both nucleotides flanking L1.LtrB are C residues (Figure 2A, left). The second nucleophilic attack of the proposed circularization pathway (Figure 1C, step 2) could thus be initiated by the last nt of the intron, misrecognizing the 5' splice site and leading to the inclusion of the last nt of exon 1 (Figure 2A, scenario 1). Alternatively, the first nt of exon 2 may accurately identify the 5' splice site leading to its inclusion at the intron circle splice junction (Figure 2A, scenario 2).

To address the origin of the non-encoded C residue at the intron circle splice junction we generated the L1.LtrB- Δ A-C/U-E1 and L1.LtrB- Δ A-C/U-E2 mutants where the two C residues flanking the intron were independently replaced by Us (Figure 2A, right). The same two mutants were also generated for L1.LtrB-LtrAMat⁻ and L1.LtrB-WT. The C residues were specifically replaced by Us in order to minimize their effect on splice site recognition and splicing efficiency. The U residues allowed for G:U instead of G:C base pair interactions at both 5' (EBS1:IBS1) and 3' splice sites (δ : δ').

Splicing efficiency of the intron mutants was assessed by poisoned primer extension (27). A P³²-labeled primer complementary to exon 2 was extended from both the precursor and ligated exons in the presence of a high concentration of ddCTP (Figure 2B, top). Since the sequence of the two RNAs are different after the exon 2 junction the first G residue encountered is at a different distance from the

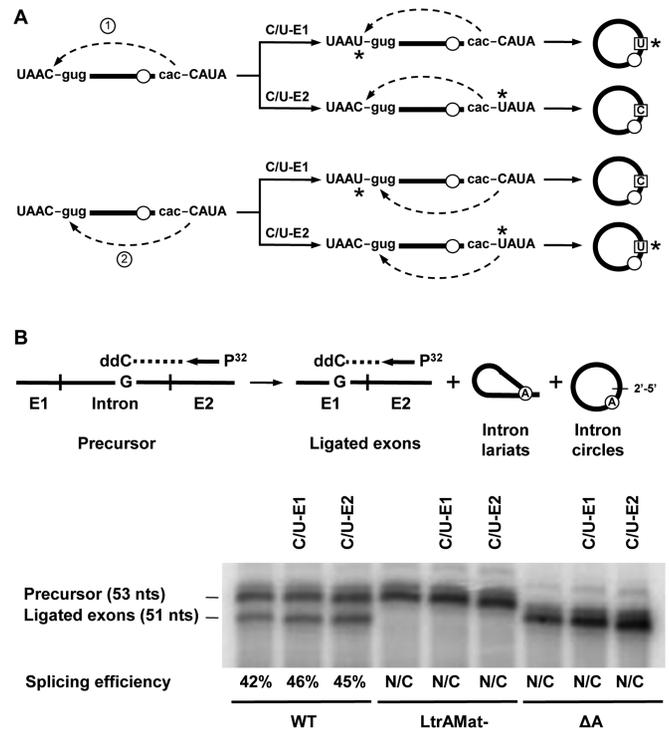


Figure 2. Splicing efficiency of various L1.LtrB constructs. (A) Schematic of the two potential transesterification reactions leading to the incorporation of an additional C residue at the L1.LtrB- Δ A circle splice junction are depicted on the left. The second transesterification reaction of the circularization pathway can be initiated by the 2'OH of either the last nt of the intron (scenario 1), leading to the incorporation of the last residue of exon 1, or the first nt of exon 2 (scenario 2) leading to its own incorporation. The potential circularization products generated for scenarios 1 and 2 are illustrated on the right. The C to U replacements in exon 1 (C/U-E1) and exon 2 (C/U-E2) are represented by asterisks in both the precursors and the released intron circles. Exon and intron sequences are depicted in upper and lower case, respectively. The absence of the branch point A residue is illustrated by an empty circle. (B) L1.LtrB splicing efficiency assessed by poisoned primer extension. This assay monitors splicing efficiency by comparing the relative abundance of precursor and ligated exons from total RNA extracts. A P³²-labeled primer (Supplementary Table S1) complementary to exon 2 was extended from both the precursor and the ligated exons in the presence of a high concentration of ddCTP (top). Since the sequence of the two RNAs are different after the exon 2 junction the first G residue encountered is at a different distance from the primer generating differently sized bands for the precursor (53 nts) and the ligated exons (51 nts). The precursor band for the three L1.LtrB- Δ A variants is 52 nts instead of 53 because the primer extension reaction extends past the intron branch point which is missing in these three introns. Splicing efficiency is calculated as the relative intensity of the ligated exons and precursor bands (ligated exons / precursor + ligated exons). N/C: not calculated.

primer generating differently sized bands for the precursor (53 nts) and ligated exons (51 nts) (Figure 2B, top). The splicing efficiency was found to be the same for the two C/U mutants compared to L1.LtrB-WT (Figure 2B, bottom). This shows that substituting either the last nt of exon 1 or the first nt of exon 2 from C to U does not affect L1.LtrB-WT splicing efficiency. For the L1.LtrB- Δ A and L1.LtrB-LtrAMat⁻ variants, that do not splice efficiently, only the precursor band was detected for all constructs showing that the C/U mutations on both sides of the intron do not influence their splicing phenotype.

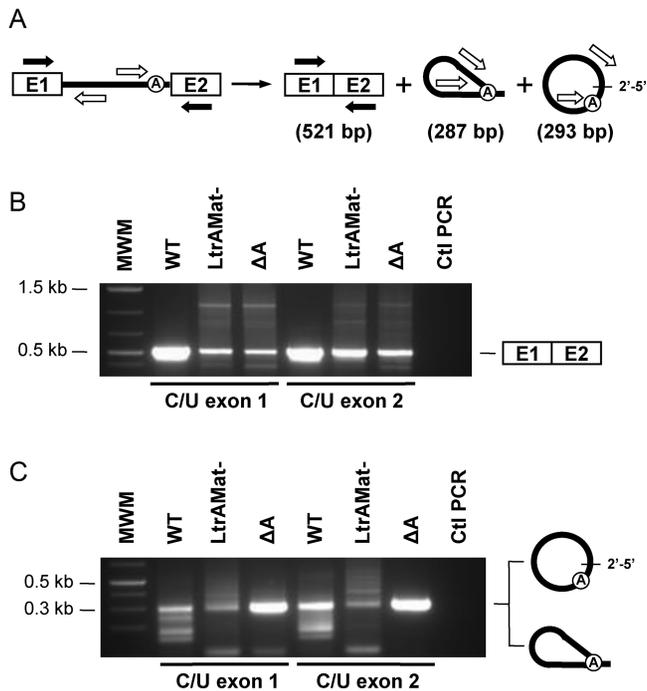


Figure 3. RT-PCR amplification of ligated exons and excised intron junctions. (A) Schematic of the LI.LtrB splicing pathway. Position of the primers (Supplementary Table S1) used to amplify ligated exons (black arrows, 521 bp) and the intron splice junction (open arrows, 287 or 293 bp) is depicted. RT-PCR amplifications (B) (ligated exons) (C) (intron splice junction) were performed on total RNA extracts from *L. lactis* (NZ9800Δ*ltrB*) harboring different LI.LtrB constructs expressed under the control of the P₂₃ constitutive promoter. RT-PCR amplicons of ligated exons (B, 521 bp) were excised and directly sequenced while the intron splice junction bands (C, 287 or 293 bp) were phosphorylated, cloned in pBS (SmaI) and sequenced.

Splicing accuracy of the C/U mutants was next assessed and compared to their wild-type intron counterparts by RT-PCR of ligated exons and released intron junctions (Figure 3) (19). Direct sequencing of the ligated exons bands (Figure 3B, 521 bp) confirmed accurate exon joining for the six C/U intron mutants. On the other hand, the RT-PCR bands of released intron junctions (Figure 3C, 287 or 293 bp) were isolated, cloned and the sequence of independent clones was obtained. Similarly to LI.LtrB-WT (19) we detected intron lariats, intron circles and alternatively circularized introns for LI.LtrB-WT-C/U-E1 and LI.LtrB-WT-C/U-E2. As previously described (19), the great majority of intron circles did not harbor additional nt at the splice junction regardless if the intron was flanked by a C or a U residue on its 5' or 3' end. This confirms that the C/U mutations flanking LI.LtrB-WT do not affect 5' and 3' splice site recognition. In contrast, both C/U mutations significantly affect the splicing accuracy of LI.LtrB-LtrAMat⁻ leading to almost exclusively alternatively spliced circles and preventing further analysis. Similarly to LI.LtrB-ΔA (19), all LI.LtrB-ΔA-C/U-E1 circles harbored an extra C at the splice junction (Figure 2A, scenario 2). However, the splice junction of the LI.LtrB-ΔA-C/U-E2 circles exclusively harbored an additional U residue instead of a C (Figure 2A, scenario 2). Interestingly, the limited number of LI.LtrB-WT-C/U-

E1 and LI.LtrB-WT-C/U-E2 circles, harboring respectively a non-encoded C or U residue, also supported scenario 2 (Figure 2A, scenario 2). The identification of an additional U at the splice junction of the LI.LtrB-ΔA-C/U-E2 and some LI.LtrB-WT-C/U-E2 circles demonstrates that the non-encoded nt is accurately read by the RT enzyme during RT-PCR and validate the identity of the C residue at the junction of LI.LtrB-ΔA-C/U-E1 and LI.LtrB-ΔA circles.

Overall, these results demonstrate that the non-encoded C residue present at the splice junction of LI.LtrB-ΔA circles correspond to the first nt of exon 2 (Figure 2A, scenario 2). The additional C residue is incorporated at the intron circle splice junction during the second step of the circularization pathway through accurate recognition of the 5' splice site by the 2'OH residue of the first nt of exon 2 (Figure 2A, scenario 2). Our data also suggest that the 3' splice site is mostly misrecognized during the first transesterification reaction of the circularization pathway (Figure 1C, step 1) leaving at least one extra nt at the intron 3' end.

Accumulation of LI.LtrB-ΔA circularization intermediates in *L. lactis*

We next compared splicing intermediates of LI.LtrB-ΔA with controls by Northern blots using a series of probes specific to either the intron (Intron), the intron lariat junction (Lariat jct), the intron circle junction (Circle WT jct), the LI.LtrB-ΔA circle junction (Circle ΔA+C jct) and the 5' splice site (Exon 1-Intron jct) (Figure 4). The four junction probes were designed to recognize the same number of nts on both sides of their respective junctions (Supplementary Table S1). The intron (Figure 4A) and 5' splice site (Figure 4E) probes revealed that the full-length precursor mRNA of LI.LtrB-ΔA (~4.2 kb) accumulates when compared to LI.LtrB-WT and LI.LtrB-ΔLtrA + LtrA that both splice very efficiently (27). A more intense LI.LtrB-ΔA band (~3.0 kb) was also detected by the same two probes (Figure 4A and E). According to its size and the fact that it is specifically revealed by the 5' splice site junction probe, this band most likely corresponds to the splicing intermediate of the circularization pathway consisting of exon 1 attached to full-length LI.LtrB-ΔA (Figure 1C). We were unable to detect LI.LtrB-ΔA circles from *L. lactis* total RNA extracts even if we used a probe specific to the circular junction of LI.LtrB-ΔA (Figure 4D). In addition, following migration through a denaturing polyacrylamide gel, no retarded bands corresponding to intron circles, in the size range of the control intron lariats, were revealed with the general intron probe (Figure 4F).

Overall these data show accumulation of full-length precursor mRNA and splicing intermediates of LI.LtrB-ΔA suggesting that both steps of the circularization pathway are not very efficient *in vivo*. They also suggest that the released intron RNA circles identified by RT-PCR (Figure 3C) (19) are present in low amounts not detectable by Northern blots.

Circularization intermediates of LI.LtrB-ΔA harbor short tails at the 3' end

To further characterize the extremities of the LI.LtrB-ΔA circularization intermediate detected by Northern blot we

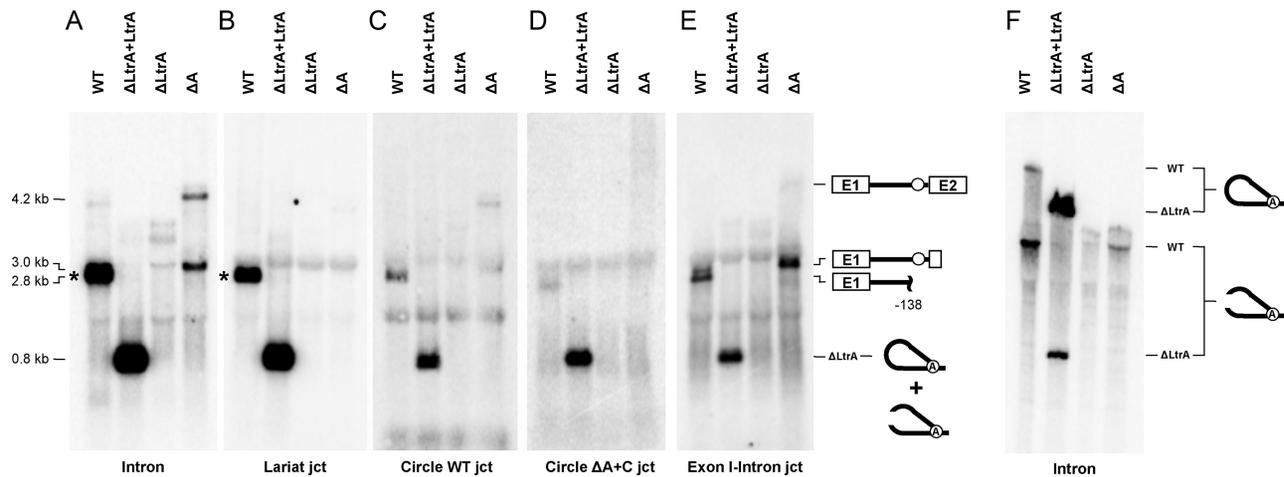


Figure 4. Detection of Ll.LtrB splicing intermediates *in vivo*. Northern blot hybridizations were performed with various P³²-labeled probes (Supplementary Table S1). *L. lactis* RNA extracts were run on denaturing agarose (A–E) or acrylamide gels (F). A general intron probe (A and F, Intron) was used to detect all splicing intermediates and intron forms while the other probes were designed to detect specific junctions recognizing the same number of nts on each side of the junction (B–E). Position of precursor mRNA, intron lariats, processed intron lariats and splicing intermediates is denoted. Position of the full-length released intron lariat (A and B) (2491 nts) is indicated by asterisks.

performed RT-PCR assays on *L. lactis* total RNA extracts. The intron 3' ends were analyzed by first extending the intron RNA with a polyA tail followed by the synthesis of a cDNA with an oligo dT. The RNA strand from the RNA/DNA duplex was removed by RNase H treatment followed by PCR amplification of the cDNA (Figure 5A). The specific band amplified for Ll.LtrB-WT (Figure 5B) corresponds to a processing site located within the *ltrA* gene at position -138 from the intron 3' end (Figure 5C). It is consistent with one of the bands detected by Northern blot (~2.8 kb) (Figure 4E, WT) and a processing site previously detected in the same area of Ll.LtrB (26). The same processing site was confirmed for Ll.LtrB-ΔA and Ll.LtrB-LtrAMat⁻ by sequencing the corresponding bands. The processing site is absent for Ll.LtrB-ΔLtrA and Ll.LtrB-ΔLtrA+LtrA where the LtrA coding gene is missing or expressed *in trans*, respectively. The higher and more intense band for Ll.LtrB-ΔA revealed the presence of at least an additional C at the 3' end of the intron (Figure 5D). In order to confirm the exact nature of the 3' end of Ll.LtrB-ΔA, the same experiment was repeated but extending a polyU instead of a polyA tail at the 3' end of the intron (Figure 5A). A similar PCR band was obtained and its sequence (Figure 5E), in conjunction with the previous sequence (Figure 5D), demonstrates that Ll.LtrB-ΔA harbors either two (CA) or three extra nts (CAU) at its 3' end corresponding to the first residues of exon 2. The intensity of the overlapping peaks at position 421 (Figure 5D) indicates that the Ll.LtrB-ΔA splicing intermediates harboring either a CA or CAU tail are in comparable amounts *in vivo*. Sequences of the corresponding bands for Ll.LtrB-ΔLtrA, Ll.LtrB-ΔLtrA+LtrA and Ll.LtrB-LtrAMat⁻ revealed the presence of similar tails at the 3' end of these intron variants (Figure 5B). Using a similar RT-PCR approach, free 5' intron end was only detected for Ll.LtrB-WT and Ll.LtrB-ΔLtrA+LtrA (Supplementary Figure S1). However, the amplified band for Ll.LtrB-ΔA was higher, more diffuse and significantly less intense (Supplementary

Figure S1B). Sequence of this band was shown to harbor the last 22 nts of exon 1 and to be heterogeneous beyond that point showing that precise 5' ends of Ll.LtrB-ΔA are not present *in vivo* (Supplementary Figure S1E). Thus, no trace of perfect linear intron was detected since Ll.LtrB-ΔA harbors either the first 2 or 3 nts of exon 2 at the 3' end and at least 22 nts of exon 1 at the 5' end.

These results demonstrate the presence, in similar amounts, of two major circularization intermediates consisting of exon 1 attached to full-length Ll.LtrB-ΔA with the two (CA) or three (CAU) first nts of exon 2 at their 3' end. They also corroborate the accumulation of these circularization intermediates *in vivo* (Figure 4) and confirm that the 3' splice site is mostly misrecognized by exon 1 during the first transesterification reaction of the Ll.LtrB-ΔA circularization pathway (Figure 1C).

Detection of Ll.LtrB-ΔA accurate circularization *in vivo* using a functional splicing assay

Analysis of the 3' end of Ll.LtrB-ΔA indicated that the 3' splice site is mostly misrecognized leading to the accumulation of circularization intermediates harboring either a CA or a CAU tail (Figure 5). Misrecognition of the 3' splice site by exon 1 during the first transesterification reaction should also lead to inaccurate exon ligation (Figure 1C, step 1). The absence of 2 nts (CA) at the exon splice junction changes the reading frame and creates a premature stop codon early in exon 2 producing a truncated relaxase (LtrB-ΔE2). The absence of 3 nts (CAU) at the splice junction leads to the deletion of a conserved catalytic Histidine residue (LtrB-ΔHis) (31). On the other hand, we detected by RT-PCR the presence of properly ligated exons (Figure 3) (19) as well as low levels of perfect intron circles, not harboring an extra C at the splice junction (Figure 3) (19) suggesting that both Ll.LtrB-ΔA splice sites are sometimes properly recognized during circularization.

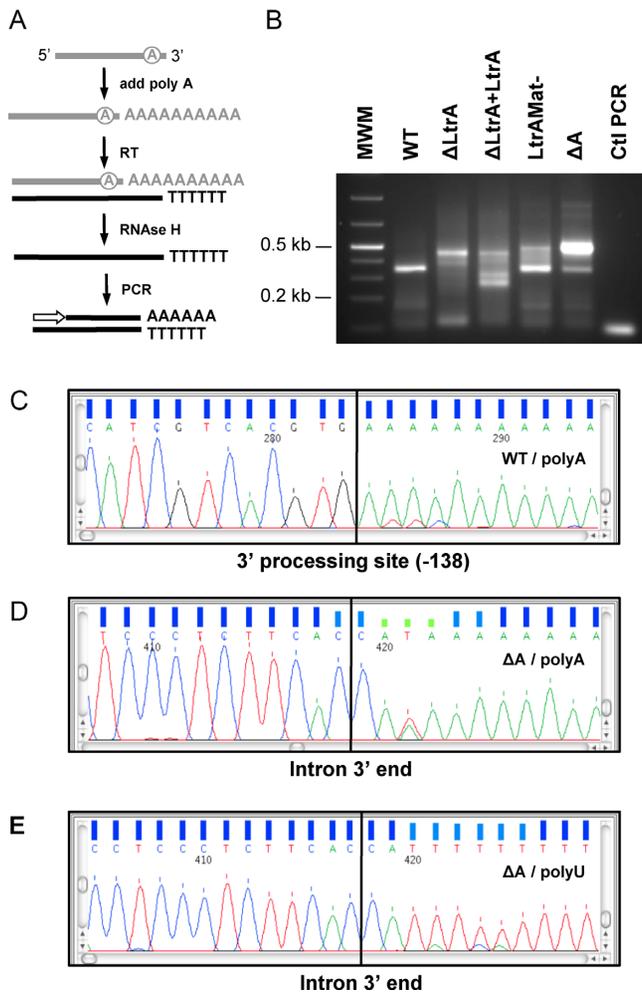
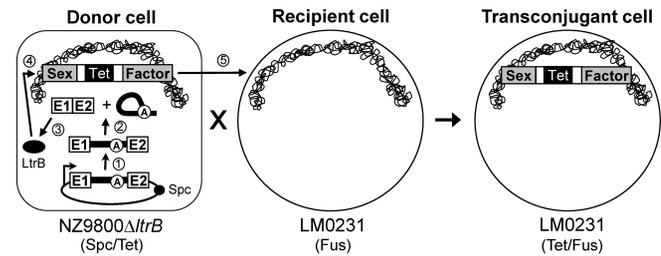


Figure 5. Identification of Ll.LtrB 3' ends *in vivo*. Free intron 3' ends were amplified by RT-PCR from *L. lactis* total RNA extracts. (A) Intron 3' ends were identified by first extending the intron RNA with a polyA tail followed by the synthesis of a cDNA with an oligo dT. The RNA strand was removed by an RNase H treatment and the single strand DNA amplified by PCR. (B) The PCR reactions were run on a 2% agarose gel and the chromatogram of some of the sequenced bands are shown (C–D). The same procedure was repeated for the Ll.LtrB- ΔA construct but extending a polyU instead of a polyA tail at the 3' end of the intron (E).

To functionally address the accurate circularization of Ll.LtrB- ΔA *in vivo* we used a very sensitive splicing/conjugation assay where Ll.LtrB splicing efficiency is monitored by the transfer rate of the chromosomal sex factor (SF) between strains of *L. lactis* (Figure 6, top) (7,29). Ll.LtrB interrupts the gene coding for a relaxase enzyme (LtrB) that initiates SF transfer by conjugation. The transfer rate of the SF between *L. lactis* strains was shown to be directly proportional to Ll.LtrB splicing efficiency from the relaxase transcript (7,30).

The level of SF conjugation efficiency supported by Ll.LtrB- ΔA was found to be similar to Ll.LtrB-LtrAMat⁻ (10^{-4}) three logs lower than Ll.LtrB-WT (10^{-1}) but significantly higher than background (10^{-9}) (Figure 6, bottom). In contrast, the mutant relaxase genes, not interrupted by the intron (LtrB- Δ Ll.LtrB- Δ E2, LtrB- Δ Ll.LtrB- Δ His), only promote SF conjugation efficiency in the 10^{-8}



Ll.LtrB constructs in pDL

pDL-P₂₃² (empty plasmid) ^a
 pDL-P₂₃²-Ll.LtrB-WT ^a
 pDL-P₂₃²-Ll.LtrB- Δ LtrA ^b
 pDL-P₂₃²-Ll.LtrB- Δ LtrA+LtrA
 pDL-P₂₃²-Ll.LtrB-LtrAMat ^b
 pDL-P₂₃²-Ll.LtrB- ΔA
 pDL-P₂₃²-LtrB- Δ Ll.LtrB
 pDL-P₂₃²-LtrB- Δ Ll.LtrB- Δ E2
 pDL-P₂₃²-LtrB- Δ Ll.LtrB- Δ His

SF conjugation efficiency

$(1.13 \pm 0.38) \times 10^{-9}$
 $(1.08 \pm 0.21) \times 10^{-1}$
 $(1.93 \pm 0.96) \times 10^{-8}$
 $(7.67 \pm 5.58) \times 10^{-2}$
 $(2.86 \pm 0.86) \times 10^{-4}$
 $(4.57 \pm 1.40) \times 10^{-4}$
 $(4.96 \pm 0.90) \times 10^{-2}$
 $(1.27 \pm 0.22) \times 10^{-8}$
 $(8.39 \pm 4.64) \times 10^{-8}$

^a Quiroga *et al.*, 2011; ^b Belhocine *et al.*, 2007

Figure 6. Ll.LtrB splicing/conjugation assay and SF conjugation rates. The splicing efficiency of Ll.LtrB is monitored by the conjugation rate of the SF between two *L. lactis* strains. The donor strain harbors the chromosomal SF with a defective relaxase gene (NZ9800 Δ ltrB::tet) while the recipient strain lacks the SF (LM0231). The Ll.LtrB intron, along with portions of its exons, was replaced in the chromosome of NZ9800 by a tetracycline resistance marker (tet), which prevents expression of the relaxase (LtrB). The LtrB deficient strain is complemented by providing the interrupted *ltrB* gene from a plasmid. Following transcription of the *ltrB* gene (step 1) Ll.LtrB splices and the flanking exons are ligated (step 2). If the exons are precisely ligated the mature mRNA will be properly translated leading to expression of the relaxase enzyme (LtrB, black oval) (step 3). LtrB will then recognize the SF origin of transfer (*oriT*) (step 4) and initiate SF transfer from a donor to a recipient cell by conjugation (step 5). Transfer efficiency of conjugative elements between *L. lactis* strains was previously shown to be directly proportional to Ll.LtrB splicing from the relaxase transcript (7,30). In variations of the assay, the pLE-P₂₃²-LtrA plasmid was cotransformed with pDL-P₂₃²-Ll.LtrB- Δ LtrA to provide LtrA in trans (pDL-P₂₃²-Ll.LtrB- Δ LtrA+LtrA) or the uninterrupted relaxase gene enzyme was expressed (pDL-P₂₃²-LtrB- Δ Ll.LtrB, pDL-P₂₃²-LtrB- Δ Ll.LtrB- Δ E2, pDL-P₂₃²-LtrB- Δ Ll.LtrB- Δ His).

range (Figure 6, bottom). This suggests that SF conjugation efficiency is supported by the presence of low levels of accurately ligated exons and not by the improperly ligated exons complementary to the two major circularization intermediates.

Taken together, our functional splicing assay confirms the presence of low levels of accurately ligated exons *in vivo* and suggests that accurate recognition of the 3' splice site by exon 1 leads to the appropriate recognition of the 5' splice site by the last nt of the intron and the precise excision of Ll.LtrB- ΔA circles not harboring additional non-encoded nts at the splice junction.

DISCUSSION

Even though a pathway for the excision of group II intron circles was initially proposed many years ago (11), this splicing mechanism was mostly ignored and not examined in details at the molecular level. Nevertheless, we recently demonstrated that, concurrently with branching, a signif-

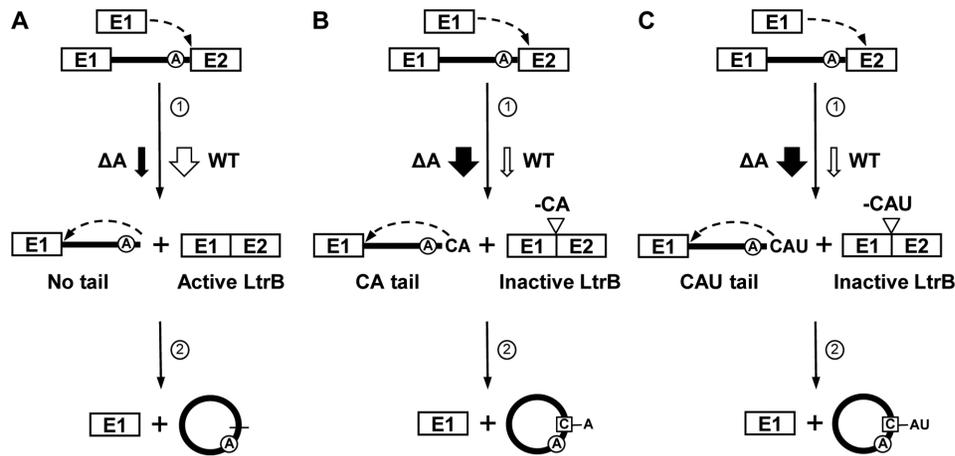


Figure 7. L1.LtrB circularization pathway. L1.LtrB-WT (WT) and L1.LtrB- Δ A (Δ A) use the same circularization pathway (A–C) *in vivo*. (A) The first nucleophilic attack at the intron-exon 2 junction is initiated by the 3'-OH of a free exon 1 (step 1). Accurate recognition of the 3' splice site properly ligates the flanking exons leading to the expression of an active relaxase (Active LtrB). The 2'-OH of the last intron residue initiates the second nucleophilic reaction at the exon 1-intron junction (step 2) resulting in precise intron circularization and the release of free exon 1 available to initiate another circularization reaction. (B and C) The 3'-OH of a free exon 1 misrecognizes the intron-exon 2 junction during the first nucleophilic attack (step 1) leaving tails of 2 (CA) (B) or 3 (CAU) (C) nts at the 3' end of the intron. Misrecognition of the 3' splice site also improperly ligates the flanking exons creating a premature stop codon just after the splice junction (-CA) (Inactive LtrB, LtrB- Δ E2) (B) or deleting an important catalytic amino acid from the LtrB relaxase enzyme (-CAU) (Inactive LtrB, LtrB- Δ His) (C). The 2'-OH of the first nt of exon 2 then accurately initiates the second nucleophilic reaction at the exon 1-intron junction (step 2) resulting in its incorporation at the intron circularization junction and the release of free exon 1 available to initiate another circularization reaction. L1.LtrB-WT circularization predominantly involves the release of perfect intron circles (A) while the great majority of L1.LtrB- Δ A circles harbor the first nt of exon 2 at their splice junction (B, C). However, low level of accurate (perfect intron circles, precisely ligated exons, SF conjugation) and inaccurate (first nt of exon 2 at circle splice junction) circularization can still be detected for L1.LtrB- Δ A and L1.LtrB-WT respectively. Removal of the branch point A residue significantly affects 3' splice site recognition tilting the balance from accurate (A) to inaccurate (B and C) 3' splice site recognition ultimately leading to the addition of the first nt of exon 2 (C) at the circle splice junction (cf. WT (open arrows) and Δ A (black arrows)).

icant proportion of the L1.LtrB group II intron from the Gram-positive bacterium *L. lactis* uses the circularization pathway *in vivo* (19). However, since the RT enzyme is not as processive moving through the 2'-5' bond at the branch point of lariats during RT-PCR, the relative abundance between intron lariats and circles cannot be precisely determined (19). In this study, we took advantage of the L1.LtrB- Δ A mutant that does not produce lariats and exclusively excises as circles to address experimentally the circularization pathway of bacterial group II introns.

Upon circularization, a majority of the L1.LtrB- Δ A circles were found to harbor an additional non-encoded C residue of unknown origin at the intron splice junction while some circular junctions were perfectly joined (19). RT-PCR assays using L1.LtrB- Δ A mutants where the flanking C residues were independently replaced by Us demonstrated that the non-encoded C residue present at the splice junction of circular introns originates exclusively from the first nt of exon 2 accurately recognizing the 5' splice site (Figure 2A, scenario 2). This result strongly suggested that the 3' splice site of L1.LtrB- Δ A is mostly misrecognized during the first transesterification reaction of the circularization pathway (Figure 1C) leaving at least the first nt of exon 2 attached to the 3' end of the intron. In accordance, we showed that the 3' end of the L1.LtrB- Δ A circularization intermediates harbor the first two (CA) or three (CAU) nts of exon 2 (Figure 5). These two main circularization intermediates, consisting of full length intron still attached to exon 1, were found to be in approximately the same amount in *L. lactis* total RNA extracts (Figure 5E) and to accumulate *in vivo* (Figure 4). As expected, free 3' in-

tron ends were not detected since we previously demonstrated that only a small proportion of L1.LtrB- Δ A circles do not harbor an extra C at the circularization junction (19). On the other hand, misrecognition of the 3' splice site suggested the generation of improperly ligated exons missing either 2 (CA) or 3 (CAU) nts at the splice junction and producing respectively a truncated (LtrB- Δ E2) or catalytically inactive (LtrB- Δ His) relaxase enzyme. Using a functional splicing assay we showed that the *ltrB* gene interrupted by L1.LtrB- Δ A supports SF conjugation efficiency significantly higher than background (10^{-4} vs 10^{-9}) (Figure 6). We concluded that the SF conjugation efficiency is supported by trace amounts of accurately ligated exons produced during accurate circularization of L1.LtrB- Δ A since both LtrB- Δ L1.LtrB- Δ E2 (10^{-8}) and LtrB- Δ L1.LtrB- Δ His (10^{-8}) mutated proteins, that should be generated through 3' splice site misrecognition, cannot support SF conjugation in the 10^{-4} range (Figure 6).

Our work suggests that the L1.LtrB- Δ A circularization pathway (Figure 7, black arrows) is mostly initiated by misrecognition of the 3' splice site by the 3'OH of the last nt of free exon 1 (Figure 7B, C, step 1). This first transesterification reaction concurrently leads to the production of two major circularization intermediates with short tails of two (CA) (Figure 7B) or three (CAU) (Figure 7C) nts at the 3' end of the intron and the inaccurate joining of flanking exons leading to the production of inactive LtrB relaxase enzymes (Figure 7B and C). However, small amounts of perfect intron circles and accurately ligated exons were detected by RT-PCR and conjugation showing that the 3' splice site is also accurately recognized albeit at a significant

lower level (Figure 7A). The second transesterification reaction, that accurately recognizes the 5' splice site, is mainly initiated by the 2'OH of the first nt of exon 2 leading to its incorporation at the intron circle splice junction (Figure 7B and C, step 2). When the intron does not harbor a tail at the 3' end (Figure 7A), the 2'OH of the last nt of the intron precisely attack the 5' splice site releasing perfect end-to-end intron circles (Figure 7A, step 2). Precise recognition of the 5' splice site releases more exon 1 that can initiate another round of intron circularization (Figure 7A–C). Accumulation of full-length precursor mRNAs (Figure 4) and circularization intermediates of Ll.LtrB- Δ A (Figures 4 and 5) suggests that both transesterification reactions are not very efficient *in vivo* when compared to the branching pathway (Figure 1A).

Taken together, our data also indicate that Ll.LtrB-WT uses the same circularization pathway as Ll.LtrB- Δ A *in vivo*. In accordance, we detected circularization intermediates with 3' tails similar to Ll.LtrB- Δ A for Ll.LtrB- Δ LtrA+LtrA, Ll.LtrB- Δ LtrA and Ll.LtrB-LtrAMat⁻ showing some level of 3' splice site misrecognition despite the presence of the branch point A residue (Figure 5B). We assume that low amounts of circularization intermediates with similar tails are also present for Ll.LtrB-WT but that the occurrence of a prominent processing site at position -138 prevented their detection (Figure 5B and C). The detection of some Ll.LtrB-WT circles also harboring the first nt of exon 2 at the splice junction is another indication that Ll.LtrB-WT and Ll.LtrB- Δ A share the same circularization pathway. Even though both Ll.LtrB variants use the same overall circularization pathway the balance between accurate and misrecognition of the 3' splice site during the first transesterification reaction was found to be considerably different. While the 3' splice site of Ll.LtrB-WT is most of time accurately recognized, leading to precise ligation of its flanking exons, and releasing almost exclusively perfect intron circles, the 3' splice site of Ll.LtrB- Δ A is generally misrecognized, leading to improperly ligated exons and releasing almost exclusively intron circles harboring the first nt of exon 2 at the splice junction (Figure 7, cf. black arrows (Ll.LtrB- Δ A) and open arrows (Ll.LtrB-WT)). Removal of the branch point A residue from Ll.LtrB thus not only prevents branching but also significantly tilts the balance of the first transesterification reaction of the circularization pathway towards misrecognition of the 3' splice site.

Interestingly, variants of the RmInt1 intron from the bacterium *Sinorhizobium meliloti* were also found to harbor an additional non-encoded C residue at the splice junction of released intron circles (21,22). Similarly to Ll.LtrB, RmInt1 harbors a C residue as the first nt of exon 2 which was proposed to be at the origin of the additional nt at the circle splice junction following 3' splice site misrecognition (32). This suggests that these two bacterial introns, even though not from the same class, share the same circularization pathway *in vivo*.

In summary, this work represents a thorough molecular characterization of the previously proposed group II intron circularization pathway (11). Our data also suggest that circularization is a conserved secondary splicing pathway for bacterial group II introns. Given the variety of known and predicted functions associated with circular RNAs from

various biological systems (33) one could envisage that group II intron RNA circles also serve a specific function in bacterial cells.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Félix LaRoche-Johnston and Deeva Uthayakumar for providing comments on the manuscript.

FUNDING

Discovery Grant from Natural Sciences and Engineering Research Council of Canada [227826 to B.C.]; William Dawson Scholar Award from McGill University [100705 to B.C.]. The open access publication charge for this paper has been waived by Oxford University Press - NAR.

Conflict of interest statement. None declared.

REFERENCES

1. Ferat, J.L. and Michel, F. (1993) Group II self-splicing introns in bacteria. *Nature*, **364**, 358–361.
2. Lambowitz, A.M. and Zimmerly, S. (2011) Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.*, **3**, a003616.
3. Zimmerly, S. and Semper, C. (2015) Evolution of group II introns. *Mob. DNA*, **6**, 7.
4. Lambowitz, A.M. and Belfort, M. (2015) Mobile bacterial group II introns at the crux of eukaryotic evolution. *Microbiol. Spectr.*, **3**, 1.
5. Belhocine, K., Plante, I. and Cousineau, B. (2004) Conjugation mediates transfer of the Ll.LtrB group II intron between different bacterial species. *Mol. Microbiol.*, **51**, 1459–1469.
6. Belhocine, K., Yam, K.K. and Cousineau, B. (2005) Conjugative transfer of the *Lactococcus lactis* chromosomal sex factor promotes dissemination of the Ll.LtrB group II intron. *J. Bacteriol.*, **187**, 930–939.
7. Belhocine, K., Mak, A.B. and Cousineau, B. (2007) *Trans*-splicing of the Ll.LtrB group II intron in *Lactococcus lactis*. *Nucleic Acids Res.*, **35**, 2257–2268.
8. Nisa-Martínez, R., Jiménez-Zurdo, J.I., Martínez-Abarca, F., Muñoz-Adelantado, E. and Toro, N. (2007) Dispersion of the RmInt1 group II intron in the *Sinorhizobium meliloti* genome upon acquisition by conjugative transfer. *Nucleic Acids Res.*, **35**, 214–222.
9. Fedorova, O. and Zingler, N. (2007) Group II introns: structure, folding and splicing mechanism. *Biol. Chem.*, **388**, 665–678.
10. Pyle, A.M. (2010) The tertiary structure of group II introns: implications for biological function and evolution. *Crit. Rev. Biochem. Mol. Biol.*, **45**, 215–232.
11. Murray, H.L., Mikheeva, S., Coljee, V.W., Turczyk, B.M., Donahue, W.F., Bar-Shalom, A. and Jarrell, K.A. (2001) Excision of group II introns as circles. *Mol. Cell*, **8**, 201–211.
12. Arnberg, A.C., Van Ommen, G.J., Grivell, L.A., Van Bruggen, E.F. and Borst, P. (1980) Some yeast mitochondrial RNAs are circular. *Cell*, **19**, 313–319.
13. Halbreich, A., Pajot, P., Foucher, M., Grandchamp, C. and Slonimski, P. (1980) A pathway of cytochrome b mRNA processing in yeast mitochondria: specific splicing steps and an intron-derived circular DNA. *Cell*, **19**, 321–329.
14. Peebles, C.L., Belcher, S.M., Zhang, M., Dietrich, R.C. and Perlman, P.S. (1993) Mutation of the conserved first nucleotide of a group II intron from yeast mitochondrial DNA reduces the rate but allows accurate splicing. *J. Biol. Chem.*, **268**, 11929–11938.
15. Peebles, C.L., Perlman, P.S., Mecklenburg, K.L., Petrillo, M.L., Tabor, J.H., Jarrell, K.A. and Cheng, H.L. (1986) A self-splicing RNA excises an intron lariat. *Cell*, **44**, 213–223.

16. van der Veen,R., Arnberg,A.C., van der Horst,G., Bonen,L., Tabak,H.F. and Grivell,L.A. (1986) Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing *in vitro*. *Cell*, **44**, 225–234.
17. Podar,M., Chu,V.T., Pyle,A.M. and Perlman,P.S. (1998) Group II intron splicing *in vivo* by first-step hydrolysis. *Nature*, **391**, 915–918.
18. Jarrell,K.A., Peebles,C.L., Dietrich,R.C., Romiti,S.L. and Perlman,P.S. (1988) Group II intron self-splicing. Alternative reaction conditions yield novel products. *J. Biol. Chem.*, **263**, 3432–3439.
19. Monat,C., Quiroga,C., Laroche-Johnston,F. and Cousineau,B. (2015) The Ll.LtrB intron from *Lactococcus lactis* excises as circles *in vivo*: insights into the group II intron circularization pathway. *RNA*, **21**, 1286–1293.
20. Dalby,S.J. and Bonen,L. (2013) Impact of low temperature on splicing of atypical group II introns in wheat mitochondria. *Mitochondrion*, **13**, 647–655.
21. Molina-Sánchez,M.D., Barrientos-Durán,A. and Toro,N. (2011) Relevance of the branch point adenosine, coordination loop, and 3' exon binding site for *in vivo* excision of the *Sinorhizobium meliloti* group II intron RmInt1. *J. Biol. Chem.*, **286**, 21154–21163.
22. Molina-Sánchez,M.D., Martínez-Abarca,F. and Toro,N. (2006) Excision of the *Sinorhizobium meliloti* group II intron RmInt1 as circles *in vivo*. *J. Biol. Chem.*, **281**, 28737–28744.
23. Li-Pook-Than,J. and Bonen,L. (2006) Multiple physical forms of excised group II intron RNAs in wheat mitochondria. *Nucleic Acids Res.*, **34**, 2782–2790.
24. Nagy,V., Pirakitikulr,N., Zhou,K.I., Chillón,I., Luo,J. and Pyle,A.M. (2013) Predicted group II intron lineages E and F comprise catalytically active ribozymes. *RNA*, **19**, 1266–1278.
25. Ichihyanagi,K., Beaugard,A., Lawrence,S., Smith,D., Cousineau,B. and Belfort,M. (2002) Retrotransposition of the Ll.LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. *Mol. Microbiol.*, **46**, 1259–1272.
26. Matsuura,M., Saldanha,R., Ma,H., Wank,H., Yang,J., Mohr,G., Cavanagh,S., Dunny,G.M., Belfort,M. and Lambowitz,A.M. (1997) A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Dev.*, **11**, 2910–2924.
27. Plante,I. and Cousineau,B. (2006) Restriction for gene insertion within the *Lactococcus lactis* Ll.LtrB group II intron. *RNA*, **12**, 1980–1992.
28. Zajac,P., Islam,S., Hochgerner,H., Lönnerberg,P. and Linnarsson,S. (2013) Base preferences in non-templated nucleotide incorporation by MMLV-derived reverse transcriptases. *PLoS One*, **31**, e85270.
29. Quiroga,C., Kronstad,L., Ritlop,C., Filion,A. and Cousineau,B. (2011) Contribution of base-pairing interactions between group II intron fragments during *trans-splicing in vivo*. *RNA*, **17**, 2212–2221.
30. Klein,J.R., Chen,Y., Manias,D.A., Zhuo,J., Zhou,L., Peebles,C.L. and Dunny,G.M. (2004) A conjugation-based system for genetic analysis of group II intron splicing in *Lactococcus lactis*. *J. Bacteriol.*, **186**, 1991–1998.
31. Pansegrau,W., Schröder,W. and Lanka,E. (1994) Concerted action of three distinct domains in the DNA cleaving-joining reaction catalyzed by relaxase (TraI) of conjugative plasmid RP4. *J. Biol. Chem.*, **269**, 2782–2789.
32. Chillón,I., Molina-Sánchez,M.D., Fedorova,O., García-Rodríguez,F.M., Martínez-Abarca,F. and Toro,N. (2014) *In vitro* characterization of the splicing efficiency and fidelity of the RmInt1 group II intron as a means of controlling the dispersion of its host mobile element. *RNA*, **20**, 2000–2010.
33. Lasda,E. and Parker,R. (2014) Circular RNAs: diversity of form and function. *RNA*, **20**, 1829–1842.