RESEARCH PAPER

RNA Biology 11:8, 1061–1071; August 2014; Published with license by Taylor & Francis Group, LLC

Insights into the strategies used by related group II introns to adapt successfully for the colonisation of a bacterial genome

Laura Martínez-Rodríguez, Fernando M García-Rodríguez, María Dolores Molina-Sánchez, Nicolás Toro, and Francisco Martínez-Abarca*

Grupo de Ecología Genética; Estación Experimental del Zaidín; Consejo Superior de Investigaciones Científicas; Granada, Spain

Keywords: maturase, RmInt1, retroelements, ribozymes, reverse transcriptase, *Sinorhizobium meliloti*, *Sinorhizobium medicae*

Abbreviations: IEP, intron encoded protein; RNP, ribonucleoprotein; EBS, exon binding site; IBS, intron binding site; RT, reverse transcriptase

Group II introns are self-splicing RNAs and site-specific mobile retroelements found in bacterial and organellar genomes. The group II intron RmInt1 is present at high copy number in *Sinorhizobium meliloti* species, and has a multifunctional intron-encoded protein (IEP) with reverse transcriptase/maturase activities, but lacking the DNA-binding and endonuclease domains. We characterized two RmInt1-related group II introns RmInt2 from *S. meliloti* strain GR4 and Sr.md.I1 from *S. medicae* strain WSM419 in terms of splicing and mobility activities. We used both wild-type and engineered intron-donor constructs based on ribozyme ΔORF-coding sequence derivatives, and we determined the DNA target requirements for RmInt2, the element most distantly related to RmInt1. The excision and mobility patterns of intron-donor constructs expressing different combinations of IEP and intron RNA provided experimental evidence for the co-operation of IEPs and intron RNAs from related elements in intron splicing and, in some cases, in intron homing. We were also able to identify the DNA target regions recognized by these IEPs lacking the DNA endonuclease domain. Our results provide new insight into the versatility of related group II introns and the possible co-operation between these elements to facilitate the colonization of bacterial genomes.

Introduction

Group II introns are catalytic RNAs and mobile retroelements that are thought to be the ancestors of nuclear spliceosomal introns and non-long-terminal repeat (non-LTR) retrotransposons in higher organisms.[1-4] They were initially identified in the mitochondrial and chloroplast genomes of lower eukaryotes and plants, and were subsequently found in bacteria and archaea.[5-9] The many bacterial genome sequences already available have provided evidence for very different spreading rates of group II introns, with the number of copies in the genome ranging from one to the 28 closely related copies of a group II intron found in the thermophilic cyanobacterium *Thermosynechococcus elongatus*.[10]

A typical group II intron consists of a catalytically active intron RNA and an intron-encoded protein (IEP). The intron RNA is highly structured and folds into a conserved three-dimensional structure of six distinct double-helical domains, DI to DVI.[11] Based on the structure of the RNA, three main classes of group II introns (IIA, IIB, and IIC) have been described.[8] The IEP is encoded by an ORF within DIV. Group II IEPs have an N-terminal reverse transcriptase (RT) domain homologous to retroviral RT sequences, followed by a putative RNA-binding domain with RNA splicing or maturase activity (domain X) and, in some cases, a C-terminal DNA-binding (D) region followed by a DNA endonuclease (En) domain.[12-15] Classification and phylogenetic analyses of group II intron based on their IEPs have resulted in the definition of several main groups, named A, B, C, D, E, F, CL1 (chloroplast-like 1), CL2 (chloroplast-like 2), and ML (mitochondrion-like),[16-19] although additional types of intron ORF have recently been identified.[15] The introns of classes A, C, D, E, and F and the newly identified *g1* introns encode proteins with no En domain.[15]

The IEP of mobile group II introns is involved in the splicing reaction in vivo (maturase activity), generating an excised intron RNA lariat that remains bound to the IEP in a ribonucleoprotein complex (RNP).[5-8] Intron mobility involves the reverse

*Correspondence to: Francisco Martínez-Abarca; Email: fmabarca@eez.csic.es
Submitted: 06/26/2014; Accepted: 07/21/2014
http://dx.doi.org/10.4161/rna.32092

splicing of the intron RNA directly into a DNA strand, reverse transcription of the inserted intron RNA by the IEP, and integration of the resulting intron cDNA into the genome by host enzymes, independently of homologous recombination.[20,21] This mechanism is used by group II introns both to retrohome into specific DNA target sites at high frequency and to retrotranspose into ectopic sites, which is thought to facilitate intron dispersal.[20,22-26]

The RNP complex recognizes the intron DNA target via both the IEP and the intron lariat RNA. The central part of the target, containing the intron insertion site, is recognized by base pairing between the exon-binding sites (EBS) in the lariat RNA and the complementary region in the DNA target, the intron-binding sites (IBS).[5] The distal parts are thought to be recognized by the IEP, thus extending the DNA target site requirements to different lengths, depending on the group II intron studied.[10,27-30]

RmInt1 is one of the best studied introns. Its IEP lacks the D and En domains and it belongs to the IIB3 (D) subclass.[6,7] It is highly abundant in *Sinorhizobium meliloti*, the nitrogen-fixing symbiotic bacterium that lives in symbiosis with the roots of leguminous plants of the genus *Medicago*. This intron has been shown to be highly mobile in vivo,[31-33] and its main mobility pathway is characterized by a bias in the orientation of the DNA target with respect to the replication fork, indicating that cDNA synthesis is primed by either the nascent lagging strand or Okazaki fragments at the DNA replication fork.[32] RmInt1 has recently been included in the group of bacterial group II introns available for gene targeting,[30,34] also known as targetrons.[35]

In this study, we characterized two abundant group II introns phylogenetically related to RmInt1: RmInt2, seven copies of which are present in *S. meliloti* strain GR4, and Sr.md.I1, four copies of which are present in *S. medicae* strain WSM419. We found that these introns were functional, as they were able to splice and retrohome into both their natural and optimised DNA target sites. The excision and mobility patterns of combinations of IEPs and intron RNAs from these introns and changes to distal of the DNA target provide information about DNA target recognition by these IEPs. Our results highlight the different strategies followed by related introns to ensure their successful dispersion in the bacterial genome.

Results

Characteristics of RmInt1-related group II introns

RmInt1-like elements have been detected in rhizobia by DNA-DNA hybridization, and by BlastN searches of the genomes of *Sinorhizobium* and *Rhizobium* species.[36,37] However, full-length known relatives of the *S. meliloti* RmInt1 intron (85–99% identity) were found only in the closest relatives of *S. meliloti*: *S. medicae*, *E. adhaerens* and *S. terangae*. Within the various strains of *S. meliloti*, the full-length RmInt1 copies currently present display considerable sequence conservation (>99% identity).

The sequencing of *S. meliloti* strain GR4,[38] the host in which RmInt1 was first identified and from which it was isolated, showed there to be 10 copies of this intron present in three of the five replicons of this strain. Additionally, we identified several copies of a related element (72% nucleotide identity), hereafter referred as to RmInt2. In contrast to RmInt1, this element was absent from the other four *S. meliloti* strains for which complete sequenced genomes are available,[39-42] suggesting that it was acquired recently. All these copies were located within the inverted repeats of several copies of the insertion sequence IS*Rm17* (**Fig. S1**). These inverted repeats provide two putative RmInt2 insertion sites for each IS element (located in opposite replication strands).

The only complete genome sequence available for *S. medicae* is that for strain WSM419.[43] A search for RmInt1-related elements showed this genome to contain four complete copies of a group II intron 88% identical to RmInt1: Sr.md.I1.[19] Three of these copies are located within an ORF annotated as transposase IS66, belonging to a repetitive unit of four ORFs (**Fig. S1**). Based on their IEPs, RmInt1, RmInt2, and Sr.md.I1 are bacterial ORF class D group II introns.[7,16] The secondary RNA structure predicted for RmInt1 (structural class IIB), used as a template, shows that many of the sequence differences between the three introns correspond to compensatory changes in stem regions (**Fig. 1A**). Thus, all in the nucleotide differences detected in Sr.md.I1 are substitutions, with the exception of a C nucleotide insertion three nucleotides downstream from the EBS2 of domain I. However, RmInt2 differed more markedly from RmInt1 than the *S. medicae* intron. In particular, four compensatory changes were identified in the catalytic domain of the ribozyme (domain V). In addition, the nucleotide changes within domain VI of RmInt2 generate a stable stem capped with a GAAU loop (4 nucleotides) rather than the six nucleotides found in the other two introns (shown as a detail in **Figure 1A**). The potential primary binding site (subdomain DIVa) for these intron IEPs also displays compensatory changes and is identical for both RmInt2 and Sr.md.I1, but the middle stem (ii) may be shorter due to the replacement of the U-A pairing of RmInt1 with a CU mismatch.

Like the intron RNAs, the IEPs were found to be very similar (72–87% identity in pairwise comparisons), with a strong conservation of amino-acid residues known to be required for the activity of RmInt1 and other group II IEPs (**Fig. 1B**). The 19 amino acids of the C-terminal extension were the most variable, but the conserved sequence corresponding to the class D motif ($LX_3AX_3PXLF[V/A]HW$)[44] was conserved in all three introns.

Sr.md.I1 and RmInt2 group II introns are mobile at their natural DNA target sites

The DNA target site of RmInt1 spans a region from positions –20 to +5 nt with respect to the intron insertion site, in which different recognition elements can be identified.[27,30] This region encompasses the IBSs, which base pair with the EBS sequences in the intron RNA, denoted EBS1, 2, and 3, and the distal 5′- and 3′-exons, which have no predicted interaction partner within the intron RNA. The comparison of all genomic sites for the Sr.md.I1 and RmInt2 introns indicated a high degree of IBS sequence conservation, consistent with insertion by retrohoming (**Fig. S2**)[32] However, particular differences may be observed in the predicted EBS2-IBS2 interaction of RmInt2. The EBS2

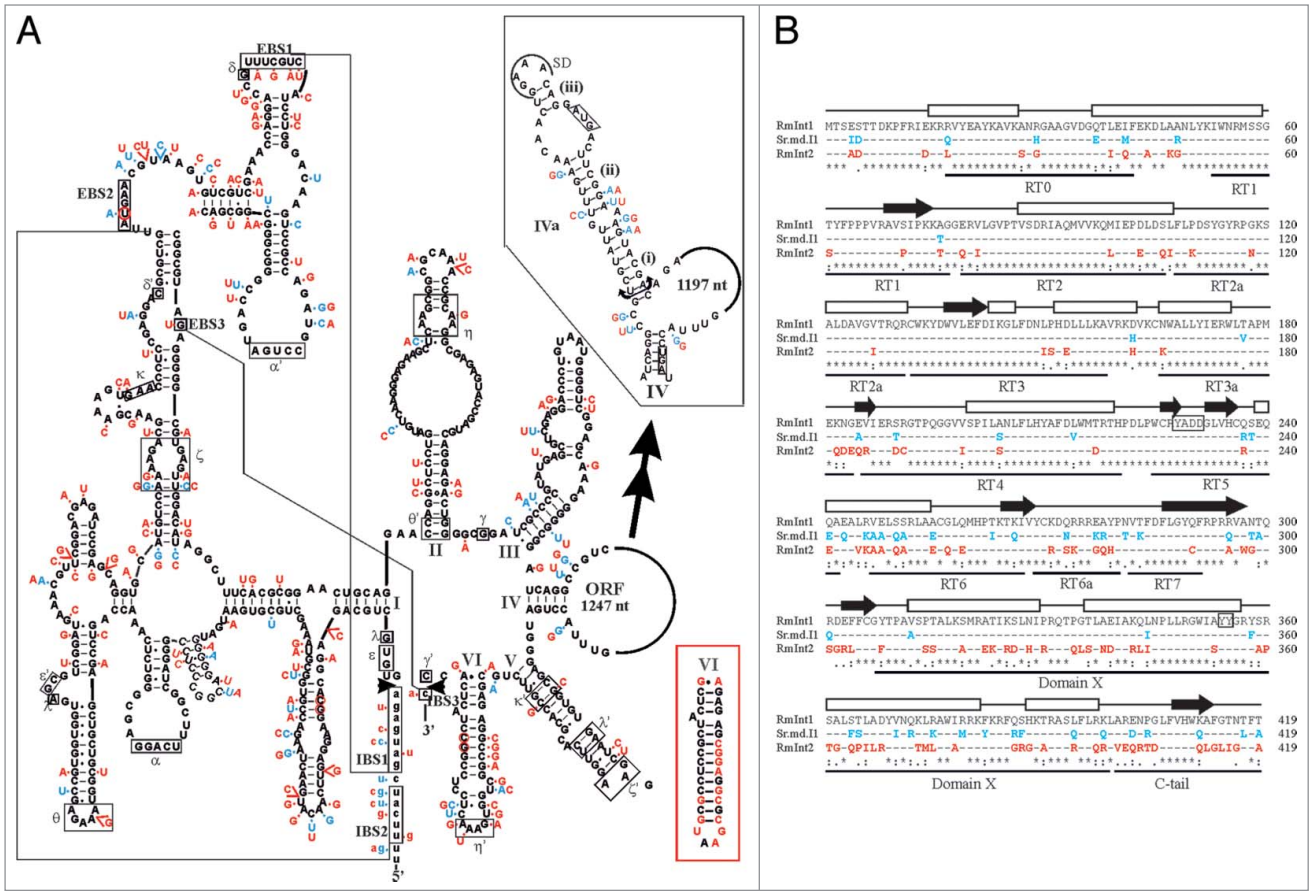

Figure 1. RmInt1-related group II introns. (**A**) Predicted secondary structure of RmInt1. Differences between RmInt1 and RmInt2 and Sr.md.I1 are indicated in red, boxed (DVI), and blue letters respectively. Nucleotide insertions and deletions are indicated with brackets and circles, respectively. Roman numerals correspond to the six conserved domains present in group II introns. Greek letters and boxes, linked by lines, indicate potential tertiary interactions between intron domains. The 5′ and 3′ exons are shown in lower-case letters and the splice sites are indicated by black arrowheads. Interactions between IBS-EBS 1, 2 and 3 of RmInt1 shown by boxes linked by lines. The predicted secondary structure of DIVa (the potential primary binding site for the intron IEPs), which contains the Shine–Dalgarno (SD) sequence and the initiation and stop codons (boxed) of all the IEP ORFs, is shown on the right. (**B**) Alignment of IEP sequences. RmInt1 IEP was aligned with RmInt2 and Sr.md.I1 IEPs, using MAFFT. Residues identical to those in the RmInt1 sequence are indicated by dashes, with changed residues indicating mismatches. RmInt1 IEP domain boundaries are shown below the alignment. The predicted secondary structure of the RmInt1 IEP, based on the JPRED folding prediction, is shown above the alignment: boxes and arrows indicate α-helices and β-strands, respectively. Critical aminoacids for RT and maturase activities[44] are boxed.

region within the RmInt2 intron RNA is shifted by one nucleotide with respect to that of RmInt1 and Sr.md.I1, and its predicted 5′exon linker corresponds to two positions (–8, –9; **Fig. 2**). Both novel introns present a mismatch, potentially affecting EBS2-IBS2 interaction. Finally, an important difference in EBS3-IBS3 recognition was observed, with Sr.md.I1 and RmInt1 making use of G-C pairing, whereas RmInt2 makes use of U-A pairing. Distal parts of the target are less conserved, but the critical nucleotide for RmInt1 retrohoming, T-15,[27] upstream from the IBS2 site, was found to be conserved in all intron target sites.

We used a plasmid mobility assay previously described for RmInt1 to assess the mobility of these new introns.[32,33,45] Engineered intron donor constructs were generated for each group II intron, flanked by a short stretch of the corresponding flanking insertion site (–20 to +5). DNA target sites (**Fig. 2**) were inserted into recipient plasmids, in both orientations with respect to the direction of DNA replication.[32] A band corresponding to the intron invasion event was clearly detected with a probe directed against the recipient plasmid (**Fig. 2C**). Intron invasion was found to be dependent on the presence of the DNA target site (not shown). Despite the presence of mismatches in the IBS2-EBS2 recognition regions of the DNA targets for both introns, these introns displayed retrohoming. As expected for class D endonuclease-minus introns, RmInt2 and Sr.md.I1 displayed a retrohoming bias toward the strand serving as a template for the nascent lagging strand in DNA replication forks (LAG constructs).[32] A second homing product, 2 kb larger than the first, was detected in the RmInt2 mobility assay (**Fig. 2C**, lane 1). Analysis of the sequence of this homing product indicated that it corresponded to an intron insertion generating two head-to-tail intron copies, possibly produced because the EBS3 (U) is complementary to the first nucleotide of the intron (G).

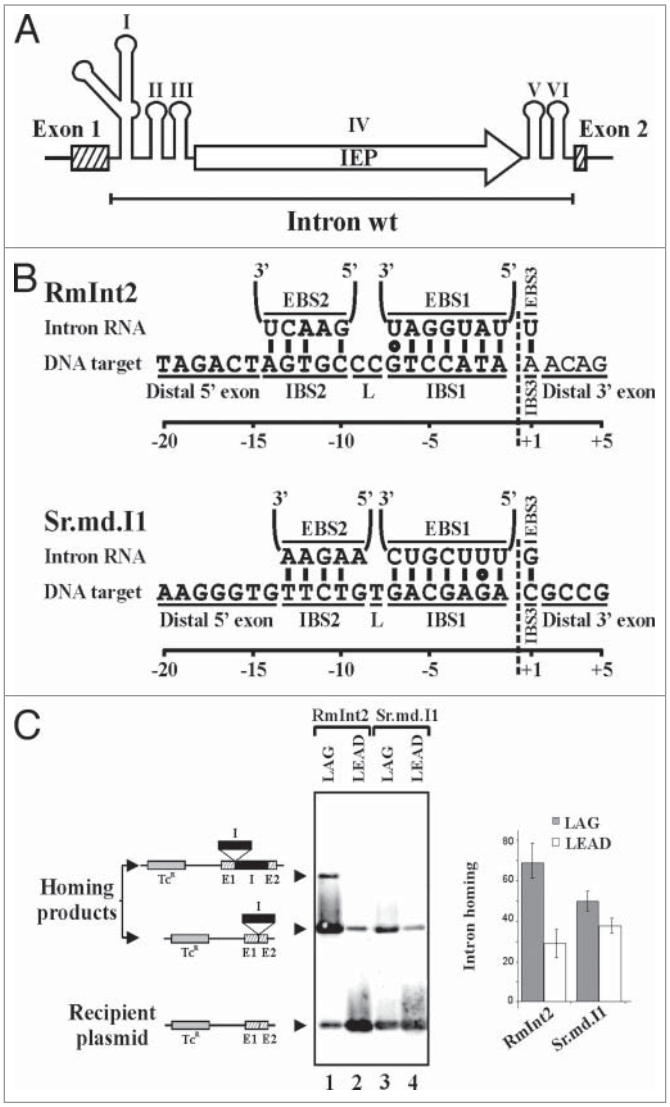

Figure 2. Retrohoming of RmInt2 and Sr.md.I1. (**A**) Representation of the intron donor construct used in this assay. (**B**) Predicted base-pairing interactions between introns and their DNA target sites. The diagram shows a -20/+5 target site with the various sequence elements and the EBS-IBS interactions predicted for RmInt2 and Sr.md.I1. The vertical dashed line indicates the intron-insertion site. (**C**) Southern blot hybridizations of SalI-digested DNA from *S. meliloti* RMO17 transconjugants harboring donor and recipient plasmids. Recipient plasmids contain specific DNA targets for each intron in both orientations with respect to the replication fork, such that the reverse transcription of the inserted intron RNA could potentially use either a nascent leading (LEAD) or lagging (LAG) DNA strand at the replication fork as a primer. Membranes were probed with a DNA fragment specific for the recipient plasmids. The homing products and recipient plasmids detected are indicated to the left of the panel. A bar diagram shows homing efficiencies is presented in the panel and these efficiencies were calculated as indicated in the Methods.

Determining RmInt2 DNA target requirements

Determining the RmInt2 EBS2 sequence

The positioning of the EBS2, the EBS2/IBS2 mismatch of RmInt2 and the linker constituted by two nucleotides raised questions about whether the EBS2 sequence of this intron had been correctly identified. We defined the EBS2-IBS2 interaction, by carrying out selection experiments (see Materials and Methods) with the two-plasmid retrohoming assay previously described for RmInt1.[30] Briefly, as donor constructs pKGE-MA4R2T7 was used, in which the IEP was encoded by a sequence upstream from the ΔORF-RNA, in which a large portion of the IEP ORF within DIV had been replaced with the T7 promoter. A plasmid library in pACELAG was used as the recipient construct. In this plasmid, natural target sites with randomized nucleotides from positions –16 to –8 (**Fig. 3A**) were inserted upstream from a promoter-less *tet*^{R} gene. The randomized positions encompassed five predicted IBS2 residues plus two 5′ and 3′ adjacent nucleotides. The recipient plasmid library and donor plasmid were introduced into *E. coli* HMS174 (DE3). After selection, target sites with intron insertions from 52 colonies grown on LB agar containing ampicillin and tetracycline were sequenced, together with 54 recipient plasmids from the initial pool (**Fig. S3A**). **Figure 3A** shows nucleotide frequencies at each randomized position in the selected target sites, already corrected for biases in the initial pool. The WebLogo representation shows that positions –9 and –8 display no selection for any specific nucleotide, consistent with their role as the linker between IBS1 and IBS2, and contrasting with the single-nucleotide linker of the RmInt1 target site. The rest of randomized positions (–16 to –10) displayed selection for specific nucleotides: 5′-CTAGTTC-3′. In the predicted structure of the intron RNA stem-loop containing the EBS2, the nucleotide sequence displaying the best pairing to the selected sequence was 5′-GAACU-3′. The first two nucleotides of the selected sequence, CT, would remain unpaired, indicating that these nucleotides do not belong to the IBS2.

Only 50% of the inserted introns had the five nucleotide residues of EBS2 base-paired with the IBS2, and 38% had four nucleotides of the EBS2 paired (**Fig. 3B**). Individually, position -11 had the lowest percentage of pairing, at 70% (**Fig. 3C**), with base-pairing rates for the other positions varying between 82% and 92%. This ability of RmInt2 to insert into targets with the pairing of just four nucleotide residues in the EBS2/IBS2 interaction accounts for the insertion of this intron into its natural targets with position -11 unpaired (**Fig. 2**). These data suggest that RmInt2 has a lower specificity for target selection than RmInt1, for which almost 86% full base-pairing between EBS2 and IBS2 is observed in insertion events, with base-pairing rates for the individual residues between EBS2 and IBS2 positions of 89% to 100%.[30]

Essential nucleotides in the distal 5′-exon and 3′-exon of the RmInt2 target site

We identified two nucleotides belonging to the distal 5′-exon (see above), but it was considered likely that other positions in this exon and in the distal 3′-exon would be required for the efficient retrohoming of RmInt2.[27] We sought to identify these critical positions, by carrying out selection experiments in which the recipient plasmid library contained cloned targets with randomized positions from -24 to -15 and +2 to +9. The IBS2 of the

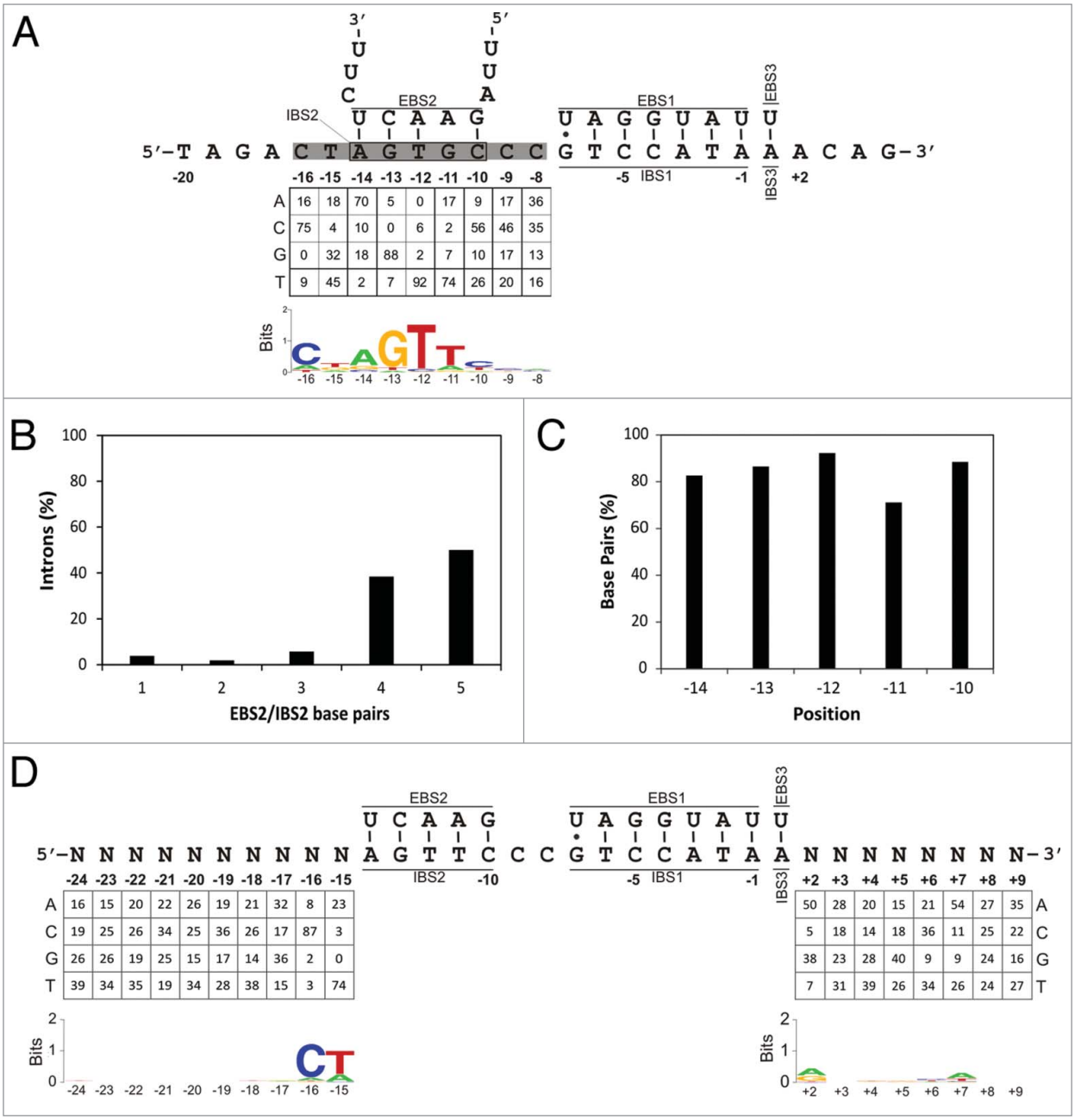

Figure 3. RmInt2 DNA target site requirements. (**A**) Panel shows the target used to define the RmInt2 EBS2, with randomized nucleotides highlighted in gray. Nucleotide frequencies at each randomized position of the selected plasmids, corrected for biases in the initial pool, are shown below and represented as a WebLogo. (**B**) Percentage of introns with different numbers of EBS2/IBS2 base pairs formed with the target site in the selected invasion events. (**C**) Percentage of base pairs formed at each EBS2/IBS2 position in the selected invasion events. (**D**) Identification of critical nucleotide residues in the distal exon regions of the RmInt2 DNA target. Randomized positions are represented as N. Nucleotide frequencies at each randomized position of the intron-inserted targets, corrected for biases in the initial pool, are shown. The WebLogo representation illustrates the corrected nucleotide frequencies.

target sites was 5′-AGTTC-3′, corresponding to the sequence emerging from the selection experiments described above. Sequences of 107 active target sites and 100 plasmids from the initial pool of recipient plasmids were used for correction for nucleotide frequency biases (**Fig. S3B**). **Figure 3D** shows the frequencies for each randomized position along with a WebLogo representation of the same data. The distal 5′-exon contained only two nucleotides selected in the active target sites, C-16 and T-15, which were already selected during definition of the EBS2 of RmInt2. These nucleotides were present in 87% and 74% of the selected targets. The distal 3′-exon contained two conserved nucleotides: A+2 and A+7 (for which the degree of conservation was lower). Therefore, by contrast to our findings for RmInt1, the minimal target site of RmInt2 was found to extend

from position-16 in the 5′-exon to position +7 in the 3′-exon and the retrohoming efficiency was only a tenth that for RmInt1 (**Table S1**).

Splicing and mobility of different combinations of RmInt1-related intron RNAs and IEPs

Group II intron RNAs and ORFs have mostly coevolved.[46] The interaction of the IEP with the intron RNA, primarily in the DIVa subdomain, with additional contacts in the catalytic core, is therefore critical for both RNA splicing and intron mobility.[47] Moreover, IEPs recognize particular sequences at the distal 5′- and 3′-exons of the DNA target sequence and they seem to have evolved to recognize different target sites.[8] The specificity of the co-operation of IEP and RNA components from full-length related mobile introns has not been investigated in bacteria. We investigated whether interactions with RmInt1-related elements were specific, by using constructs encoding the IEP upstream from the ΔORF-RNAas intron donors,[33] thereby facilitating the combination of intron RNAs and IEPs in splicing and mobility assays. Mobility assays with donor plasmids expressing the ΔORF-RNA construct with the corresponding IEP (**Fig. 4A**, lanes 1, 5, and 9) showed the retrohoming efficiencies of introns RmInt2 (69%) and Sr.md.I1 (46%) to be lower than that of RmInt1 (86%), probably due to the suboptimal natural target sites of the new introns. Interestingly, combinations of different IEPs and intron RNAs (ΔORFs) displayed different patterns of specificity. The Sr.md.I1 IEP significantly promoted RmInt1 mobility, which reached an efficiency of 49%, similar to that of the canonical $S1_{IEP}/S1_{\Delta ORF}$ (Sr.md.I1) combination (lanes 5 and 8), but different from that of RmInt2 (lane 2). The RmInt2 and RmInt1 IEPs displayed a higher degree of specificity in mobility, as little or no mobility was observed for non-native combinations (lane 1 vs. 4 and 7 for RmInt2 IEP and lane 9 vs. 3 and 6 for RmInt1 IEP). Mobility may be impaired due to the inability of these IEPs to promote heterologous intron RNA splicing. We assessed the degree of intron excision for every combination, by carrying out primer extension analyses on total RNA with a common primer complementary to the 5′ ends of all intron RNAs (**Fig. 4B**). In all combinations, a major band corresponding to the excised intron was detected (105 nt for RmInt2 and 104 nt for the RmInt1 and Sr.md.I1 introns). The highest level of excision was observed with RmInt1, for which excision levels were 3 to 10 times higher than for the other two introns. The different combinations differed in their degree of specificity. The RmInt1 and Sr.md.I1 IEPs displayed the highest levels of excision with their canonical RNAs (lane 5 vs. lanes 4 and 6; lane 9 vs. lanes 7 and 8), but a different pattern was observed for RmInt2. The level of excision of RmInt2 (ΔORF) by heterologous IEPs from either Sr.md.I1 or RmInt1 was two to three times higher than that obtained with the canonical $R2_{IEP}/R2_{\Delta ORF}$ combination (lanes 2 and 3 vs. lane 1). This inhibition of RmInt2 excision by its own IEP do not occur at transcriptional level since similar level of RNA precursors are detected in all three combinations (data not shown) and it seems unlikely that the RNA preparation used selectively discarded the RNPs formed between the RmInt2 intron and its IEP. This result suggests that putative differences

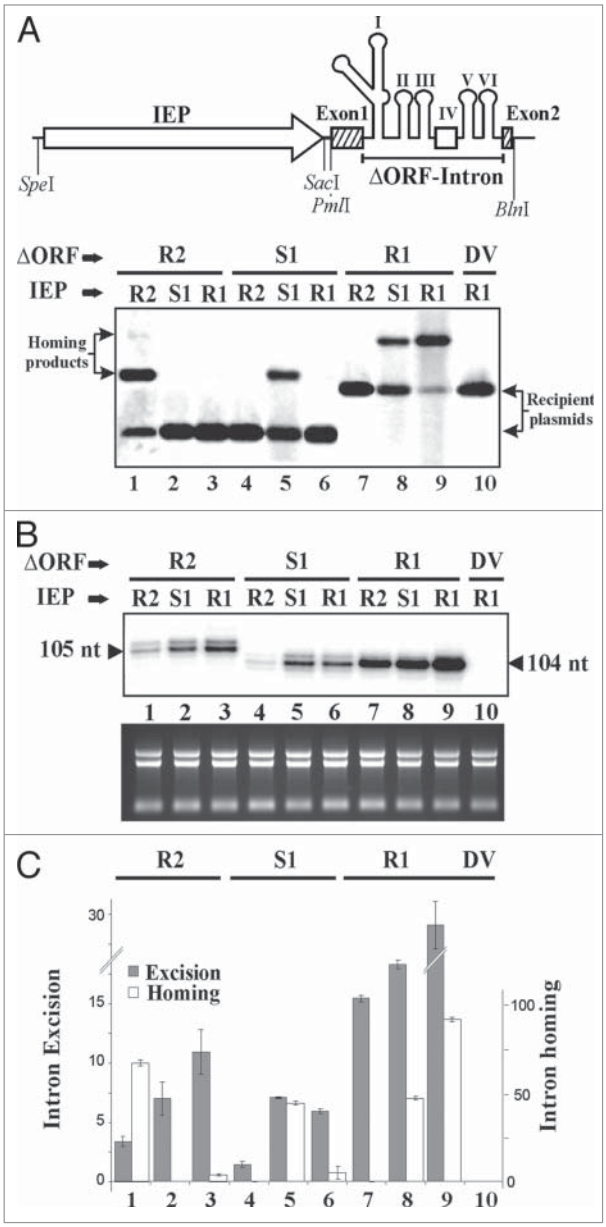

Figure 4. Excision and retrohoming of intron-donor constructs expressing different combinations of IEPs and ribozymes. (**A**) Schematic diagram of the intron donor construct used in these assays, in which the intron IEP sequence is located upstream from the intron ribozyme (ΔORF) sequence, and retrohoming assays. Southern blot hybridizations of SalI-digested DNA from *S. meliloti* RMO17 transconjugants harboring the indicated intron donor and recipient plasmids, in which the specific DNA target was inserted in the LAG orientation. The membranes were probed with a DNA fragment specific for the recipient plasmids. (**B**) Intron excision assays. Primer extension assays were performed on total RNA extracted from RMO17 cells carrying the donor construct. The IEP and ΔORF are indicated. RNA was reverse-transcribed with a ^{32}P [γ-ATP] 5′-labeled primer complementary to the first 104 nt of RmInt1 and Sr.md.I1, and to the first 105 nt of RmInt2. As a negative control (lanes 10), we used a splicing-defective RmInt1 dV mutant intron construct.[57] (**C**) Excision and homing efficiencies, calculated as indicated in the methods, are shown in gray and white, respectively. Data are the means of determinations in at least three independent assays, with the corresponding standard error indicated by thin lines.

in the binding affinity of the heterologous IEPs to the RmInt2 intron RNA could favor an increase in intron RNA splicing. Surprisingly, these combinations abolished intron mobility, suggesting that either the positioning of the heterologous IEPs for the initiation of cDNA synthesis[48] or DNA target recognition by the IEPs inhibited RmInt2 mobility.

Role of RmInt1-related IEPs in DNA target recognition

For group II introns which have IEPs containing D and En domains, mutations in the distal 5′-exon inhibit both reverse splicing and second-strand cleavage, whereas mutations in the 3′-exon inhibit second-strand cleavage only.[8] The DNA target site requirements for the homing of RmInt1 define distal 5′ and 3′ exon regions displaying no predictable base-pairing with the intron RNA. This has led to suggestions that nucleotides within these regions are recognized by the IEP.[27] Despite the high level of splicing of the $R2_{\Delta ORF}$-$R1_{IEP}$ combination, a lack of insertion of the RmInt2 intron into its DNA target raised questions about the possible impairment of DNA target recognition by the RmInt1 IEP. We analyzed this aspect, by generating a chimaeric DNA target construct (target –20/+5) in which the distal part of the target (–20 to −15 and +2 to +5) corresponds to the RmInt1 target site (IS*Rm2011–2* DNA target), with the IBS1, 2 and 3 (–14 to +1) corresponding to the RmInt2 target site (IS*Rm17* DNA target) (**Fig. 5A**). A new target (target –21/+5) was generated with an extra T at position –16, to increase the distance between the EBS2/IBS2 pairing and the rest of the distal 5′-exon (as occurs in the canonical RmInt1 target site). Homing assays with both chimaeric targets (**Fig. 5B**) revealed that the donor construct $R2_{\Delta ORF}$-$R1_{IEP}$ retrohomed to the target if the distal sites were modified in line with the sequence requirements of the RmInt1 IEP. Intron insertion at the predicted site was confirmed by sequencing the retrohoming product (data not shown). Interestingly, the –21/+5 DNA target with the extra T-16 displayed almost no detectable invasion, indicating that the part of the distal 5′−exon recognized by the IEP was also physically constrained by the distance to the intron insertion site, rather than by the positioning of the EBS2 of the intron RNA at the target site.

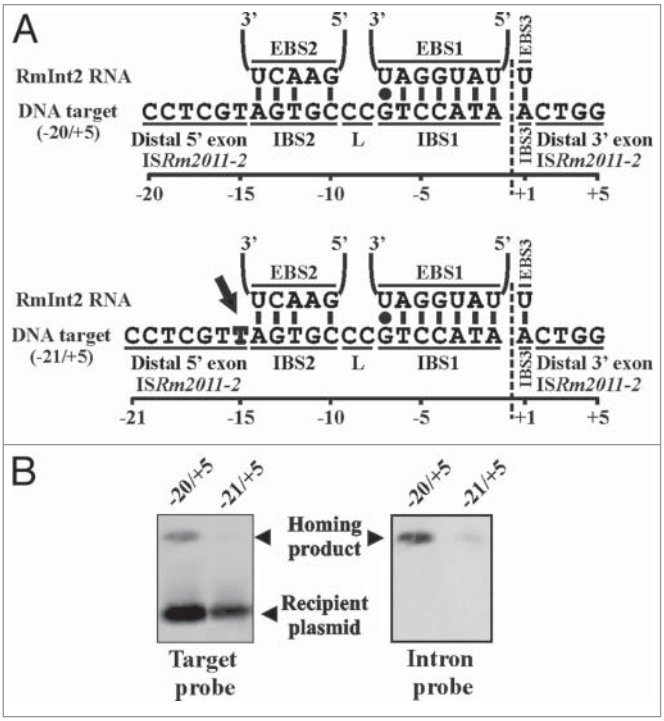

Figure 5. Retrohoming on chimaeric DNA targets. (**A**) Base-pairing interaction between RmInt2 and two chimaeric targets consisting of the IBS region (-14 to +1) from the IR of the IS*Rm17* and the distal exons from IS*Rm2011–2*. (-21/20 to -15 and +2 to +5). The vertical dashed line indicates the intron-insertion site. (**B**) Representative Southern-blot hybridizations of SalI-digested DNA from *S. meliloti* RMO17 transconjugants harboring the $R1_{IEP}$-$R2_{\Delta ORF}$ intron donor and recipient plasmids containing the -20/+5 and -21/+5 DNA targets in the LAG orientation. Membranes were hybridized with a DNA probe specific for the recipient plasmids (target probe) and the intron (intron probe).

Discussion

We obtained evidence that two RmInt1-related introns, RmInt2 from *S. meliloti* strain GR4 and Sr.md.I1 from the genome of *S. medicae,* were functional introns able to splice and to retrohome into their DNA target sites. The critical residues in the distal 5′- and 3′-exon regions of the RmInt2 DNA target sequence and EBS2-IBS2 pairing were identified by selection experiments. RmInt2, and probably Sr.md.I1, displayed a lower specificity for target site selection than RmInt1, allowing mispairing in the EBS2/IBS2 interaction. Splicing and retrohoming assays combining intron RNA and IEP components from these introns indicated that complementation between these introns was possible for splicing and, in some cases, even for intron mobility. The IEP recognizes distal parts of the DNA target site, and distal 5′-exon region recognition seems to play a particularly important role, being constrained by essential nucleotide positions, but also by the distance to the intron insertion site.

Bacterial groups II introns home preferentially to sites outside of functional genes, generally within intergenic regions,[18] or to other mobile genetic elements controlling their spread,[49] possibly following a selection-driven extinction model predicting the elimination of highly colonised genomes from the population by purifying selection.[50] It has also been suggested that group II introns may have influenced bacterial evolution, with some intron fragments remaining and continuing to evolve in the genome as functional regulatory elements.[37] RmInt1, RmInt2 and Sr.md.I1 are closely related elements, but they probably represent different intron evolutionary scenarios in extant *S. meliloti* and *S. medicae* species. RmInt1 is present in 90% of *S. meliloti* isolates[24,31] and may currently be in the process of colonising these bacterial species.[51] Less information is available for Sr.md. I1, but hybridization data suggest that this intron may also be spreading in *S. medicae*.[36] By contrast, RmInt2 seems to be a more recent acquisition, because all copies harbored by strain GR4 were found to be identical and no copies of this retroelement have been reported in the complete genome sequences of

other *S. meliloti* strains. All these introns are mobile and display a bias toward insertion into the strand serving as the template for the nascent lagging strand in DNA replication forks, suggesting a similar mechanism of spread within the bacterial genome.

In *S. meliloti* GR4, the two related introns RmInt1 and RmInt2 successfully coexist as 17 full-length intron copies. We found that the IEPs of RmInt1, RmInt2 and Sr.md.I1 were potentially interchangeable, to some extent, promoting intron RNA splicing, and that the Sr.md.I1 IEP could support RmInt1 intron homing. Previous reports have indicated that the IEP (LtrA) of the *Lactococcus lactis* Ll.ltrB intron initially binds to subdomain DIVa, subsequently making weaker contacts with conserved catalytic core regions, to stabilize the active intron RNA structure for RNA splicing.[47,52] Residual RNA splicing occurs in RNA constructs where DIVa is deleted indicating a direct binding of the IEP to the catalytic core.[47,52,53] The deletion of DIVa from the yeast intron aI2 inhibits splicing, but the level of residual splicing is considerably higher than for Ll.ltrB.[54] As expected, given the co-evolution of intron RNA and IEP components, most of the different combinations of heterologous RmInt1-related IEPs and intron RNAs resulted in splicing impairment, potentially due to the lower affinity of the IEPs for the heterologous intron RNAs (DIVa and/or catalytic core). Surprisingly, the IEPs from RmInt1 and Sr.md.I1 promoted higher levels of RmInt2 intron RNA splicing than the IEP of RmInt2. These results suggest that IEP-RNA interactions evolved in a different manner and reveal not only an interesting mechanism of downregulation of RmInt2 at the splicing level (that could be used by other group II introns) but that the interactions between the RmInt2 intron and its IEP can undergo conformational rearrangements to ensure a very efficient DNA targetting step. Moreover, these results may be of biological significance, because tighter control over the splicing reaction would result in stronger control over host insertion sequence mobility.

For the Ll.ltrB and aI2 introns, the deletion of DIVa almost entirely abolishes intron mobility, which is also inhibited by stronger or weaker IEP-DIVa interactions.[48] It is therefore unsurprising that most of the heterologous IEP-RNA combinations were associated with an abolition of intron mobility. However, the Sr.md.I1 IEP promoted the efficient mobility of RmInt1. In this case, mobility efficiency was very similar to that of wild-type RmInt1, suggesting similarity between DNA-target recognition by the Sr.md.I1 IEP and that mediated by the RmInt1 IEP. Detailed examination of the natural target of Sr.md.I1 revealed that the critical nucleotides T-15 and G-16 in the distal 5′-exon of the RmInt1 DNA target site were conserved. By contrast, the G+4 nucleotide of the distal 3′-exon was not conserved.[27,30] Thus, the Sr.md.I1 IEP probably has more relaxed recognition rules at the 3′ exon than the RmInt1 IEP, because the RmInt1 IEP was unable to support the mobility of Sr.md.I1.

RmInt2 provides an illustration of the versatility of group II introns in terms of survival and spread in the bacterial genome. The use of the inverted repeats of insertion sites for targeting doubles the number of target sites and ensures that at least one such site is on the lagging strand. This intron also displays a partial relaxation of target site specificity, and the capacity of RmInt2 to insert into targets with the pairing of just four nucleotide residues in the EBS2/IBS2 interaction accounts for the efficient insertion of wild-type RmInt2 into its natural target (seven intron copies inserted at DNA targets with position –11 unpaired). Moreover, the U-A EBS3/IBS3 pairing increases the possibility of intron insertion by generating head-to-tail intron copies. Related introns within the same host have thus adopted different and opposite strategies for successful dispersal and survival in the bacterial genome. RmInt1 displays high target-site specificity and retrohoming efficiency with abundant DNA targets in *S. meliloti* genomes (IS*Rm2011–2*). By contrast, RmInt2 displays relaxed specificity and reduced retrohoming efficiency, and has increased the probability of survival and spread in the genome by using IS inverted repeats as DNA targets.

The ability of the $RmInt2_{\Delta ORF}$/$RmInt1_{IEP}$ combination to invade a chimaeric RmInt2 target, in which the distal parts of the 5′ and 3′ exon flanking sequences are those of the RmInt1 target, indicates that these regions are recognized by the IEP. Interestingly, the recognition of the distal 5′ exon is not constrained by the distance to EBS2/IBS2 pairing, suggesting that the IEP is tightly bound to the intron RNA in the RNP particles and is entirely unable to distort the DNA target site, consistent with the lack of D and En regions in class D group II intron ORFs. The closely related IEPs of RmInt1, RmInt2, and Sr.md.I1 are similar in size, at 419 amino acids, but the minimal RmInt2 target (–16 to +7) is shifted toward the 3′ exon with respect to that of RmInt1 (–20 to +5). Thus, changes in the intron RNA and IEP of closely related elements may lead to the selection of new targets, a process to which the IEP makes a significant contribution. In this sense, meanwhile the maturase domain and the C-tail accumulate most such changes between these IEP's, they could play an important role not only in splicing but also in mobility.[44]

The RmInt1 IEP can act in trans[33] promoting intron RNA splicing and retrohoming activity. Our findings therefore raise questions about the possible contribution of such co-operation between IEPs and intron RNAs to the survival and spread of closely related introns in bacterial genomes, rendering them less prone to extinction. Our results provide new insight into the versatility of group II introns in terms of their ability to colonise bacterial genomes through conservative changes to the intron RNA and IEP and possible co-operation between these components from different related introns. These properties may be useful for further structural studies and biotechnological applications.

Materials and Methods

Bacterial strains and growth conditions

S. meliloti GR4 and RMO17 and *S. medicae* WSM419 strains were grown at 28°C on complete TY or defined minimal medium (MM). *E. coli* DH5α, used for the cloning and maintenance of plasmid constructs, was grown at 37 °C on LB medium. Antibiotics were added to the medium when required, at the following concentrations: kanamycin, 200 μg/ml for rhizobia and 50 μg/ml for *E. coli*; tetracycline, 10 μg/ml; ampicillin, 200 μg/ml.

Oligonucleotides

The oligonucleotides used in this study are listed in **Table S2**.

Construction of intron donor and recipient target site plasmids

The RmInt2 intron donor consisted of a 1.9-kb DNA fragment containing the wild-type RmInt2 intron and short flanking exons (–18/+5) generated by the amplification, with the AccuPrime™ high-fidelity *Taq* DNA polymerase (Invitrogen), of total DNA from *S. meliloti* GR4, with the primers 1F_R2_IS17 and 1R_R2_IS17. The purified DNA fragment was inserted into pGEM-T Easy Vector (Promega) to generate pGEMtR2. Similarly, the Sr.md.I1 intron donor consisted of a 1.9-kb DNA fragment containing the wild-type Sr.md.I1 intron flanked by short exons (–20/+5), generated by amplifying total DNA from *S. medicae* WSM419 with the primers 1F_S1_IS66 and 1R_S1_IS66. The resulting 1.9-kb DNA fragment was then inserted into the pGEM-T Easy Vector, to generate pGEMtS1. Both DNA inserts were completely sequenced, to ensure that no adventitious mutations had been introduced. Subsequently, a BamHI DNA fragment containing either the RmInt2 or the Sr.md.I1 intron from pGEMtR2 or pGEMtS1, was inserted in pKG0[45] to obtain the donor plasmids, pKGRmInt2 and pKGSr.md.I1, respectively.

ΔORF intron donors and chimaeras were also constructed. ΔORF derivatives had an internal deletion of intron domain IV, and resembled the pKGEMA4 construct described for RmInt1 (pKGIEPΔORF20/5).[33] They were generated by inverse PCR, using the above mentioned pGEMt derivative plasmids, pGEMtR2 and pGEMtS1, as a template and the divergent primers dORF1_R2 and dORF2_R2 for deletion of RmInt2 intron nucleotide positions 615–1763 and the primers dORF1_S1 and dORF2_S1 primers for deletion of the Sr.md.I1 intron nucleotide positions 611–1761. The 3.7 kb PCR-amplified fragments were subjected to XhoI digested and self-ligation to generate the pGEMtΔORFR2 and pGEMtΔORFS1 plasmids, respectively. Intron deletions were confirmed by DNA sequencing. The pGEMtΔORFR2 plasmid contains an extra "CGGGCCGCTC GAG" sequence at its engineered XhoI site. An intact IEP ORF for each intron was engineered by PCR with either the IEP_SpeI_R2 and IEP_SacI_R2 primers for the pGEMtR2 template, or the IEP_SpeI_S1 and IEP_SacI_S1 primers for the pGEMtS1 template, in each case generating a 1.3 kb DNA fragment with SpeI-SacI restriction ends.

Final intron donors and chimaeras were constructed with pKGEMA4 as a vector cassette, in which the IEP ORFs and ΔORF intron DNA segments were replaced via the engineered SpeI and SacI single cloning sites for the various IEPs, and the BlnI and PmlI sites for the various ΔORF introns. Thus, the canonical RNA and IEP combinations, pKGEMA4RmInt2 ($R2_{\Delta ORF}$-$R2_{IEP}$) and pKGEMA4Sr.md.I1 ($S1_{\Delta ORF}$-$S1_{IEP}$), were generated, together with non-canonical combinations, resulting in the assay of nine types of donor constructs. The pKG4DV plasmid, a mutant derivative of pKGEMA4, in which the catalytic triad GTT is replaced with CGA in domain V of the RmInt1 ribozyme,[55] was used as a negative control in intron excision and mobility assays.

The intron recipient plasmids were pJBTc19 derivatives[56] in which the target site was inserted in both orientations with respect to the direction of the replication fork. The corresponding plasmids for Sr.md.I1 (pJBISSme3LEAD and pJBISSme3LAG), which resembled the recipient plasmids for RmInt1 (pJB0.6LAG and pJB0.6LEAD),[32] were generated by inserting a 44 nt PCR fragment into the NotI site, in both orientations. This fragment extended from positions –22 to +22 nt of the intron insertion site (containing the minimal target site –20/+5) and was obtained by PCR with the appropriate combination of primers, on total DNA from *S. medicae* WSM419. Recipient plasmids for RmInt2 (pJBISRm17LEAD and pJBISRm17LAG), and chimaeric target site plasmids (pJBChim20/5 and pJBChim21/5) were generated by annealing phosphorylated oligonucleotides. The ISRm17_F/ISRm17_R, Chimera-20/+5_F/Chimera-20/+5_R, and Chimera-21/+5_F/ Chimera-21/+5_R primer pairs were hybridized such that the double-stranded molecules were flanked by protuberant NotI restriction sites. Further ligation to a NotI-digested and dephosphorylated pJBTc19 vector resulted in the definitive recipient plasmid. The correct orientation of the insert was confirmed by digestion and sequencing.

Functional intron assays

In vivo intron excision was assessed by primer extension analysis. Primer extension reactions were performed as previously described,[57] with slight modifications. Total RNA (15 μg) extracted from free-living *S. meliloti* RMO17 cells containing the various intron donor plasmids (wild-type, ΔORF-IEP derivatives, and chimaeric constructs) were reverse-transcribed with the AMV reverse transcriptase (Roche Diagnostics GmbH) and a radioactively labeled 104–5mer oligonucleotide. Samples were then resolved by electrophoresis in a denaturing 6% polyacrylamide gel. The corresponding excision band (104 nt for RmInt1 and Sr.md.I1 introns, and 105 nt for RmInt2) was quantified with the Quantity One software package (BioRad Laboratories). The results were normalized (100 × [band signal intensity/addition of overall band intensities]) and plotted.

Intron mobility was determined by the double-plasmid assay, in *S. meliloti* RMO17.[32] *S. meliloti* RMO17 was transformed with the appropriate combination of donor and recipient plasmids, by successive triparental conjugation. Plasmid pools were isolated from four exponentially growing cultures in TY and were digested with SalI. The digested plasmids were resolved by electrophoresis in a 0.8% agarose gel and vacuum-blotted onto a nylon membrane, according to the manufacturer's instructions (Pall Corporation). DNA hybridization was performed under high-stringency conditions, with DIG-labeled probes generated by PCR with the Probe-TetA-F and Probe-TetA-R primers specific for the tetracycline resistance gene of the recipient plasmids or with the Probe-R2-F and Probe-R2-R primers specific for the RmInt2 group II intron ribozyme. Hybridizing bands corresponding to homing products and recipient plasmids were analyzed with

Quantity One software (Bio-Rad Laboratories) and invasion efficiency was calculated as 100 × [Homing product/(Homing product +Recipient plasmid)].[45]

Experimental design to determine the DNA target site requirements of RmInt2

For the construction of an intron donor plasmid harboring the phage T7 promoter within intron domain IV, we used pKGEMA4RmInt2 as a template for site-specific insertion by two-step PCR. The first PCR was performed with the primers AgeI and R2_T7F, whereas the second reaction involved primers Probe_R2F and R2_T7R. A T7 promoter sequence was attached to the end of the PCR fragment with primers R2_T7F and R2_T7R. These PCR products were purified on Ilustra™ Microspin™ S-300 HR columns, mixed and amplified with the external primers AgeI and Probe_R2F. Once generated, the amplicon was sequenced, digested with enzymes cleaving at two internal sites (AvrII and SacII) and inserted into pKGEMA4RmInt2 digested with the same enzymes, to create pKGEMA4R2T7.

Randomized target sites were inserted into the recipient plasmid library used to define the EBS2 of RmInt2 by annealing primers DInt2–5′N and DInt2–3′ and filling in the gaps with the Klenow fragment of DNA polymerase I. The fragments were then inserted between the PstI and XbaI sites of pACELAG.[34] The recipient plasmid library used to identify critical nucleotides in the distal parts of both exons was obtained by inserting targets obtained by annealing the D5′3'ExR2 and DInt2–3′ primers between the PstI and XbaI sites of pACELAG and filling in the gaps with the Klenow fragment of DNA polymerase I. *E. coli* DH10B cells were then electroporated with the libraries and plated on LB plus ampicillin and incubated overnight at 37°C. The number of clones integrating into each library was found to be 6.09×10^4 for the EBS2 library and 9.6×10^4 for the distal exons library.

Selection experiments were performed by electroporating HMS174 (DE3) cells harboring pKGEMA4R2T7 with the appropriate library, and transformants were incubated for 1 h at 28°C in SOC medium. Cells were then diluted 1/200 in LB medium containing kanamycin plus ampicillin and grown overnight at 28°C. They were then serially diluted and plated on LB agar containing tetracycline and ampicillin or ampicillin only and incubated at 28°C. The colonies grown on these media were analyzed by colony PCR and sequencing. "Pool" or "Selected" (**Fig. S3**) indicates that the data were obtained from colonies grown on medium containing ampicillin or ampicillin plus tetracycline, respectively. The nucleotide frequency at each of the randomized positions was corrected for biases in the initial pool, as previously described.[30] Normalized frequencies were used to generate a sample set of 100 active target sites displayed in WebLogo format.[58]

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Funding

This work was supported by research grants CSD 2009–0006 of the Consolider-Ingenio program including ERDF (European Regional Development Funds), and BIO2011–24401 from the Spanish *Ministerio de Ciencia e Innovación*, currently the *Ministerio de Economía y Competitividad*. LRM was supported by predoctoral fellowships (Program JAE-Predoc from Consejo Superior de Investigaciones Científicas).

Supplemental Materials

Supplemental data for this article can be accessed on the publisher's website.

References

1. Sharp PA. On the origin of RNA splicing and introns. Cell 1985; 42:397-400; PMID:2411416; http://dx.doi.org/10.1016/0092-8674(85)90092-3
2. Cech TR. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. Cell 1986; 44:207-10; PMID:2417724; http://dx.doi.org/10.1016/0092-8674(86)90751-8
3. Cavalier-Smith T. Intron phylogeny: a new hypothesis. Trends Genet 1991; 7:145-8; PMID:2068786; http://dx.doi.org/10.1016/0168-9525(91)90102-V
4. Eickbush TH. Origin and evolutionary relationships of retroelements. The evolutionary biology of viruses. Raven Press, New York 1994.
5. Michel F, Ferat JL. Structure and activities of group II introns. Annu Rev Biochem 1995; 64:435-61; PMID:7574489; http://dx.doi.org/10.1146/annurev.bi.64.070195.002251
6. Dai L, Zimmerly S. ORF-less and reverse-transcriptase-encoding group II introns in archaebacteria, with a pattern of homing into related group II intron ORFs. RNA 2003; 9:14-9; PMID:12554871; http://dx.doi.org/10.1261/rna.2126203
7. Toro N. Bacteria and Archaea Group II introns: additional mobile genetic elements in the environment. Environ Microbiol 2003; 5:143-51; PMID:12588294; http://dx.doi.org/10.1046/j.1462-2920.2003.00398.x
8. Lambowitz AM, Zimmerly S. Mobile group II introns. Annu Rev Genet 2004; 38:1-35; PMID:15568970; http://dx.doi.org/10.1146/annurev.genet.38.072902.091600
9. Toro N, Jiménez-Zurdo JI, García-Rodríguez FM. Bacterial group II introns: not just splicing. FEMS Microbiol Rev 2007; 31:342-58; PMID:17374133; http://dx.doi.org/10.1111/j.1574-6976.2007.00068.x
10. Mohr G, Ghanem E, Lambowitz AM. Mechanisms used for genomic proliferation by thermophilic group II introns. PLoS Biol 2010; 8:e1000391; PMID:20543989; http://dx.doi.org/10.1371/journal.pbio.1000391
11. Michel F, Costa M, Westhof E. The ribozyme core of group II introns: a structure in want of partners. Trends Biochem Sci 2009; 34:189-99; PMID:19299141; http://dx.doi.org/10.1016/j.tibs.2008.12.007
12. Mohr G, Perlman PS, Lambowitz AM. Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. Nucleic Acids Res 1993; 21:4991-7; PMID:8255751; http://dx.doi.org/10.1093/nar/21.22.4991
13. San Filippo J, Lambowitz AM. Characterization of the C-terminal DNA-binding/DNA endonuclease region of a group II intron-encoded protein. J Mol Biol 2002; 324:933-51; PMID:12470950; http://dx.doi.org/10.1016/S0022-2836(02)01147-6
14. Lambowitz AM, Zimmerly S. Group II introns: mobile ribozymes that invade DNA. Cold Spring Harb Perspect Biol 2011; 3:a003616; PMID:20463000; http://dx.doi.org/10.1101/cshperspect.a003616
15. Toro N, Martínez-Abarca F. Comprehensive phylogenetic analysis of bacterial group II intron-encoded ORFs lacking the DNA endonuclease domain reveals new varieties. PLoS One 2013; 8:e55102; PMID:23355907; http://dx.doi.org/10.1371/journal.pone.0055102
16. Zimmerly S, Hausner G, Wu Xc. Phylogenetic relationships among group II intron ORFs. Nucleic Acids Res 2001; 29:1238-50; PMID:11222775; http://dx.doi.org/10.1093/nar/29.5.1238
17. Toro N, Molina-Sánchez MD, Fernández-López M. Identification and characterization of bacterial class E group II introns. Gene 2002; 299:245-50; PMID:12459272; http://dx.doi.org/10.1016/S0378-1119(02)01079-X
18. Simon DM, Clarke NAC, McNeil BA, Johnson I, Pantuso D, Dai L, Chai D, Zimmerly S. Group II introns in eubacteria and archaea: ORF-less introns and new varieties. RNA 2008; 14:1704-13; PMID:18676618; http://dx.doi.org/10.1261/rna.1056108
19. Candales MA, Duong A, Hood KS, Li T, Neufeld RAE, Sun R, McNeil BA, Wu L, Jarding AM, Zimmerly S. Database for bacterial group II introns. Nucleic Acids Res 2012; 40:D187-90; PMID:22080509; http://dx.doi.org/10.1093/nar/gkr1043
20. Martínez-Abarca F, Toro N. RecA-independent ectopic transposition in vivo of a bacterial group II

intron. Nucleic Acids Res 2000; 28:4397-402; PMID: 11058141; http://dx.doi.org/10.1093/nar/28.21.4397
21. Smith D, Zhong J, Matsuura M, Lambowitz AM, Belfort M. Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming. Genes Dev 2005; 19:2477-87; PMID:16230535; http://dx.doi.org/10.1101/gad.1345105
22. Yang J, Mohr G, Perlman PS, Lambowitz AM. Group II intron mobility in yeast mitochondria: target DNA-primed reverse transcription activity of aI1 and reverse splicing into DNA transposition sites in vitro. J Mol Biol 1998; 282:505-23; PMID:9737919; http://dx.doi.org/10.1006/jmbi.1998.2029
23. Dickson L, Huang HR, Liu L, Matsuura M, Lambowitz AM, Perlman PS. Retrotransposition of a yeast group II intron occurs by reverse splicing directly into ectopic DNA sites. Proc Natl Acad Sci U S A 2001; 98:13207-12; PMID:11687644; http://dx.doi.org/10.1073/pnas.231494498
24. Muñoz E, Villadas PJ, Toro N. Ectopic transposition of a group II intron in natural bacterial populations. Mol Microbiol 2001; 41:645-52; PMID:11532132; http://dx.doi.org/10.1046/j.1365-2958.2001.02540.x
25. Ichiyanagi K, Beauregard A, Lawrence S, Smith D, Cousineau B, Belfort M. Retrotransposition of the Ll. LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. Mol Microbiol 2002; 46:1259-72; PMID:12453213; http://dx.doi.org/10.1046/j.1365-2958.2002.03226.x
26. Ichiyanagi K, Beauregard A, Belfort M. A bacterial group II intron favors retrotransposition into plasmid targets. Proc Natl Acad Sci U S A 2003; 100:15742-7; PMID:14673083; http://dx.doi.org/10.1073/pnas.2536659100
27. Jiménez-Zurdo JI, García-Rodríguez FM, Barrientos-Durán A, Toro N. DNA target site requirements for homing in vivo of a bacterial group II intron encoding a protein lacking the DNA endonuclease domain. J Mol Biol 2003; 326:413-23; PMID:12559910; http://dx.doi.org/10.1016/S0022-2836(02)01380-3
28. Zhong J, Karberg M, Lambowitz AM. Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. Nucleic Acids Res 2003; 31:1656-64; PMID:12626707; http://dx.doi.org/10.1093/nar/gkg248
29. Zhuang F, Karberg M, Perutka J, Lambowitz AM. EcI5, a group IIB intron with high retrohoming frequency: DNA target site recognition and use in gene targeting. RNA 2009; 15:432-49; PMID:19155322; http://dx.doi.org/10.1261/rna.1378909
30. García-Rodríguez FM, Hernández-Gutiérrez T, Díaz-Prado V, Toro N. Use of the computer-retargeted group II intron RmInt1 of *Sinorhizobium meliloti* for gene targeting. RNA Biol 2014; 11:391-401; PMID: 24646865; http://dx.doi.org/10.4161/rna.28373
31. Martínez-Abarca F, Zekri S, Toro N. Characterization and splicing in vivo of a *Sinorhizobium meliloti* group II intron associated with particular insertion sequences of the IS630-Tc1/IS3 retroposon superfamily. Mol Microbiol 1998; 28:1295-306; PMID:9680217; http://dx.doi.org/10.1046/j.1365-2958.1998.00894.x
32. Martínez-Abarca F, Barrientos-Durán A, Fernández-López M, Toro N. The RmInt1 group II intron has two different retrohoming pathways for mobility using predominantly the nascent lagging strand at DNA replication forks for priming. Nucleic Acids Res 2004; 32:2880-8; PMID:15155857; http://dx.doi.org/10.1093/nar/gkh616
33. Nisa-Martínez R, Jiménez-Zurdo JI, Martínez-Abarca F, Muñoz-Adelantado E, Toro N. Dispersion of the RmInt1 group II intron in the *Sinorhizobium meliloti* genome upon acquisition by conjugative transfer. Nucleic Acids Res 2007; 35:214-22; PMID: 17158161; http://dx.doi.org/10.1093/nar/gkl1072
34. García-Rodríguez FM, Barrientos-Durán A, Díaz-Prado V, Fernández-López M, Toro N. Use of RmInt1, a group IIB intron lacking the intron-encoded protein endonuclease domain, in gene targeting. Appl Environ Microbiol 2011; 77:854-61; PMID: 21115708; http://dx.doi.org/10.1128/AEM.02319-10
35. Enyeart PJ, Mohr G, Ellington AD, Lambowitz AM. Biotechnological applications of mobile group II introns and their reverse transcriptases: gene targeting, RNA-seq, and non-coding RNA analysis. Mob DNA 2014; 5:2; PMID:24410776; http://dx.doi.org/10.1186/1759-8753-5-2
36. Fernández-López M, Muñoz-Adelantado E, Gillis M, Willems A, Toro N. Dispersal and evolution of the *Sinorhizobium meliloti* group II RmInt1 intron in bacteria that interact with plants. Mol Biol Evol 2005; 22:1518-28; PMID:15814827; http://dx.doi.org/10.1093/molbev/msi144
37. Toro N, Martínez-Rodríguez L, Martínez-Abarca F. Insights into the history of a bacterial group II intron remnant from the genomes of the nitrogen-fixing symbionts *Sinorhizobium meliloti* and Sinorhizobium medicae. Heredity (Edinb) 2014; (Forthcoming); PMID:24736785; http://dx.doi.org/10.1038/hdy.2014.32
38. Martínez-Abarca F, Martínez-Rodríguez L, López-Contreras JA, Jiménez-Zurdo JI, Toro N. Complete genome sequence of the alfalfa symbiont Sinorhizobium/Ensifer meliloti strain GR4. Genome Announc 2013; 1:e00174-12; PMID:23409262; http://dx.doi.org/10.1128/genomeA.00174-12
39. Galibert F, Finan TM, Long SR, Pühler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, et al. The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science 2001; 293:668-72; PMID:11474104; http://dx.doi.org/10.1126/science.1060966
40. Galardini M, Mengoni A, Brilli M, Pini F, Fioravanti A, Lucas S, Lapidus A, Cheng JF, Goodwin L, Pitluck S, et al. Exploring the symbiotic pangenome of the nitrogen-fixing bacterium *Sinorhizobium meliloti*. BMC Genomics 2011; 12:235; PMID:21569405; http://dx.doi.org/10.1186/1471-2164-12-235
41. Schneiker-Bekel S, Wibberg D, Bekel T, Blom J, Linke B, Neuweger H, Stiens M, Vorhölter FJ, Weidner S, Goesmann A, et al. The complete genome sequence of the dominant *Sinorhizobium meliloti* field isolate SM11 extends the S. meliloti pan-genome. J Biotechnol 2011; 155:20-33; PMID:21396969; http://dx.doi.org/10.1016/j.jbiotec.2010.12.018
42. Weidner S, Baumgarth B, Göttfert M, Jaenicke S, Pühler A, Schneiker-Bekel S, Serrania J, Szczepanowski R, Becker A. Genome sequence of *Sinorhizobium meliloti* Rm41. Genome Announc 2013; 1:e00013-12; PMID:23405285; http://dx.doi.org/10.1128/genomeA.00013-12
43. Reeve W, Chain P, O'Hara G, Ardley J, Nandesena K, Bräu L, Tiwari R, Malfatti S, Kiss H, Lapidus A, et al. Complete genome sequence of the Medicago microsymbiont Ensifer (Sinorhizobium) medicae strain WSM419. Stand Genomic Sci 2010; 2:77-86; PMID: 21304680; http://dx.doi.org/10.4056/sigs.43526
44. Molina-Sánchez MD, Martínez-Abarca F, Toro N. Structural features in the C-terminal region of the *Sinorhizobium meliloti* RmInt1 group II intron-encoded protein contribute to its maturase and intron DNA-insertion function. FEBS J 2010; 277:244-54; PMID:19951359; http://dx.doi.org/10.1111/j.1742-4658.2009.07478.x
45. Martínez-Abarca F, García-Rodríguez FM, Toro N. Homing of a bacterial group II intron with an intron-encoded protein lacking a recognizable endonuclease domain. Mol Microbiol 2000; 35:1405-12; PMID:10760141; http://dx.doi.org/10.1046/j.1365-2958.2000.01804.x
46. Toor N, Hausner G, Zimmerly S. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. RNA 2001; 7:1142-52; PMID:11497432; http://dx.doi.org/10.1017/S1355838201010251
47. Wank H, SanFilippo J, Singh RN, Matsuura M, Lambowitz AM. A reverse transcriptase/maturase promotes splicing by binding at its own coding segment in a group II intron RNA. Mol Cell 1999; 4:239-50; PMID:10488339; http://dx.doi.org/10.1016/S1097-2765(00)80371-8
48. Watanabe K, Lambowitz AM. High-affinity binding site for a group II intron-encoded reverse transcriptase/maturase within a stem-loop structure in the intron RNA. RNA 2004; 10:1433-43; PMID: 15273321; http://dx.doi.org/10.1261/rna.7730104
49. Chillón I, Martínez-Abarca F, Toro N. Splicing of the *Sinorhizobium meliloti* RmInt1 group II intron provides evidence of retroelement behavior. Nucleic Acids Res 2011; 39:1095-104; PMID:20876688; http://dx.doi.org/10.1093/nar/gkq847
50. Leclercq S, Cordaux R. Selection-driven extinction dynamics for group II introns in Enterobacteriales. PLoS One 2012; 7:e52268; PMID:23251705; http://dx.doi.org/10.1371/journal.pone.0052268
51. Biondi EG, Toro N, Bazzicalupo M, Martínez-Abarca F. Spread of the group II intron RmInt1 and its insertion sequence target sites in the plant endosymbiont *Sinorhizobium meliloti*. Mob Genet Elements 2011; 1:2-7; PMID:22016840; http://dx.doi.org/10.4161/mge.1.1.15316
52. Matsuura M, Noah JW, Lambowitz AM. Mechanism of maturase-promoted group II intron splicing. EMBO J 2001; 20:7259-70; PMID:11743002; http://dx.doi.org/10.1093/emboj/20.24.7259
53. Cui X, Matsuura M, Wang Q, Ma H, Lambowitz AM. A group II intron-encoded maturase functions preferentially in cis and requires both the reverse transcriptase and X domains to promote RNA splicing. J Mol Biol 2004; 340:211-31; PMID:15201048; http://dx.doi.org/10.1016/j.jmb.2004.05.004
54. Huang H-R, Chao MY, Armstrong B, Wang Y, Lambowitz AM, Perlman PS. The DIVa maturase binding site in the yeast group II intron aI2 is essential for intron homing but not for in vivo splicing. Mol Cell Biol 2003; 23:8809-19; PMID:14612420; http://dx.doi.org/10.1128/MCB.23.23.8809-8819.2003
55. Molina-Sánchez MD, Barrientos-Durán A, Toro N. 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 2011; 286:21154-63; PMID:21521690; http://dx.doi.org/10.1074/jbc.M110.210013
56. Blatny JM, Brautaset T, Winther-Larsen HC, Haugan K, Valla S. Construction and use of a versatile set of broad-host-range cloning and expression vectors based on the RK2 replicon. Appl Environ Microbiol 1997; 63:370-9; PMID:9023917
57. Molina-Sánchez MD, Martinez-Abarca F, Toro N. Excision of the *Sinorhizobium meliloti* group II intron RmInt1 as circles in vivo. J Biol Chem 2006; 281:28737-44; PMID:16887813; http://dx.doi.org/10.1074/jbc.M602695200
58. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. Genome Res 2004; 14:1188-90; PMID:15173120; http://dx.doi.org/10.1101/gr.849004