



RESEARCH ARTICLE

Bacterial group II introns generate genetic diversity by circularization and *trans*-splicing from a population of intron-invaded mRNAs

Félix LaRoche-Johnston, Caroline Monat, Samy Coulombe , Benoit Cousineau *

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

* benoit.cousineau@mcgill.ca



 OPEN ACCESS

Citation: LaRoche-Johnston F, Monat C, Coulombe S, Cousineau B (2018) Bacterial group II introns generate genetic diversity by circularization and *trans*-splicing from a population of intron-invaded mRNAs. *PLoS Genet* 14(11): e1007792. <https://doi.org/10.1371/journal.pgen.1007792>

Editor: Navtej Toor, Department of Chemistry and Biochemistry, University of California, UNITED STATES

Received: June 27, 2018

Accepted: October 26, 2018

Published: November 21, 2018

Copyright: © 2018 LaRoche-Johnston et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The RNA-Seq data are available from Sequence Read Archive (SRA) (NCBI/NIH) (accession number PRJNA503115). All other relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by a discovery grant from the Natural Sciences and Engineering Research Council of Canada (<http://www.nserc-crsng.gc.ca>) to BC (227826). FLJ previously received a Graduate Excellence Fellowship from

Abstract

Group II introns are ancient retroelements that significantly shaped the origin and evolution of contemporary eukaryotic genomes. These self-splicing ribozymes share a common ancestor with the telomerase enzyme, the spliceosome machinery as well as the highly abundant spliceosomal introns and non-LTR retroelements. More than half of the human genome thus consists of various elements that evolved from ancient group II introns, which altogether significantly contribute to key functions and genetic diversity in eukaryotes. Similarly, group II intron-related elements in bacteria such as abortive phage infection (Abi) retroelements, diversity generating retroelements (DGRs) and some CRISPR-Cas systems have evolved to confer important functions to their hosts. In sharp contrast, since bacterial group II introns are scarce, irregularly distributed and frequently spread by lateral transfer, they have mainly been considered as selfish retromobile elements with no beneficial function to their host. Here we unveil a new group II intron function that generates genetic diversity at the RNA level in bacterial cells. We demonstrate that LI.LtrB, the model group II intron from *Lactococcus lactis*, recognizes specific sequence motifs within cellular mRNAs by base pairing, and invades them by reverse splicing. Subsequent splicing of ectopically inserted LI.LtrB, through circularization, induces a novel *trans*-splicing pathway that generates exon 1-mRNA and mRNA-mRNA intergenic chimeras. Our data also show that recognition of upstream alternative circularization sites on intron-interrupted mRNAs release LI.LtrB circles harboring mRNA fragments of various lengths at their splice junction. Intergenic *trans*-splicing and alternative circularization both produce novel group II intron splicing products with potential new functions. Overall, this work describes new splicing pathways in bacteria that generate, similarly to the spliceosome in eukaryotes, genetic diversity at the RNA level while providing additional functional and evolutionary links between group II introns, spliceosomal introns and the spliceosome.

McGill University (www.mcgill.ca), a Canada Graduate Scholarship-Masters from the Natural Sciences and Engineering Research Council of Canada (<http://www.nserc-crsng.gc.ca>) and a Master's Research Scholarship from Fonds de Recherche en Nature et Technologies du Québec (<http://www.frqnt.gouv.qc.ca/accueil>); and currently holds an Alexander Graham Bell Canada Graduate Scholarship-Doctoral from the Natural Sciences and Engineering Research Council of Canada (<http://www.nserc-crsng.gc.ca>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

A hallmark of eukaryotic cells is the ability of their numerous introns, sequences that interrupt genes, to generate genetic diversity through the expression of several different protein variants from a single gene. These eukaryotic introns share a common ancestor with bacterial group II introns, which are mobile, scarce and maintained at low copy-levels. Because of this stark contrast, bacterial group II introns are considered as selfish mobile elements with no beneficial function to their hosts. As a challenge to this long-standing view, we describe here a new group II intron function that expands the genetic diversity and overall complexity of its bacterial host transcriptome. This ability of combining disparate mRNA fragments expands the repertoire of functional products at the disposal of the bacterial host. Our work thus provides a potential explanation of why group II introns were maintained in bacteria over the course of evolution, despite strong selective pressure to streamline their genomes. Moreover, generating new combinations of RNA molecules as a by-product of their otherwise selfish mobility pathway provides a new evolutionary link into how group II introns later evolved to fully support this function in the nuclei of eukaryotes.

Introduction

Bacterial group II introns are large RNA enzymes that mostly behave as retromobile elements [1–5]. Following their autocatalytic excision from interrupted RNA transcripts, they can reinsert within identical or similar DNA target sequences by retrohoming or retrotransposition, respectively [6–8]. These retromobile genetic elements are present in archaea, bacteria, and bacterial-derived organelles such as plant and fungal mitochondria, and plant chloroplasts [9]. While group II introns are somewhat infrequent in archaea, roughly one quarter of all sequenced bacterial genomes harbor one to a few copies displaying a broad phylogenetic distribution in the bacterial kingdom [10]. In sharp contrast, no functional group II introns were yet described in the nuclear genome of eukaryotes where they seem to be functionally excluded [11]. Although mitochondrial and chloroplast group II introns mainly interrupt housekeeping genes, bacterial group II introns are generally found in non-coding sequences and associated with other mobile genetic elements [5]. Organellar group II introns thus primarily function as classic intervening sequences while bacterial group II introns behave like mobile elements. Bacterial group II introns were also shown to propagate by conjugation within and between species, invading the chromosome or resident plasmids of their new hosts using either the retrohoming or retrotransposition pathways [12–14].

Group II introns require the assistance of RNA binding proteins called maturases to adopt their active three-dimensional conformation and self-splice *in vivo* [15]. Specific sequence motifs within group IIA introns mediate the accurate recognition of the 5' and 3' splice sites. Exon binding sequence 1 (EBS1) and 2 (EBS2) identify the 5' splice site by base pairing with complementary intron binding sequence 1 (IBS1) and 2 (IBS2) situated at the 3' extremity of the upstream exon. The 3' splice site is recognized by the δ - δ' base pairing interaction at the 5' extremity of the downstream exon. Group II introns self-splice from interrupted RNA transcripts through three different splicing pathways (Fig 1) [15]. The branching (Fig 1A), hydrolysis (Fig 1B) and circularization (Fig 1C) pathways release the intron as either branched structures called lariats, in linear forms or as closed circles, respectively. Each of these three splicing pathways involve two consecutive transesterification reactions (Fig 1, steps

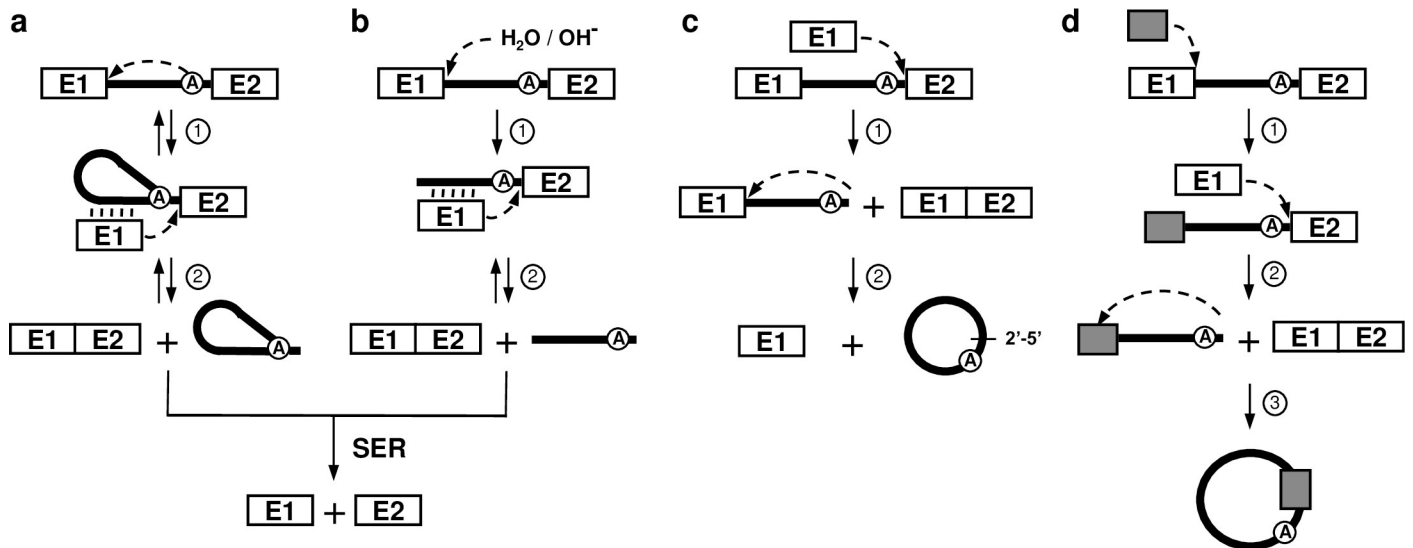


Fig 1. Group II intron splicing pathways. (a) Branching pathway. Following transcription of the interrupted gene, the 2'-OH residue of the branch-point nucleotide (A) initiates the first nucleophilic attack at the exon 1-intron junction (step 1). This transesterification reaction connects the 5' end 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. A 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. To explain the presence of additional nts at the splice junction of intron circles, the external nucleophilic attack pathway (d) was previously proposed [17, 19]. The 3'OH residue of a block of external nts (grey box) attacks the exon 1-intron junction, ligating it to the intron 5' end while concurrently displacing exon 1 (step 1). The 3'OH at the end of exon 1 then attacks the intron-exon 2 junction releasing ligated exons and a linear intron harboring external nts at its 5' end (step 2). The third transesterification reaction is initiated by the 2'-OH of the last intron residue (step 3). The position of this final nucleophilic attack thus dictates how many additional nts are incorporated at the junction of intron circles.

<https://doi.org/10.1371/journal.pgen.1007792.g001>

1 and 2). Branching, however, is the only splicing pathway that is completely reversible where intron lariats can recognize single- and double-stranded nucleic acid substrates (RNA/DNA) through base pairing and reinsert themselves by reverse splicing (Fig 1A, double arrows) [15, 16]. Since reverse splicing is the initial step of both group II intron mobility pathways, retrohoming and retrotransposition, only released intron lariats are active mobile elements [16].

We recently unveiled and characterized at the molecular level the circularization pathway of Ll.LtrB, the model group II intron, from the gram-positive bacterium *Lactococcus lactis* [17, 18]. Our work showed that the intron excises simultaneously through the branching and circularization pathways *in vivo* leading to the accumulation of both intron lariats and circles respectively. While the majority of the excised intron circles were found to have their 5' and 3' ends perfectly joined, we identified Ll.LtrB RNA circles harboring additional nucleotides at their splice junction. Here we describe novel group II intron splicing pathways in which the release of intron circles, harboring or not mRNA fragments of various lengths at their splice junctions, occurs concurrently with the generation of intergenic E1-mRNA and mRNA-mRNA chimeras *in vivo*. Overall, this study unveils that, similarly to spliceosomal introns in eukaryotes, bacterial group II introns generate genetic diversity at the RNA level, producing novel splicing products with potential new functions.

Results

Some excised Ll.LtrB RNA circles harbor mRNA fragments of various lengths at their splice junction

To study the splicing pathway leading to the incorporation of additional nucleotides at the splice junction of group II intron circles [17] we performed an RT-PCR reaction across the Ll.LtrB- Δ LtrA+LtrA lariat and circle splice junctions (Fig 2) [17, 18]. We cloned and sequenced the amplicons located in the faint smear above the RT-PCR band that corresponds to perfect lariat and circle splice junctions (Fig 2C). They revealed excised intron circles harboring additional nucleotides (nts) between the first and the last nts of the intron (Fig 3A). The stretch of additional nts greatly varied in size (20–576 nts), originated from the *L. lactis* chromosome or the two plasmids used to express the intron (Fig 2A) [17, 18] and mapped to the transcribed strand of annotated genes. Some sequences were identified more than once while others corresponded to different portions of the same gene.

Additional nts within the same size range (26–593 nts) and with identical characteristics (S1 Fig) were identified at the circle splice junction of Ll.LtrB-WT (Fig 2C). Taken together, these data show that mRNA fragments are incorporated at the splice junction of Ll.LtrB RNA circles during circularization, regardless if LtrA, the intron-encoded protein, is expressed in *trans* (Fig 3A) or in *cis* (S1 Fig).

IBS1/2-like sequences are present upstream of both extremities of the mRNA fragments incorporated at the Ll.LtrB circle splice junction

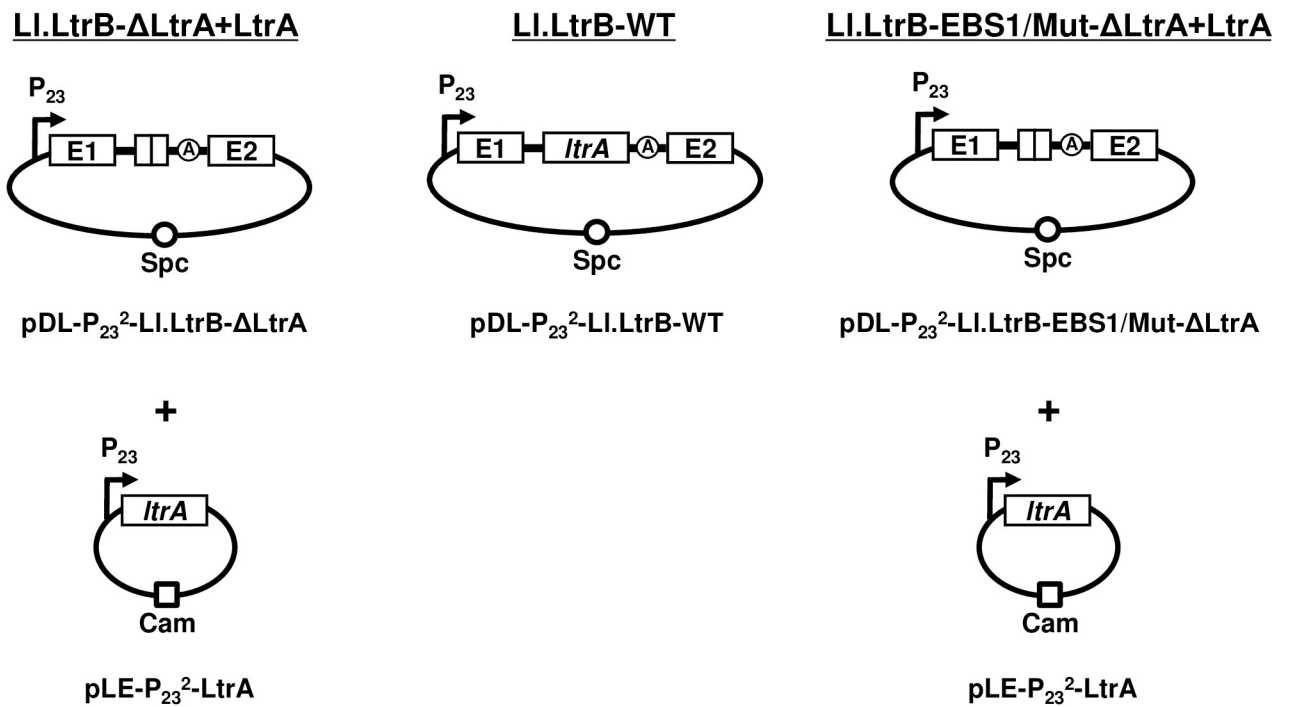
The flanking sequences on both sides of the mRNA fragments incorporated at the Ll.LtrB- Δ LtrA+LtrA (Fig 3A) and Ll.LtrB-WT (S1 Fig) circle splice junctions were retrieved, compiled and analyzed. Directly upstream from the 5' and 3' junctions we identified IBS1/2-like sequences partly complementary to the EBS1/2 sequences for both introns (Figs 3A and S1). Consensus sequences of 30 nts spanning the 5' and 3' junctions of the mRNA fragments confirmed the presence of IBS1/2-like sequence motifs. The IBS1-like motifs are better defined than the IBS2-like motifs, whereas the upstream IBS1/2-like motifs are stronger for both Ll.LtrB- Δ LtrA+LtrA and Ll.LtrB-WT (Fig 4A–4C).

Ll.LtrB recognizes both extremities of the mRNA fragments present at intron circle splice junctions through base pairing

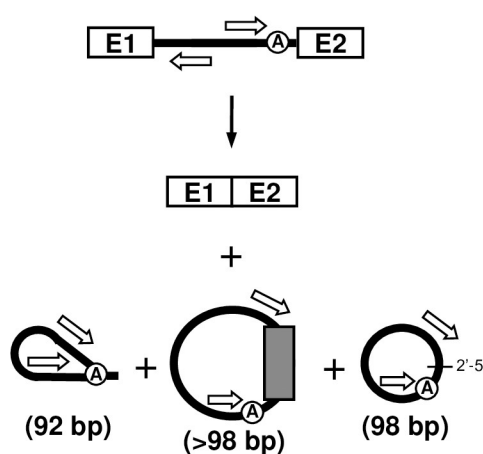
Comparable mRNA fragments of various lengths (43–452 nts) (Fig 3B) were also found at the circle splice junction of Ll.LtrB-EBS1/Mut- Δ LtrA+LtrA (Fig 2C), for which the EBS1 sequence was modified from 5'-GUUGUG-3' to 5'-CAACAC-3' (Fig 4D). Accordingly, the IBS1-like consensus sequence motifs upstream from both mRNA junctions were found to be different from Ll.LtrB- Δ LtrA+LtrA and Ll.LtrB-WT and complementary to the mutated EBS1 sequence (Fig 4E). In addition, similarly to Ll.LtrB- Δ LtrA+LtrA and Ll.LtrB-WT, the IBS1-like sequence motifs are both better defined than the IBS2-like motifs and the upstream IBS1/2-like motif much stronger.

The base pairing potential of Ll.LtrB-EBS1/Mut- Δ LtrA+LtrA is more stringent than Ll.LtrB- Δ LtrA+LtrA and Ll.LtrB-WT because its EBS1 sequence (5'-CAACAC-3') can perfectly recognize only 1 sequence (5'-GUGUUG-3'). In contrast, both introns harboring the wild-type EBS1 sequence (5'-GUUGUG-3') can base pair perfectly with 64 different sequence combinations using G = U wobble base pairings. Consequently, the more stringent EBS1 sequence led to a fainter RT-PCR smear (Fig 2C), the identification of fewer mRNA fragments at the intron circle splice junction (Fig 3B), and to much stronger flanking consensus motifs when

a



b



c

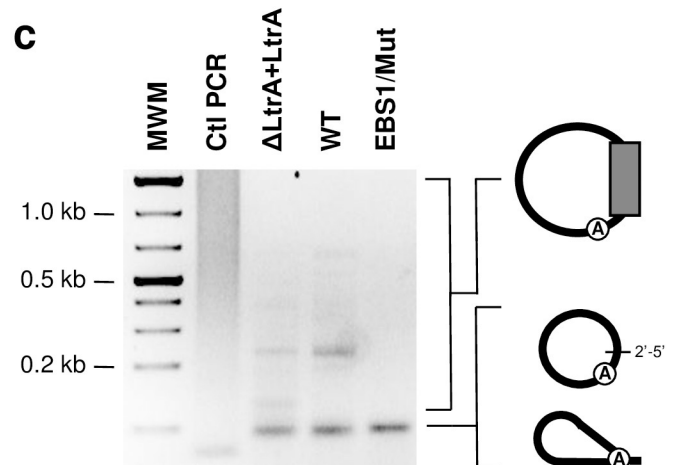


Fig 2. Detection of mRNA fragments at the splice junction of excised intron RNA circles. (a) Various LL.LtrB constructs used in this study where the LtrA protein is provided either in *trans* (LI.LtrB- Δ LtrA+LtrA, LI.LtrB-EBS1/Mut- Δ LtrA+LtrA) or in *cis* (LI.LtrB-WT) (b) Schematic of LL.LtrB self-splicing. Position of the primers (open arrows) (S2 Table) used to amplify the splice junction of excised introns by RT-PCR is depicted (92 bp (ariat), 98 bp (circle) or >98 bp (circle harboring additional nts)). (c) RT-PCR amplifications of intron splice junctions. Amplifications were performed on total RNA extracts from *L. lactis* (NZ9800 Δ ltrB) harboring different LL.LtrB constructs expressed under the control of the P_{23} constitutive promoter (Δ LtrA+LtrA: pDL- P_{23}^2 -LI.LtrB- Δ LtrA and pLE- P_{23}^2 -LtrA)(WT: pDL- P_{23}^2 -LI.LtrB-WT)(EBS1/Mut: pDL- P_{23}^2 -LI.LtrB-EBS1/Mut- Δ LtrA and pLE- P_{23}^2 -LtrA). The EBS1/Mut intron variant was shown to splice accurately and efficiently *in vivo* by RT-PCR amplifications of both released introns and ligated exons. Additional nts incorporated at the junction of intron circles are represented by a gray box.

<https://doi.org/10.1371/journal.pgen.1007792.g002>

a) Ll.LtrB-ΔLtrA+LtrA

Gene name	5' flanking		additional nts	3' flanking	
	EBS2 GUGUA	EBS1 GUGUUG		EBS2 GUGUA	EBS1 GUGUUG
Cam resistance (pLE-P ₂₃ ² -LtrA)	(2) AAAC CUCAA UACAGC	/UUUUGAACUGGU-----UAC CAUA GCAGCG	/AGAGUUAGGUUAUUG		
Cam resistance (pLE-P ₂₃ ² -LtrA)	(2) AAAC CUCAA UACAGC	/UUUUGAACUGGU-----UAC CAUA GCAGCG	/GAGAGUUAGGUUAUUG		
Cam resistance (pLE-P ₂₃ ² -LtrA)	(1) UAU UCCAUG GACUUC	/AUUUCUGG-----GUUU ACCU UAAA	/UAUCAUAUUAUUG		
Cam resistance (pLE-P ₂₃ ² -LtrA)	(1) GCC ACUUUA UACAAU	/UUUUGAUGGUGUAUC-----UAA ACAU UCUCUGG	/UAUUUGGACUCCUGU		
Unknown (pLE-P ₂₃ ² -LtrA)	(8) CAAG ACACAC CACACC	/AAAAUCAAAAAUUCC-17-CCAG GUAA CCACAAG	/AACUAUCCAAUCCA		
Unknown (pLE-P ₂₃ ² -LtrA)	(1) CAAG ACACAC CACACC	/AAAAUCA-----AA AAU UCACUAC	/UUUUGAUUAAAAACC		
Unknown (pDL-P ₂₃ ² -Ll.LtrB-ΔLtrA)	(2) AUG CGCAA CCCAACC	/CUUGGCAGAACAUUUC-16-UCC AGCAG CGCAACG	/CGGCGCAUCUCGGGC		
Unknown (pDL-P ₂₃ ² -Ll.LtrB-ΔLtrA)	(1) AUG CGCAA CCCAACC	/CUUGGCAGAACAUUUC-12-CAU CUCCAG CGAGCCG	/CAGCGGGCGCAUCUC		
Unknown (pDL-P ₂₃ ² -Ll.LtrB-ΔLtrA)	(1) AUG CGCAA CCCAACC	/CUUGGCAGAA-----CAU UCCAU CCG	/GUCCGCCAUUCUCCAG		
<i>ltrB</i> EI (pDL-P ₂₃ ² -Ll.LtrB-ΔLtrA)	(1) UAC ACAAA CACA <u>UU</u>	/AUUGUUCAUAAA <u>UU</u> ----AA AA CAU UAC CGCCAG	/GCAAAAGACUAUGUA		
Glucose-6 phosphate isomerase	(1) AU ACACU UCUACAAA	/UGUUCACUGAAAUGAU-13-CUG CACU UGCAUAU	/UCCUUUACCGUAAAG		
Glucose-6 phosphate synthetase	(1) AGC GACU UACACA <u>U</u>	/GUUAAUUGGAGCUGGU-23-AAA AGCU UACACUGa	/UCGAGUUCAACUUU		
Glucose-6 phosphate synthetase	(1) AGC GACU UACACA <u>U</u>	/GUUAAUUGGAGCUGGU-23-AAA AGCU UACACUGG	/UCGAGUUCAACUUU		
Glucose-6 phosphate synthetase	(1) AGC GACU UACACA <u>U</u>	/GUUAAUUGGAGCUGGU-24-AAA AGCU UACACUGGU	/CAGAUUGCAACUUUG		
Uncharacterized protein	(1) AAG CACA AGCACAAG	/CACAGUUGAU-----AG CUU CGCAUCA	/AAAGUUGACAGCUU		
Uncharacterized protein	(1) AAG CACA AGCACAAG	/CACAGUUGAU-----U AGCU UGCAUUC	/AAAAGUUGACAGCUU		
Beta-lactamase	(1) CUC CAACA ACACGAU	/GUGGAUACAAAAAU--21-AG CCAU GAACAUGCA	/AGUCCCUUUUUGAU		
Permease/transporter (DMU)	(1) UUA CACAC UCAUCC	/UUGCCACUUUUAA-----AG ACA UUCUCCAC	/UAAUUGUCGUCGCUU		
Glutamine synthetase	(1) UCU ACUUU ACCAAC	/GCGGUUAGGCUUUGC--5-GAC ACA UUGUAACG	/GAGCAGUCCGCGAA		
Glutamine synthetase	(1) CAU CACA ACCCAAC	/AGUUAAUCUA-----U ACA UACGUUUG	/GUUCCUGGCUAUGAA		
Cation transport ATPase	(1) GUU CAAAU UACAAA	/AUUUUUCAAACU-----UG GUUU ACAACG	/AAUAAUUGUUUUUG		
Ribosomal protein L4	(1) ACU UGACU UGCAGC	/UCCAAAAACGUCGAA-16-CAG CACU UUC <u>CAU</u> CG	/AACGUAAAGUACUUG		
Ribosomal protein L21	(1) AAA ACA UCCACAC	/GUAACAAGGCUACC---GUA ACA UUC <u>CAU</u> CAUA	/AGUUGUUAUCAAUC		
Glyceraldehyde-3P dehydrogenase	(3) AAAG UCAU ACCAUUC	/AAACCUUGUUCGU-----A CACU UGCAU <u>CU</u>	/UCGUAAAAUCGCUA		
Glycerol uptake facilitator	(5) CAC ACGU UGCUCA <u>U</u>	/ACAUUUAUUGCGCAAG-546-CAU CACU UGCAGCCG	/UUGCAUUUUUUAAA		
Glycerol uptake facilitator	(2) CAC ACGU UGCUCA <u>U</u>	/ACAUUUAUUGCGCAAG-546-CAU CACU UGCAGCCa	/UUGCAUUUUUUAAA		
Uncharacterized protein	(1) GGAA AAU UACAAC	/UCAGAAUUUACAGA-475-CAA UACA UUGCA <u>CA</u> CA	/AGGAUUGUUCUACA		
Alanyl-tRNA synthetase (alaS)	(2) ACU CACU UACU <u>CAU</u>	/GCUUGCUUGCAUAAU-274-GAA CACA UACGCA <u>AA</u>	/AACUUUGAAAAUUG		
Ser/Thr protein kinase	(1) AGG AAA AGCGCA <u>U</u>	/CAGAAUCAUCAUA-211-ACA ACA AGCACA <u>AC</u>	/CGUUGCCGAGGUAA		
Ser/Thr protein kinase	(1) AAA CAU UACU <u>CAU</u>	/CAAAAGUCUCCUCUA-166-CU U CACU <u>UACU</u> CAAA	/CAAUACAAUGUUUG		
Ser/Thr protein kinase/phos	(1) AAA CACA AGCAGG	/AGGUUAGUUGAUGAU-371-UGG CUCAU U <u>CAU</u> CAU	/GAAUUUUCACCGUG		
Ribosomal protein L19	(1) UCA ACCG UGCGCA <u>AC</u>	/UUGGCACUGAUUACC-173-CAG U UAC <u>ACU</u> CCAC	/GUGUUUGAAAAUUG		
Ribosomal protein S3	(1) AAA CAAGU UACA <u>UC</u>	/AACAUUGUUGAAUUC-242-AGA AGCAG ACA <u>AC</u>	/UUAUGGUAAACUCGG		
Ribosomal protein L13/S9	(2) ACU CACAC U <u>CAU</u> UC	/ACCCAGGUGGAAUUGA-202-ACA UAU UG <u>UACA</u> AG	/UACAAUUGCCGCGCA		
Ribosomal protein L13	(1) ACC ACA UACAC <u>AC</u>	/ACAUGCUGAUACAGG-236-ACA CACU U <u>CAGC</u> UC	/AAAAACCUAGAUCC		
2',3'-cyclic phosphodiesterase	(1) ACA CACU UCCAA <u>AA</u>	/AAAGAUAUUUAGCAG-185-AUG ACAU U <u>UACU</u> CG	/UGAGGAACUUCGCGCA		
Enolase (enoA)	(1) UUC CACU U <u>UACA</u> CU	/ACCUUGGCGGAUUCA-137-GAA AUCU U <u>CCAGC</u> Ca	/CUUAGAAAAUUCUU		
DNA gyrase	(1) UCU CUUCU U <u>CACA</u> AC	/ACGCAACGGAAUUGU-160-GCC AUUGU U <u>CGCGA</u> U	/AUGGGCCGUCGUGCA		

b) Ll.LtrB-EBS1/Mut-ΔLtrA+LtrA

Gene name	5' flanking		additional nts	3' flanking	
	EBS2 GUGUA	EBS1 CACAAC		EBS2 GUGUA	EBS1 CACAAC
Threonine dehydrogenase	(2) AAC CAGCU AG <u>GUGUUG</u>	/ACAUGGCAAAAUAG-132-UUG AUAA AG <u>GUUUUG</u>	/AUACUUUGAUUCAUC		
Hydrolase	(1) CAG CAU AG <u>GUGUUG</u>	/CUAUGGCGAUUGCUG-18--ACC ACA U <u>GUGUUG</u>	/CUCAUUUUUUGAAA		
Membrane protease	(1) CGU U CA <u>GUGUUG</u>	/GGAUUGCCGAACAAC-45--AAA ACU U <u>GUGUUG</u>	/CACUUGAUGAAGAAC		
Ribosomal protein S12/S7	(3) AAG ACAC AG <u>GUGUUG</u>	/UACUUUCUGUGGUG-161-AAA CGU AG <u>UUUUUG</u>	/GCAGAUCCAAUGUAC		
Elongation factor Tu	(2) CUU AA CA <u>AGCAGAC</u>	/CUUGUUGAUGAUGAA-112-AAC CACA U <u>GCGUUG</u>	/CUAAAGUUGAAGAAU		
Phosphomannomutase	(5) GAA CAGAC AG <u>GUGUUG</u>	/GCGGAGAAGCAAUG-189-CGA CACU U <u>GUGUUG</u>	/CGUAUUUGGUAAAAA		
Signal recognition particle	(1) AAA CAU U <u>GUGUUG</u>	/AUUGAUACGGCAGGU-178-ACA CACU U <u>GUGUUG</u>	/CGGCUUUUAUCAUUC		
Uracil Permease	(9) CAC ACA UGA <u>GUGUUA</u>	/CAAAUUUUUAGGUUC-241-GUA UACCG A <u>UGUUGC</u>	/AAAUCUAAAAGGAUA		
Ribose-P pyrophosphokinase	(1) AUU AA CA <u>UUGUCUUA</u>	/CCUUCUAUUGGUUUAU-13--GUA AGCU U <u>GUGUCU</u>	/GUGAACCAAUACAUA		
Relaxase (<i>ltrB</i>)	(1) AAC ACA U <u>UUGUUC</u>	/AUAAAUUAAAACAUAU-422-GAA CACU U <u>GUGCCG</u>	/CAUAUCAUUUUUAAU		
Putative Fe-S oxidoreductase	(1) UAC CGCAU AG <u>UUGUUG</u>	/UAGAAUAUUUAGCAG-185-AUG AUGU U <u>GAAAAU</u>	/GCGGCCGUUAGUGUU		
Putative transport protein	(1) GGU ACCGU G <u>GUGUCG</u>	/CACCAUUUGGUCAAG-147-ACU UAUA U <u>CUUUUA</u>	/GUGCGCCAGCAGUUG		

Fig 3. mRNA fragments identified at the splice junction of LL.LtrB circles. Additional nts are shown along with their flanking sequences (5' flanking) (3' flanking), their origin (Gene name) and frequency of identification between parentheses for LL.LtrB- Δ LtrA+LtrA (a) and LL.LtrB-EBS1/Mut- Δ LtrA+LtrA (b) circles. The junctions between the additional nts and their flanking regions (/) as well as the IBS1- (yellow) and IBS2- (green) like sequences are denoted. The bolded nts represent residues from the IBS1- and IBS2-like sequences that can potentially base pair with the intron's EBS1 and EBS2 sequences specified above. Sequences spanning two genes and including a short intergenic region are underlined. The genes in bold (*alaS*, *enoA*, *S12/S7*) were further studied for LL.LtrB reverse splicing analyses and the detection of E1-mRNA and mRNA-mRNA chimeras (Figs 7 and 8).

<https://doi.org/10.1371/journal.pgen.1007792.g003>

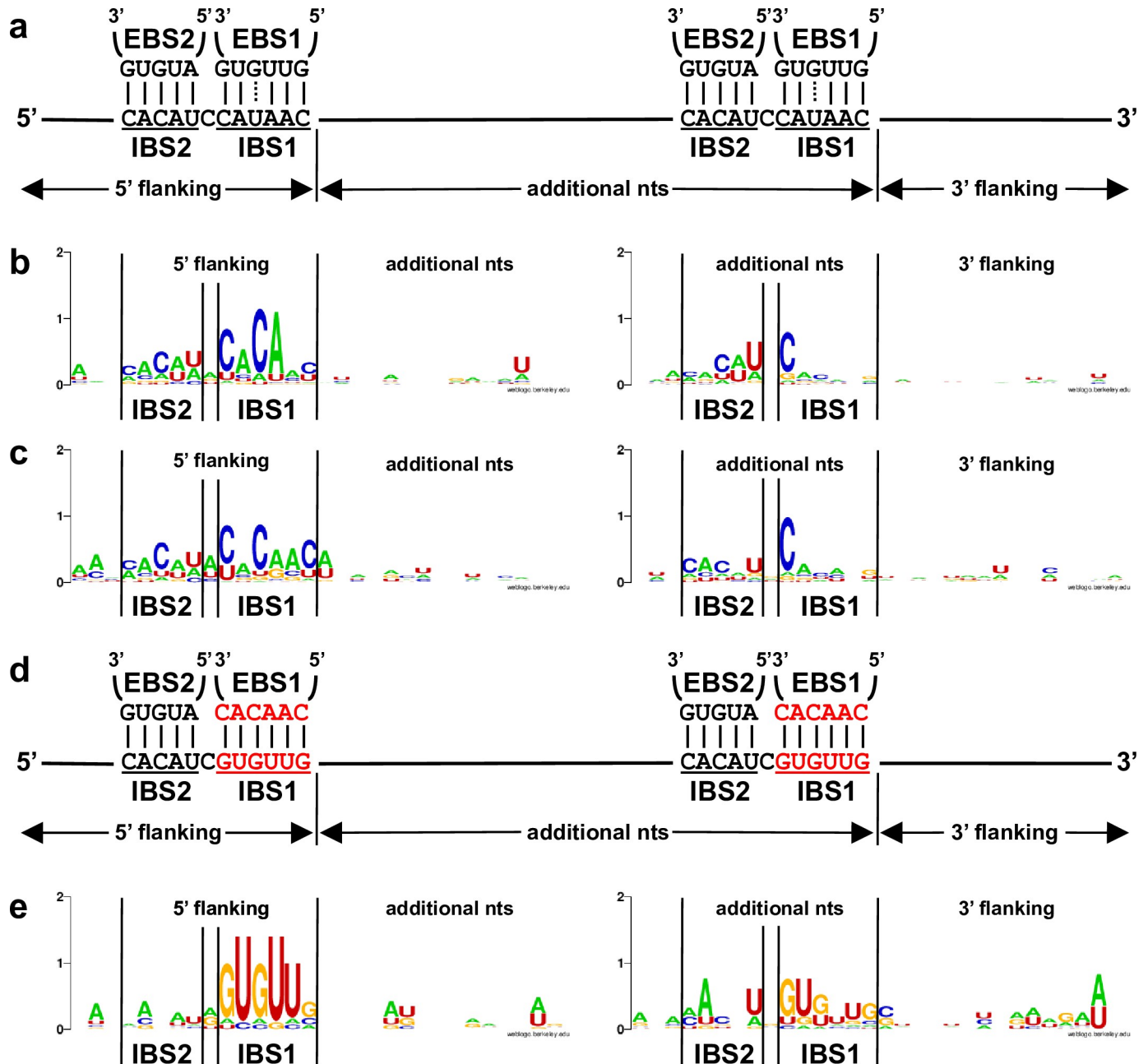


Fig 4. Logo representation of the consensus sequences (30 nts) around the 5' and 3' extremities of the mRNA fragments identified at intron circle splice junctions. The EBS1-IBS1 and EBS2-IBS2 base pairing interactions for LL.LtrB- Δ LtrA+LtrA and LL.LtrB-WT (a) as well as LL.LtrB-EBS1/Mut- Δ LtrA+LtrA (d) are depicted. The consensus sequences are shown for LL.LtrB- Δ LtrA+LtrA (Fig 3A)(b), LL.LtrB-WT (S1 Fig)(c) and LL.LtrB-EBS1/Mut- Δ LtrA+LtrA (Fig 3B)(e).

<https://doi.org/10.1371/journal.pgen.1007792.g004>

compared to Ll.LtrB- Δ LtrA+LtrA and Ll.LtrB-WT (Fig 4). These data confirm that both junctions of the incorporated mRNA fragments at intron circle splice junctions are recognized by the EBS1/2 motifs of Ll.LtrB through base pairing interactions during circularization.

Consensus sequences are slightly but consistently stronger when flexibility is allowed at both junctions of the mRNA fragments for all three constructs suggesting that Ll.LtrB does not always process mRNAs precisely downstream from the recognized IBS1/2-like motifs (S2–S5 Figs). We also identified mRNA fragments, at intron circle splice junctions, that either contained untranslated sequences or spanned two genes including the short intergenic regions of polycistronic mRNAs (Figs 3 and S1). This further supports our conclusion that Ll.LtrB can capture *L. lactis* transcripts at intron circle splice junctions during circularization.

Models of mRNA fragment incorporation at group II intron circle splice junctions

Our findings indicate that cellular mRNAs can somehow be incorporated at the Ll.LtrB circle splice junction during the circularization pathway. Two models can explain how mRNA fragments could be incorporated at the splice junction of group II intron circles (Fig 5).

The external nucleophilic attack pathway (Fig 5A) was previously proposed to explain how short stretches of additional nts could be incorporated at the circle splice junction during intron circularization. However, the pathway of integration and the origin of the additional nts were never demonstrated [17, 19]. Taking into consideration the data presented here, Ll.LtrB would recognize, through base pairing interactions, an IBS1/2-like sequence on an *L. lactis* mRNA and guide its hydrolysis downstream of the recognized sequence (step 1). Next, the 3'-OH of the processed mRNA would induce a transesterification reaction at the exon 1-intron splice junction resulting in its ligation to the 5' end of the intron and the release of exon 1 (step 2). The 3'-OH of exon 1 would then initiate the next transesterification reaction at the intron-exon 2 splice junction, releasing ligated exons and a linear intron harboring an mRNA fragment at its 5' end (step 3). The final transesterification reaction would be induced at the intron 5' end (step 4a) or within the mRNA (step 4b) by the 2'-OH of the last nt of the linear intron, just downstream from IBS1/2-like sequences, resulting in the release of either a head-to-tail circular intron or an intron circle harboring an mRNA fragment at its splice junction respectively.

An alternative pathway (Fig 5B) would rather be initiated by the reverse splicing of an intron lariat within an *L. lactis* mRNA downstream of an IBS1/2-like sequence (step 1). The ectopically inserted group II intron would then excise from the mRNA through circularization (steps 2–4). The 3'-OH of free exon 1 would first attack the phosphodiester bond at the 3' splice site between the last nt of the intron and the 3' segment of the mRNA (step 2). This would generate a chimeric mRNA consisting of the *ltrB*-exon 1 (E1) linked to the 3' segment of the mRNA (E1-mRNA) and a circularization intermediate where the linear intron is still attached to the 5' segment of the mRNA. The final transesterification reaction would then be induced at the intron 5' end (step 3a) or within the mRNA fragment (step 3b) by the 2'-OH of the last nt of the intron, just downstream from IBS1/2-like sequences, resulting in the release of either a head-to-tail circular intron or an intron circle harboring an mRNA fragment at its splice junction respectively.

Ll.LtrB lariats reverse splice within *L. lactis* mRNAs downstream of IBS1/2-like sequences

To investigate the proposed models we looked for unique intermediates of the reverse splicing pathway: the 3' junction of Ll.LtrB reverse-spliced within mRNAs and chimeric E1-mRNAs

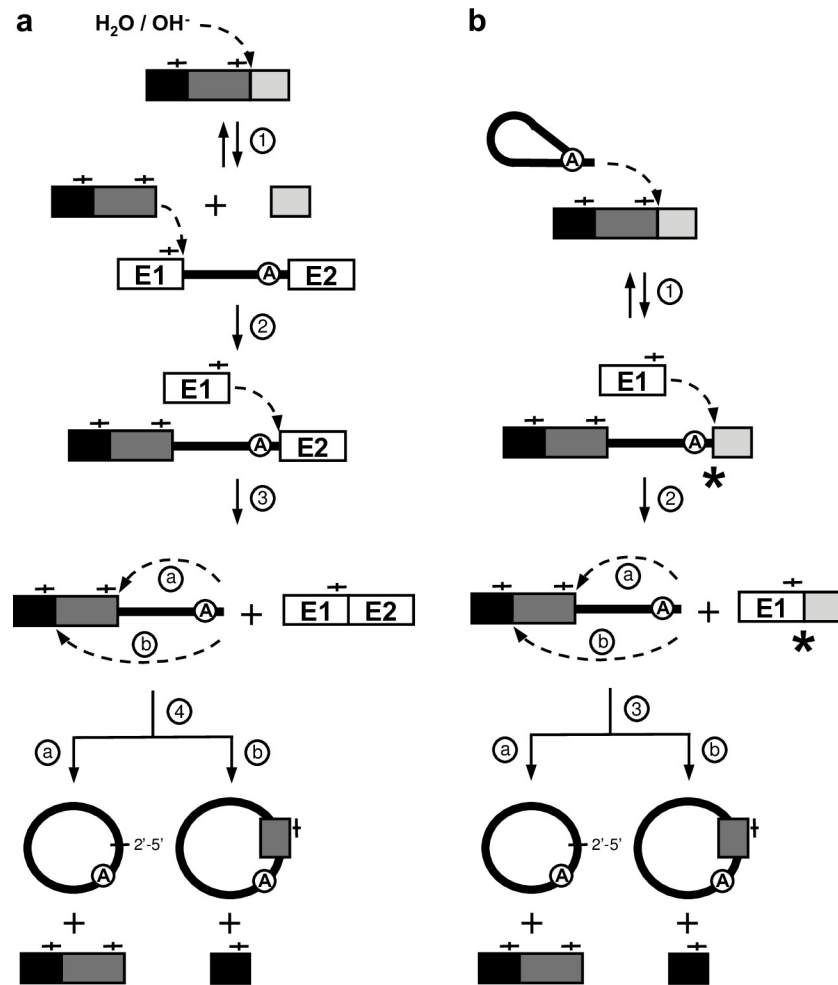


Fig 5. Models for the incorporation of mRNA fragments at the splice junction of intron RNA circles. (a) External nucleophilic attack pathway [17, 19]. The LI.LtrB group II intron recognizes, through base pairing interactions, an IBS1/2-like sequence (—|—) on an mRNA and guides the first nucleophilic attack induced by an hydroxyl ion or a water molecule downstream of the recognized sequence (step 1). Next, the 3'-OH of the processed mRNA induces a nucleophilic attack at the exon 1-intron splice junction resulting in its ligation to the 5' end of the intron and the release of exon 1 (step 2). The 3'-OH of exon 1 is then free to initiate the second transesterification reaction at the intron-exon 2 splice junction, releasing ligated exons and a linear intron harboring a fragment of mRNA at its 5' end (step 3). The final transesterification reaction is induced at the intron 5' end (a) or within the mRNA (b) by the 2'-OH of the last nt of the linear intron, just downstream from IBS1/2-like sequences (—|—), resulting in the release of either a head-to-tail circular intron (step 4a) or an intron circle harboring an mRNA fragment at its splice junction (step 4b). (b) Reverse splicing pathway. This pathway is initiated by the reverse splicing of an intron lariat within a non-cognate mRNA downstream of an IBS1/2-like sequence (—|—)(step 1). The 3'-OH of free exon 1 then attacks the phosphodiester bond at the 3' splice site between the last nt of the intron and the 3' segment of the mRNA (step 2). This generates a chimeric mRNA consisting of the *ltrB*-exon 1 (E1) linked to the 3' segment of the mRNA (E1-mRNA) and a circularization intermediate where the linear intron is still attached to the 5' segment of the mRNA. The third transesterification reaction is induced at the intron 5' end (a) or within the mRNA fragment (b) by the 2'-OH of the last residue of the linear intron, just downstream from IBS1/2-like sequences (—|—), resulting in the release of either a head-to-tail circular intron (step 3a) or an intron circle harboring an mRNA fragment at its splice junction (step 3b). The 3' junction of reverse-spliced introns and the chimeric E1-mRNAs are unique splicing intermediates that distinguish both pathways (asterisks).

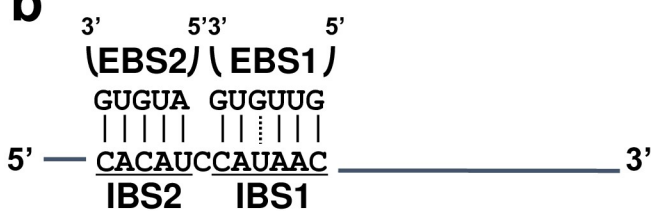
<https://doi.org/10.1371/journal.pgen.1007792.g005>

(Fig 5B, asterisks). We first detected by RNA-Seq intron-interrupted mRNAs for LI.LtrB- Δ LtrA+LtrA and LI.LtrB-EBS1/Mut- Δ LtrA+LtrA but not for the LI.LtrB- Δ A- Δ LtrA+LtrA control which lacks the essential branch point A residue required for branching and reverse

a

LL.LtrB Variants	WT	EBS1/Mut	ΔA
Independent Reverse Splicing Events	344	19	0
Number of Intron-Interrupted mRNAs	135	13	0

b



c

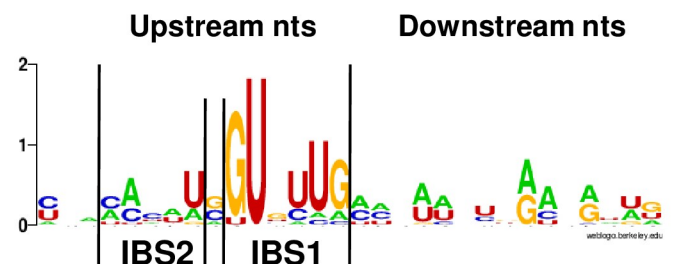
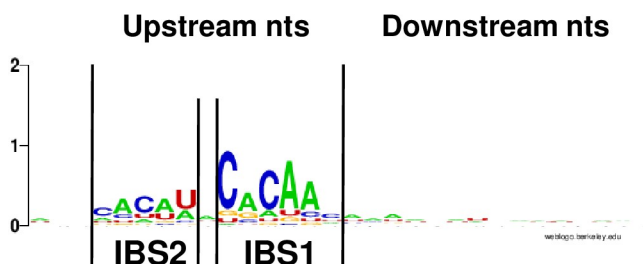
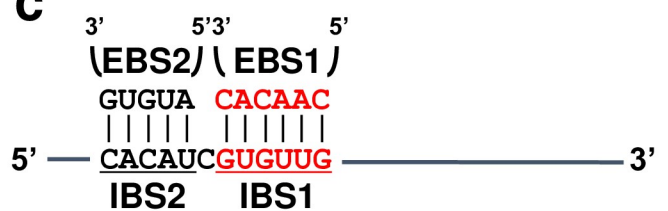


Fig 6. LL.LtrB reverse splicing within *L. lactis* mRNAs. (a) Independent LL.LtrB reverse splicing events were identified by total RNA-Seq for LL.LtrB- Δ LtrA+LtrA, LL.LtrB-EBS1/Mut- Δ LtrA+LtrA and the control LL.LtrB- Δ A- Δ LtrA+LtrA that exclusively splices through the circularization pathway and cannot reverse splice. The EBS1-IBS1 and EBS2-IBS2 base pairing interactions for LL.LtrB- Δ LtrA+LtrA (b) and LL.LtrB-EBS1/Mut- Δ LtrA+LtrA (c) are depicted. Logo representations of the consensus sequences upstream (15 nts) and downstream (15 nts) from the intron reverse splice sites within the various *L. lactis* mRNAs are also shown.

<https://doi.org/10.1371/journal.pgen.1007792.g006>

splicing [17, 18, 20] (Fig 6A). The reverse splice sites of LL.LtrB- Δ LtrA+LtrA (Fig 6B) and LL.LtrB-EBS1/Mut- Δ LtrA+LtrA (Fig 6C) were shown to be immediately preceded by consensus IBS1/2-like sequence motifs complementary to their respective EBS1/2 sequences. On the other hand, similarly to the junctions between intron circles and mRNA fragments (Figs 4 and S5), we did not detect a δ' -like sequence on the 3' side of the intron insertion sites (Fig 6). This shows that LL.LtrB can recognize IBS1/2-like sequences on various mRNAs by base pairing with its EBS1/2 sequences and invade them by reverse splicing, generating a population of intron-interrupted mRNAs in *L. lactis*. As expected, the more stringent EBS1 sequence of LL.LtrB-EBS1/Mut- Δ LtrA+LtrA led to the identification of fewer intron-interrupted mRNAs and a stronger IBS1/2-like consensus sequence upstream of the intron insertion sites compared to LL.LtrB- Δ LtrA+LtrA.

We next studied in further details the reverse splicing of LL.LtrB- Δ LtrA+LtrA within the Enolase (*enoA*) and Alanine-tRNA synthetase (*alaS*) mRNAs. The *enoA* (167 nts) and *alaS* (304 nts) mRNA fragments, previously identified at the LL.LtrB- Δ LtrA+LtrA circle splice junction,

are both flanked by a strong (10/11 nts) and a weak (7/11 and 8/11 nts) IBS1/2-like sequence motif (Fig 3A). We amplified by RT-PCR the 5' (Fig 7A and 7F) and 3' (Fig 7B and 7E) junctions between the intron and the two mRNAs. Sequences of the four amplicons confirmed reverse splicing of the intron precisely downstream of the strong IBS1/2-like sequence within the *enoA* (Fig 7C, large open arrowhead) and *alaS* (Fig 7G, large open arrowhead) mRNAs. Importantly, no amplifications were detected for the reverse splicing deficient control, Ll.LtrB- ΔA - $\Delta LtrA$ +LtrA. Next, the faint smears above (Fig 7A and 7E) and below (Fig 7B and 7F) the main amplicons were cloned and shown to correspond to several independent 5' and 3' junctions of the intron inserted downstream of different weak IBS1/2-like sequences (7-9/11 nts) (Fig 7C and 7G, black and grey arrowheads). The weak IBS1/2-like sequences flanking the mRNA fragments previously identified within intron circles (Fig 3A), were also found invaded by the intron for both *enoA* (Fig 7C, small open arrowhead) and *alaS* (Fig 7G, small open arrowhead). Similarly, the Ll.LtrB-EBS1/Mut- $\Delta LtrA$ +LtrA variant was shown to reverse splice at specific strong and weak IBS1/2-like sequences within the *S12/S7* transcript (Fig 8A–8C). The identified reverse splice sites also include the strong and weak IBS1/2-like sequences flanking the *S12/S7* mRNA fragment (161 nts) previously identified at the intron circle splice junction (Fig 3B).

Collectively, these results show that IBS1/2-like sequences are widespread within *L. lactis* mRNAs, providing abundant targets for Ll.LtrB reverse splicing. They also support the proposed alternative circularization model by which introns, reverse-spliced at ectopic sites within mRNAs, can circularize alternatively by recognizing upstream IBS1/2-like sequences leading to the capture of mRNA fragments at their splice junction (Fig 5B, step 3b). Accordingly, when additional nts are found at intron circle splice junctions, the upstream IBS1/2-like consensus sequences are consistently stronger (Figs 4 and S5) suggesting that when the intron reverse splices at a weak IBS1/2-like sequence, it is more likely to release intron circles harboring mRNA fragments by recognizing a stronger upstream alternative IBS1/2-like sequence.

Ll.LtrB circularization from intron-interrupted mRNAs generates E1-mRNA and mRNA-mRNA chimeras

The second distinguishing splicing intermediate between the two proposed models is a chimeric mRNA consisting of *ltrB*-exon 1 (E1) *trans*-spliced to an *L. lactis* mRNA fragment (E1-mRNA) (Fig 5B, asterisk). We specifically screened for E1-*enoA* and E1-*alaS* mRNA chimeras by RT-PCR. In both cases we detected, exclusively for the reverse splicing-competent intron, E1-mRNA chimeras ligated precisely downstream from the strong IBS1/2-like sequences (Fig 7D and 7H), the exact sites previously identified at one of the extremities of the mRNA fragments identified at intron circle splice junctions (Fig 3A) and invaded by reverse splicing (Fig 7A, 7B, 7E and 7F). The intron-catalyzed EBS1/2-specific generation of E1-mRNA chimeras was corroborated with the Ll.LtrB-EBS1/Mut- $\Delta LtrA$ +LtrA variant again at the previously identified strong IBS1/2-like sequence of the *S12/S7* transcript (Fig 8D). These results show that Ll.LtrB, reverse-spliced at IBS1/2-like sequences of various mRNAs, can recruit free E1 through EBS-IBS base pairing interactions, and catalyze the formation of E1-mRNA chimeras.

Ll.LtrB splicing *via* circularization, from a population of intron-interrupted mRNAs, generates processed mRNA fragments harboring IBS1/2-like sequences at their 3' end (Fig 5B, step 3a and 3b). We next examined if these splicing products could be recruited by Ll.LtrB, similarly to free E1 through EBS-IBS base pairing, and used to generate intergenic mRNA-mRNA chimeras (Fig 9). We detected by RT-PCR both *alaS-enoA* (Fig 7I) and *enoA-alaS* (Fig 7J) mRNA-mRNA intergenic chimeras joined at specific IBS1/2-like sequences for Ll.LtrB- $\Delta LtrA$ +LtrA but not for the Ll.LtrB- ΔA - $\Delta LtrA$ +LtrA control. These data show that Ll.LtrB, reverse-

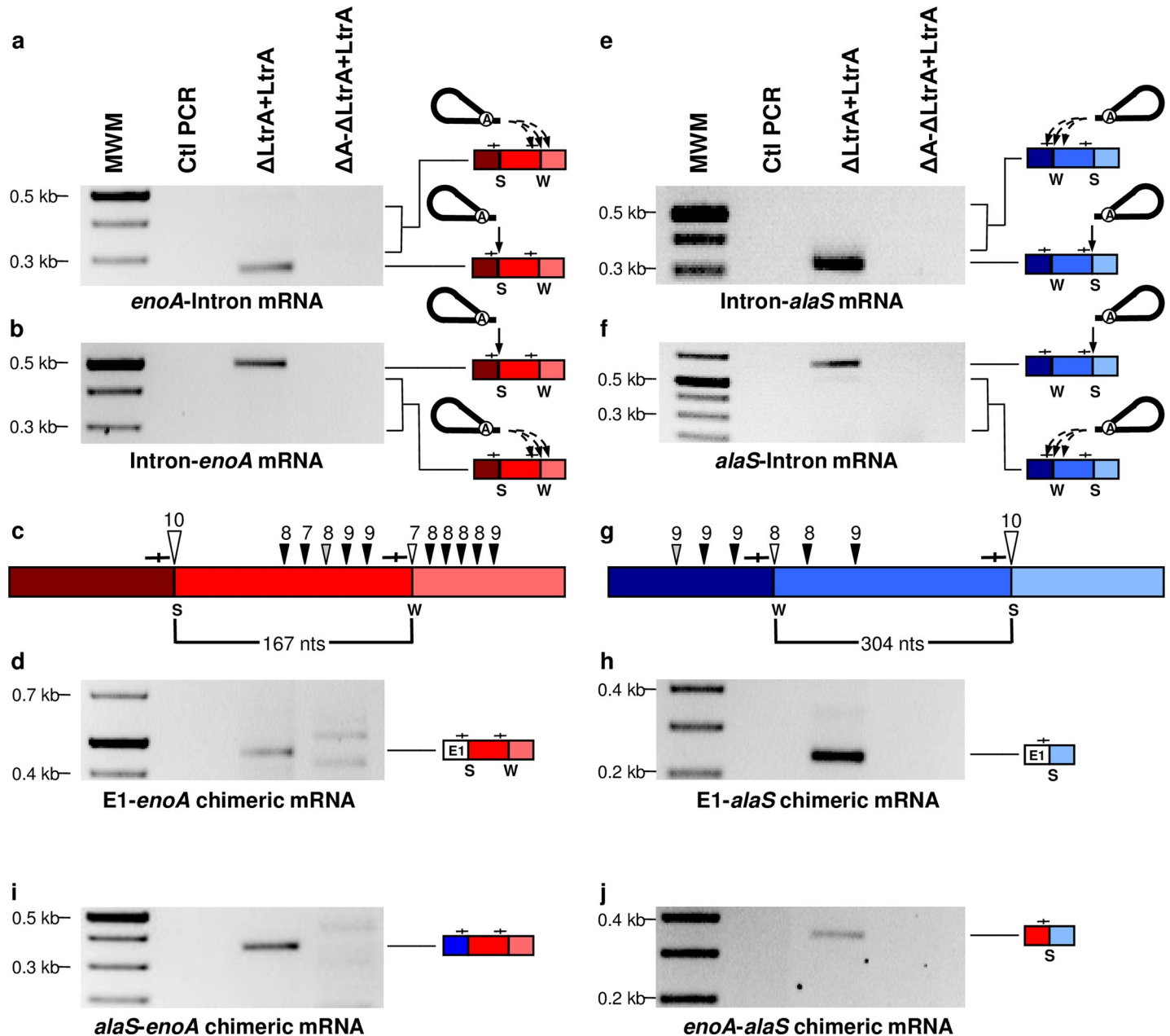


Fig 7. Detection of intermediates unique to the reverse splicing pathway: LI.LtrB reverse-spliced within *L. lactis* mRNAs, E1-mRNA and mRNA-mRNA chimeras. RT-PCR assays were performed to detect the 5' and 3' junctions of LI.LtrB- Δ LtrA+LtrA and LI.LtrB- $\Delta\Delta$ - Δ LtrA+LtrA reverse splicing events within the *enoA* (red boxes) (a, b) and *alaS* (blue boxes) (e, f) mRNAs. Complete and dashed arrows indicate reverse splicing of LI.LtrB lariats within strong (S) and weak (W) IBS1/2-like sequences (—|—) respectively. The strong (S)(10/11 nts)(large arrowhead) and weak (W)(7-9/11 nts)(small arrowhead) IBS1/2-like sequences invaded by reverse splicing are represented (c, g). The sites flanking the mRNA fragments (c, 167 nts and g, 304 nts) initially detected at intron circle splice junctions (Fig 3A) are indicated by open arrowheads. The LI.LtrB insertion sites were identified in conditions where the *enoA* or the *alaS* genes were overexpressed (small open and black arrowheads) or not (large open and small gray arrowheads) from a P_{23} constitutive promoter. mRNA chimeras between *ltrB*-exon 1 (E1) and *L. lactis* mRNAs (d, E1-*enoA*)(h, E1-*alaS*) as well as between *L. lactis* mRNAs (i, *alaS*-*enoA*)(j, *enoA*-*alaS*) were also detected by RT-PCR at IBS1/2-like sequences.

<https://doi.org/10.1371/journal.pgen.1007792.g007>

spliced within various mRNAs, can recruit through base pairing processed mRNA fragments, harboring IBS1/2-like sequences at their 3' end, to initiate the circularization splicing pathway (Fig 9, step 2). LI.LtrB can thus catalyze the shuffling of coding sequences within a population of intron-interrupted mRNAs by a new intergenic *trans*-splicing pathway (Fig 9).

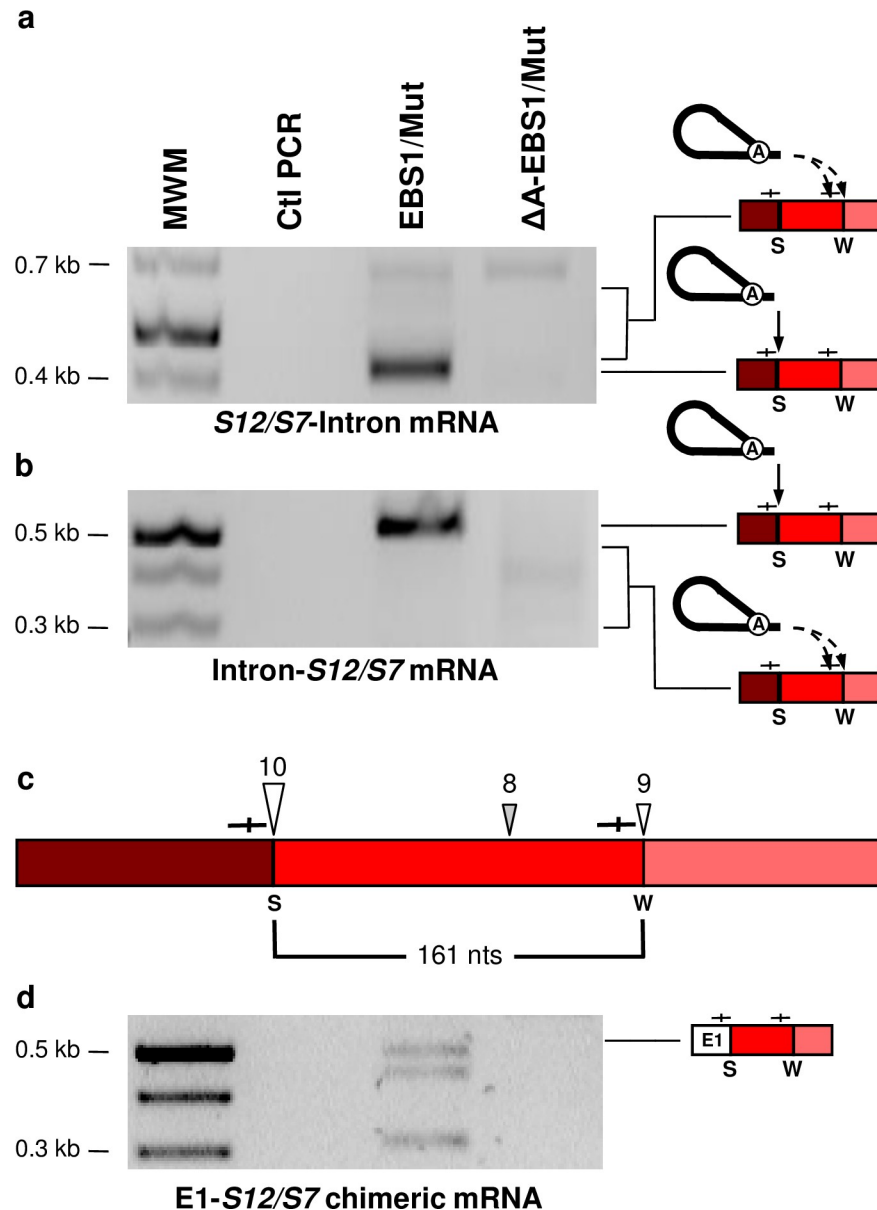


Fig 8. Detection of intermediates unique to the reverse splicing pathway: LI.LtrB reverse-spliced within an *L. lactis* mRNA and an E1-mRNA chimera. RT-PCR assays were performed to detect the 5' and 3' junctions of LI.LtrB-EBS1/Mut-ΔLtrA+LtrA and LI.LtrB-ΔA-EBS1/Mut-ΔLtrA+LtrA reverse splicing events within the Ribosomal Protein *S12/S7* mRNA (a, b). Complete and dashed arrows indicate reverse splicing of LI.LtrB lariats within strong (S) and weak (W) IBS1/2-like sequences (—|—) respectively. The strong (S)(10/11 nts)(large arrowhead) and weak (W)(8-9/11 nts)(small arrowheads) IBS1/2-like sequences invaded by reverse splicing are represented (c). The sites flanking the mRNA fragment (c, 161 nts) initially detected at intron circle splice junctions (Fig 3B) are indicated by open arrowheads. mRNA chimeras between *ltrB*-exon 1 (E1) and *L. lactis* mRNAs (d, E1-*S12/S7*) were also detected by RT-PCR at IBS1/2-like sequences.

<https://doi.org/10.1371/journal.pgen.1007792.g008>

Discussion

One quarter of currently sequenced bacterial genomes harbor one to a few group II introns [10]. This paucity, coupled with their irregular distribution and frequent lateral transfer [4], has led to the suggestion that they are selfish retromobile elements with no beneficial function

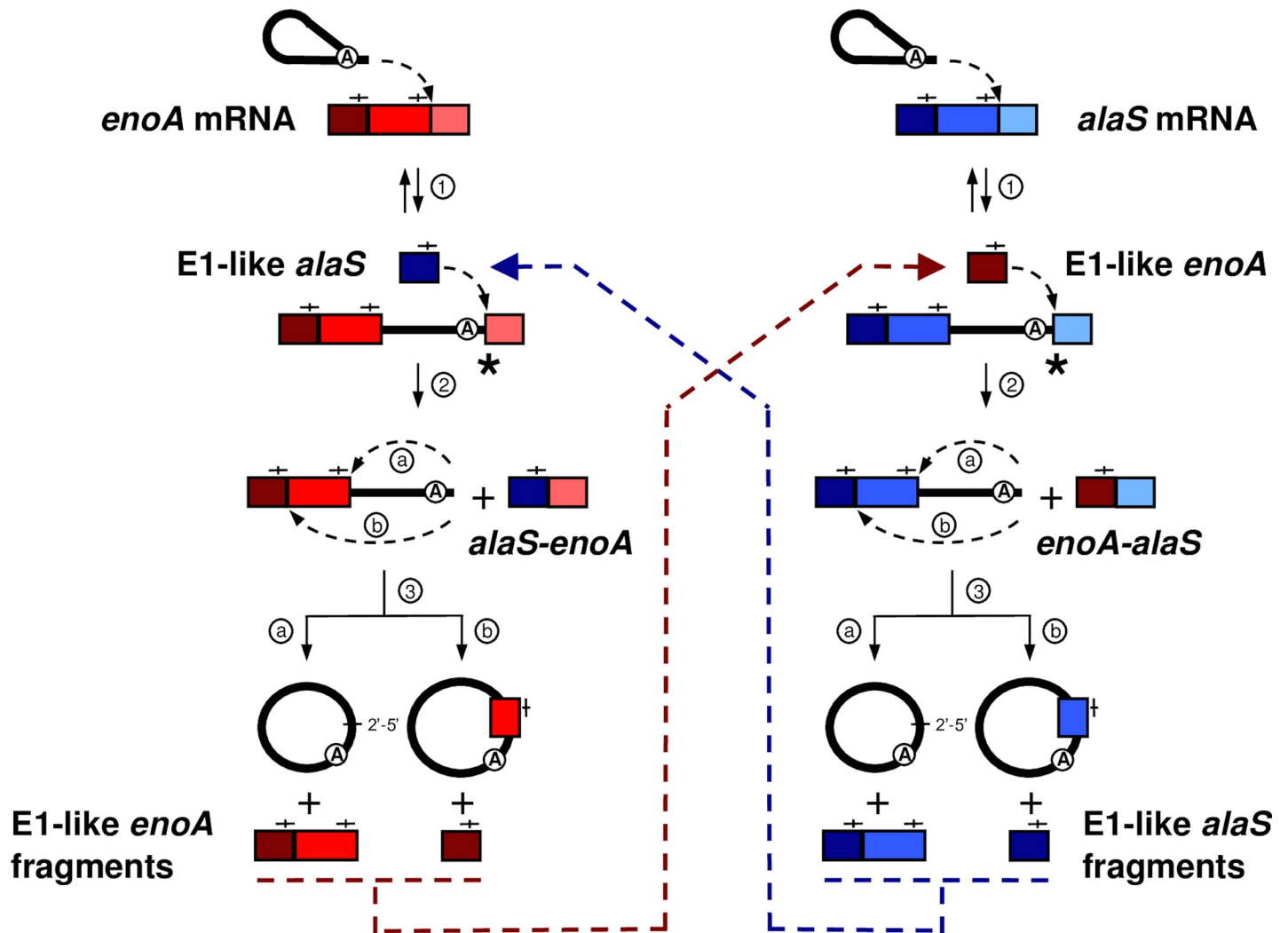


Fig 9. The intergenic group II intron trans-splicing pathway leading to the generation of *alaS-enoA* and *enoA-alaS* mRNA chimeras. LL.LtrB first recognizes by base pairing and invades by reverse splicing IBS1/2-like sequences on the *enoA* (red) and *alaS* (blue) mRNAs (step 1). The 3'OH of intron-processed *alaS* mRNA fragments (E1-like *alaS* fragments), harboring IBS1/2-like sequences (—|—) at their 3' ends (dark blue and blue)(dark blue), can, similarly to free E1, be recruited by the intron to initiate the circularization pathway. These fragments can attack the intron-exon 2 splice junction of the LL.LtrB-interrupted *enoA* mRNA (red) leading to the generation of *alaS-enoA* mRNA chimeras (step 2). Subsequent excision of the intron by circularization releases *enoA* mRNA fragments (E1-like *enoA* fragments) with IBS1/2-like sequences at their 3' ends (step 3a, dark red and red)(step 3b, dark red), which can in turn initiate the generation of *enoA-alaS* mRNA chimeras with intron-interrupted *alaS* mRNA (blue).

<https://doi.org/10.1371/journal.pgen.1007792.g009>

to their host [5]. In contrast, many group II intron derivatives provide important functions in both eukaryotes and prokaryotes [1–3]. For example, the abundant spliceosomal introns, descendants of group II introns, generate significant genetic diversity and transcriptomic complexity via alternative splicing [21], intergenic trans-splicing [22], RNA circle formation [23] and by creating new genes through exon shuffling [24].

Even though the LL.LtrB group II intron is present at only one copy in the *L. lactis* genome, the new splicing pathways described here (Figs 5B and 9) expand the genetic diversity and complexity of its host transcriptome. This stems from the ability of LL.LtrB, following its release as RNP particles, to generate a population of intron-interrupted mRNAs through reverse splicing, which we were able to detect by RNA-Seq (Fig 6) and gene-specific RT-PCR (Figs 7 and 8). LL.LtrB was recently shown to interact with its cognate ligated exons at the IBS1/2 site in

vivo, leading to either complete reverse splicing or negative regulation of targeted mRNA through hydrolysis and degradation [25]. However, when we contrasted the counts per million (CPM) of Ll.LtrB-WT, Ll.LtrB-EBS1/Mut- Δ LtrA+LtrA and Ll.LtrB- Δ A- Δ LtrA+LtrA constructs for *alaS*, the most abundant target for reverse splicing that we identified by RNA-Seq, we obtained differential expression ratios that showed very little change in the abundance of the *alaS* transcript: 0.97 between Ll.LtrB-WT and EBS1/Mut- Δ LtrA+LtrA and 0.96 between Ll.LtrB-WT and Ll.LtrB- Δ A- Δ LtrA+LtrA. This suggests that the IBS1/2-like sites we identified within host mRNAs are not efficient targets for hydrolysis, but rather seem to be used for reverse splicing. Interestingly, several of the reverse-splicing sites found by RNA-Seq were also identified independently at the extremity of mRNA fragments captured at intron circle splice junctions (Figs 3A and S1), yet there was only a small overlap of IBS1/2-like motifs between these two sets of data. Moreover, when we analysed the *enoA* and *alaS* genes in greater detail, we found a multitude of additional IBS1/2-like motifs that were used as targets for Ll.LtrB reverse splicing and whose base pairing interactions with the intron varied from strong (11/11 nts) to weak (7/11 nts) (Fig 7C and 7G). Overall, our data thus suggest that the reverse-splicing of group II introns into ectopic sites within host mRNAs is a widespread, dynamic and transient process whose exact scope is hard to determine.

We demonstrated that circularization of Ll.LtrB from interrupted mRNAs, using free E1 or mRNA fragments harboring IBS1/2-like sequences at their 3' end, generates two types of *trans*-spliced transcripts: E1-mRNA (Fig 5B) and mRNA-mRNA (Fig 9) chimeras respectively. Ll.LtrB was recently found to generate free E1 *in vivo* through hydrolysis of ligated cognate exons at the IBS1/2 site [25]. This Spliced Exon Reopening (SER) reaction (Fig 1) could thus produce the initial source of E1 required to initiate Ll.LtrB circularization from both its cognate exons and ectopic insertion sites. In addition, we found that alternative circularization of Ll.LtrB from interrupted mRNAs releases intron circles harboring mRNA fragments at their splice junction (Fig 5B, step 3b). These novel bacterial splicing products, generated by alternative circularization and intergenic *trans*-splicing, may have and/or lead to novel biological functions for their host cell. For instance, chimeric RNAs, intron circles and different circular RNAs that accumulate *in vivo* have been recently associated to a variety of interesting new functions such as RNA sponges, protein sponges and transcriptional regulators in various biological systems [23, 26, 27]. Moreover, the *trans*-spliced E1-mRNA and mRNA-mRNA chimeras could be reclaimed by the host and potentially lead to the creation of new genes. Group II introns may thus serve a beneficial function for their hosts by increasing the complexity and genetic diversity of their transcriptomes (Fig 10) which could explain why they were retained in bacteria.

Our work also unveils two additional functional and evolutionary links between group II introns, spliceosomal introns and the spliceosome. First, the *trans*-splicing of E1 at the 5' end of various mRNA fragments is analogous to the second step of the spliced leader (SL) *trans*-splicing pathway, which has a patchy evolutionary distribution amongst eukaryotes and whose origin has remained enigmatic [28, 29]. Second, we showed that group II introns, similarly to the spliceosome [22], can catalyze the *trans*-splicing of intergenic mRNA-mRNA chimeras in bacteria. Since group II introns are considered as the progenitors of both spliceosomal introns and the snRNAs of the spliceosome [1–3], our findings suggest that the spliceosome-dependent formation of SL *trans*-spliced transcripts and intergenic mRNA-mRNA chimeras in eukaryotes both consist of ancient group II intron splicing functions still shared with their contemporary bacterial relatives.

Overall, we described here new group II intron splicing pathways that generate and expand the genetic diversity and complexity of its host transcriptome which represents a new function for these bacterial retroelements. Our work also unveils new functional and evolutionary links

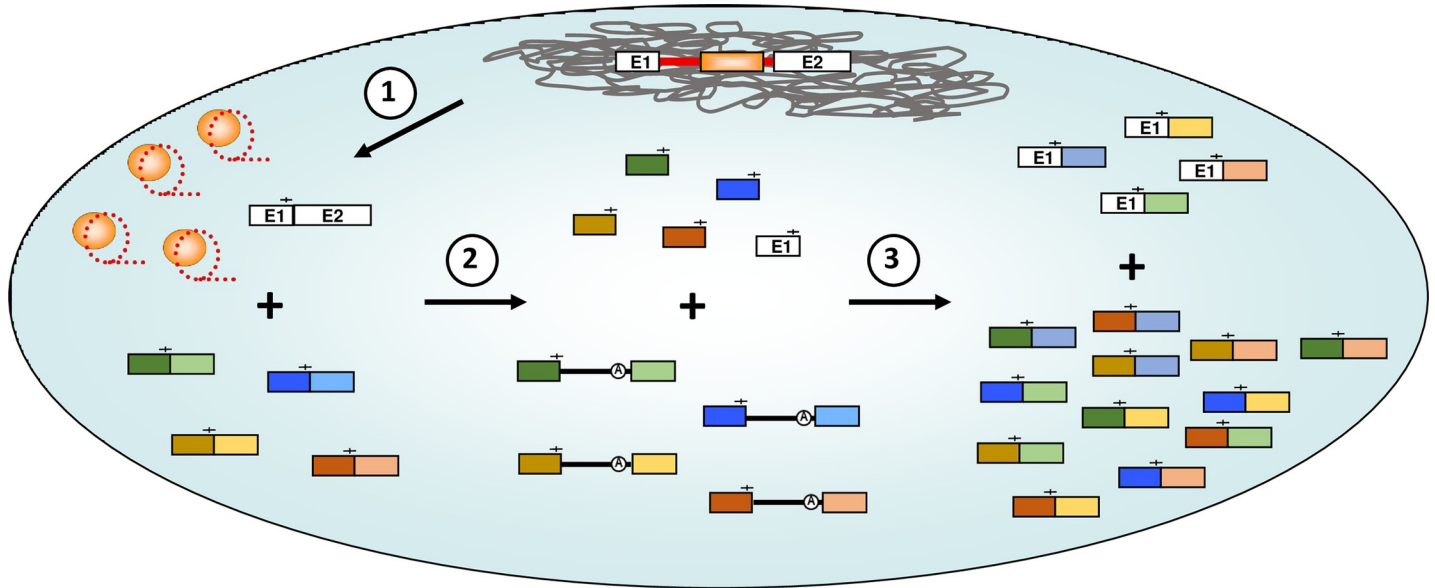


Fig 10. Model for group II intron-catalysed genetic diversity. Upon expression of group II intron-interrupted genes in bacteria, the ribozymes self-splice using the conventional branching pathway, releasing a mix of RNPs (lariats + LtrA) and accurately ligated flanking exons (step 1). Excised RNPs next interact with cellular mRNA transcripts through specific base pairing with IBS1/2-like sequences (—|—). This interaction leads either to complete reverse splicing or hydrolysis at the IBS1/2-like sites, producing a population of intron-invaded mRNA transcripts or hydrolysed mRNA fragments with a free 3'-OH, respectively (step 2). When introns interrupting an ectopic site self-splice using the circularization pathway, they recruit either processed ectopic mRNA fragments or their processed cognate E1, which can both act as external nucleophiles in an intergenic *trans*-splicing reaction (step 3). This produces two distinct populations of chimeric mRNA transcripts: E1-mRNA and mRNA-mRNA products, which together increase the overall diversity of the bacterial host's transcriptome. The presence of a series of group II intron-interrupted mRNAs may potentially lead to a multitude of chimeric mRNA-mRNA combinations.

<https://doi.org/10.1371/journal.pgen.1007792.g010>

with their nuclear relatives in eukaryotes, and provide a potential explanation of why group II introns were maintained in bacteria.

Materials and methods

Bacterial strains and plasmids

Lactococcus lactis strain NZ9800 Δ ltrB (Tet^R) [8] was grown in M17 media supplemented with 0.5% glucose (GM17) at 30°C without shaking. The *Escherichia coli* strain DH10 β , used for cloning purposes, was grown in LB at 37°C with shaking. Antibiotics were used at the following concentrations: chloramphenicol (Cam^R), 10 μ g/ml; spectinomycin (Spc^R), 300 μ g/ml. Previously constructed plasmids (pDL-P₂₃²-Ll.LtrB- Δ LtrA [30], pDL-P₂₃²-Ll.LtrB-WT [30], pLE-P₂₃²-LtrA [31]) were used to study Ll.LtrB splicing. Additional variants were constructed by site-directed mutagenesis (New England Biolabs Q5 Site-Directed-Mutagenesis Kit): pDL-P₂₃²-Ll.LtrB- Δ A- Δ LtrA, pDL-P₂₃²-Ll.LtrB-EBS1/Mut- Δ LtrA. The alanyl tRNA synthetase (*alaS*) and enolase (*enoA*) genes were cloned in pLE-P₂₃²-LtrA (BssHIII), downstream of the second P₂₃ promoter, and expressed with the intron in a two-plasmid system. Primers used for mutagenesis and cloning are in S2 Table.

RNA extraction, PCR and RT-PCR

Total RNA was isolated from NZ9800 Δ ltrB harboring various plasmid constructs as previously described [31]. RT-PCR reactions [17, 18] were performed on total RNA preparations of NZ9800 Δ ltrB harboring various intron constructs (primers in S2 Table).

RNA-Seq

RNA-Seq was performed on rRNA-depleted total RNA from *L. lactis* (NZ9800 Δ ltrB) expressing Ll.LtrB- Δ LtrA+LtrA, Ll.LtrB- Δ A- Δ LtrA+LtrA, or Ll.LtrB-EBS1/Mut- Δ LtrA+LtrA using the Illumina HiSeq 2500 paired-end sequencing system [32].

Sequence consensus

Aligned and adjusted consensus sequences were prepared using the WebLogo software [33]. Adjusted consensus sequences were determined using a code that calculated contiguous nucleotides with the highest capacity of base pairing to EBS1 and EBS2 (total of 11 nucleotides), separated from each other by 0–2 nucleotides, in a region that spanned -14, +4 nts around the junction with the intron.

Supporting information

S1 Fig. mRNA fragments identified at the splice junction of Ll.LtrB-WT circles.
(PDF)

S2 Fig. mRNA fragments identified at the splice junction of Ll.LtrB- Δ LtrA+LtrA circles.
(PDF)

S3 Fig. mRNA fragments identified at the splice junction of Ll.LtrB-WT circles.
(PDF)

S4 Fig. mRNA fragments identified at the splice junction of Ll.LtrB-EBS1/Mut- Δ LtrA+LtrA circles.
(PDF)

S5 Fig. Logo representation of the consensus sequences (30 nts) around the 5' and 3' extremities of the mRNA fragments identified at intron circle splice junctions.
(PDF)

S1 Table. Plasmids.
(PDF)

S2 Table. Primers.
(PDF)

Acknowledgments

We thank G. Burger for comments on the manuscript.

Author Contributions

Conceptualization: Félix LaRoche-Johnston, Caroline Monat, Benoit Cousineau.

Formal analysis: Félix LaRoche-Johnston, Caroline Monat, Samy Coulombe, Benoit Cousineau.

Funding acquisition: Benoit Cousineau.

Investigation: Félix LaRoche-Johnston, Caroline Monat, Samy Coulombe, Benoit Cousineau.

Methodology: Félix LaRoche-Johnston, Caroline Monat, Samy Coulombe, Benoit Cousineau.

Project administration: Benoit Cousineau.

Supervision: Benoit Cousineau.

Writing – original draft: Benoit Cousineau.

Writing – review & editing: Félix LaRoche-Johnston, Benoit Cousineau.

References

- Novikova O, Belfort M. Mobile Group II Introns as Ancestral Eukaryotic Elements. *Trends in genetics: TIG*. 2017; 33(11):773–83. <https://doi.org/10.1016/j.tig.2017.07.009> PMID: 28818345
- Lambowitz AM, Belfort M. Mobile Bacterial Group II Introns at the Crux of Eukaryotic Evolution. *Microbiol Spectr*. 2015; 3(1).
- Zimmerly S, Semper C. Evolution of group II introns. *Mob DNA*. 2015; 6:7. <https://doi.org/10.1186/s13100-015-0037-5> PMID: 25960782
- LaRoche-Johnston F, Monat C, Cousineau B. Recent horizontal transfer, functional adaptation and dissemination of a bacterial group II intron. *BMC Evol Biol*. 2016; 16(1):223. <https://doi.org/10.1186/s12862-016-0789-7> PMID: 27765015
- Dai L, Zimmerly S. Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. *Nucleic acids research*. 2002; 30(5):1091–102. PMID: 11861899
- Cousineau B, Smith D, Lawrence-Cavanagh S, Mueller JE, Yang J, Mills D, et al. Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell*. 1998; 94(4):451–62. PMID: 9727488
- Cousineau B, Lawrence S, Smith D, Belfort M. Retrotransposition of a bacterial group II intron. *Nature*. 2000; 404(6781):1018–21. <https://doi.org/10.1038/35010029> PMID: 10801134
- Ichiyanagi K, Beauregard A, Lawrence S, Smith D, Cousineau B, Belfort M. Retrotransposition of the LI.LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. *Molecular microbiology*. 2002; 46(5):1259–72. PMID: 12453213
- Lambowitz AM, Zimmerly S. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harbor perspectives in biology*. 2011; 3(8):a003616. <https://doi.org/10.1101/cshperspect.a003616> PMID: 20463000
- Candales MA, Duong A, Hood KS, Li T, Neufeld RA, Sun R, et al. Database for bacterial group II introns. *Nucleic acids research*. 2012; 40(Database issue):D187–90. <https://doi.org/10.1093/nar/gkr1043> PMID: 22080509
- Chalamcharla VR, Curcio MJ, Belfort M. Nuclear expression of a group II intron is consistent with spliceosomal intron ancestry. *Genes & development*. 2010; 24(8):827–36.
- Belhocine K, Plante I, Cousineau B. Conjugation mediates transfer of the LI.LtrB group II intron between different bacterial species. *Molecular microbiology*. 2004; 51(5):1459–69. <https://doi.org/10.1111/j.1365-2958.2004.03923.x> PMID: 14982638
- Belhocine K, Yam KK, Cousineau B. Conjugative transfer of the *Lactococcus lactis* chromosomal sex factor promotes dissemination of the LI.LtrB group II intron. *Journal of bacteriology*. 2005; 187(3):930–9. <https://doi.org/10.1128/JB.187.3.930-939.2005> PMID: 15659671
- Nisa-Martinez R, Jimenez-Zurdo JI, Martinez-Abarca F, Munoz-Adelantado E, Toro N. Dispersion of the RmInt1 group II intron in the *Sinorhizobium meliloti* genome upon acquisition by conjugative transfer. *Nucleic acids research*. 2007; 35(1):214–22. <https://doi.org/10.1093/nar/gkl1072> PMID: 17158161
- Fedorova O, Zingler N. Group II introns: structure, folding and splicing mechanism. *Biological chemistry*. 2007; 388(7):665–78. <https://doi.org/10.1515/BC.2007.090> PMID: 17570818
- Pyle AM. Group II Intron Self-Splicing. *Annu Rev Biophys*. 2016; 45:183–205. <https://doi.org/10.1146/annurev-biophys-062215-011149> PMID: 27391926
- Monat C, Quiroga C, Laroche-Johnston F, Cousineau B. The LI.LtrB intron from *Lactococcus lactis* excises as circles in vivo: insights into the group II intron circularization pathway. *Rna*. 2015; 21(7):1286–93. <https://doi.org/10.1261/rna.046367.114> PMID: 25956521
- Monat C, Cousineau B. Circularization pathway of a bacterial group II intron. *Nucleic acids research*. 2016; 44(4):1845–53. <https://doi.org/10.1093/nar/gkv1381> PMID: 26673697
- Li-Pook-Than J, Bonen L. Multiple physical forms of excised group II intron RNAs in wheat mitochondria. *Nucleic acids research*. 2006; 34(9):2782–90. <https://doi.org/10.1093/nar/gkl328> PMID: 16717283
- van der Veen R, Kwakman JH, Grivell LA. Mutations at the lariat acceptor site allow self-splicing of a group II intron without lariat formation. *The EMBO journal*. 1987; 6(12):3827–31. PMID: 2828039

21. Bush SJ, Chen L, Tovar-Corona JM, Urrutia AO. Alternative splicing and the evolution of phenotypic novelty. *Philos Trans R Soc Lond B Biol Sci.* 2017; 372(1713).
22. Lei Q, Li C, Zuo Z, Huang C, Cheng H, Zhou R. Evolutionary Insights into RNA trans-Splicing in Vertebrates. *Genome Biol Evol.* 2016; 8(3):562–77. <https://doi.org/10.1093/gbe/evw025> PMID: 26966239
23. Chen LL. The biogenesis and emerging roles of circular RNAs. *Nature reviews Molecular cell biology.* 2016; 17(4):205–11. <https://doi.org/10.1038/nrm.2015.32> PMID: 26908011
24. Franca GS, Cancherini DV, de Souza SJ. Evolutionary history of exon shuffling. *Genetica.* 2012; 140(4–6):249–57. <https://doi.org/10.1007/s10709-012-9676-3> PMID: 22948334
25. Qu G, Piazza CL, Smith D, Belfort M. Group II intron inhibits conjugative relaxase expression in bacteria by mRNA targeting. *eLife.* 2018; 7.
26. Hsiao KY, Sun HS, Tsai SJ. Circular RNA—New member of noncoding RNA with novel functions. *Exp Biol Med (Maywood).* 2017; 242(11):1136–41.
27. Cortes-Lopez M, Miura P. Emerging Functions of Circular RNAs. *Yale J Biol Med.* 2016; 89(4):527–37. PMID: 28018143
28. Krchnakova Z, Krajcovic J, Vesteg M. On the Possibility of an Early Evolutionary Origin for the Spliced Leader Trans-Splicing. *J Mol Evol.* 2017; 85(1–2):37–45. <https://doi.org/10.1007/s00239-017-9803-y> PMID: 28744787
29. Hastings KE. SL trans-splicing: easy come or easy go? *Trends in genetics: TIG.* 2005; 21(4):240–7. <https://doi.org/10.1016/j.tig.2005.02.005> PMID: 15797620
30. Matsuura M, Saldanha R, Ma H, Wank H, Yang J, Mohr G, et al. A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes & development.* 1997; 11(21):2910–24.
31. Belhocine K, Mak AB, Cousineau B. Trans-splicing of the LI.LtrB group II intron in *Lactococcus lactis*. *Nucleic acids research.* 2007; 35(7):2257–68. <https://doi.org/10.1093/nar/gkl1146> PMID: 17389638
32. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature.* 2008; 456(7218):53–9. <https://doi.org/10.1038/nature07517> PMID: 18987734
33. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome research.* 2004; 14(6):1188–90. <https://doi.org/10.1101/gr.849004> PMID: 15173120