Characterization of the short c-terminal region of the *Sinorhizobium meliloti* group II intron-encoded protein that lacks the DNA endonuclease domain

María Dolores Molina-Sánchez, Francisco Martínez-Abarca and Nicolás Toro*

Grupo de Ecología Genética, Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas, Calle Profesor Albareda 1, 18008 Granada, Spain

*Correspondence

N Toro, Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas, Calle Profesor Albareda 1, 18008 Granada, Spain

Phone: (+34) 958181600

Fax: (+34) 958129600

E-mail: nicolas.toro@eez.csic.es

*Running title*: Group II intron maturase C-terminal region

*Key words*: catalytic RNAs, maturase, reverse transcriptase, splicing, retroelements
Summary

Group II introns are both catalytic RNAs and mobile retroelements that move through a process catalyzed by a ribonucleoprotein (RNP) complex that consists of an intron-encoded protein (IEP) and the spliced intron lariat RNA. Group II IEPs are multifunctional and contain an N-terminal reverse transcriptase (RT) domain, which functions in intron mobility, followed by a putative RNA-binding domain (domain X) associated with RNA splicing or maturase activity. These IEPs also have a C-terminal DNA binding (D)/DNA endonuclease (En) region. The IEP encoded by the mobile group II intron RmInt1, which lacks the D/En region, has only a short C-terminal extension (C-tail) after a typical domain X, apparently unrelated to the C-terminal regions of other group II IEPs. Multiple sequence alignments identified features of the C-terminal portion of the RmInt1 IEP that are conserved throughout evolution suggesting a subgroup-specific function distinct of maturase activity. The functional importance of these features was demonstrated by analyses of deletions and mutations affecting conserved amino acid residues. Within the C-terminal portion of the RmInt1 IEP maturase domain, we identified conserved residues in a putative α-helical region that are specifically required for the insertion of the intron into DNA targets in the leading strand template for DNA synthesis. These findings suggest that these group II intron IEPs may have adapted to function in mobility to optimize their insertions coupled to major insertion pathway at the replication fork.
Introduction

Group II introns are large catalytic RNAs found in organelle and bacterial genomes that splice via a lariat intermediate, in a mechanism similar to that of spliceosomal introns ([1]. The intron RNA folds into a conserved three-dimensional structure consisting of six distinct domains, DI to DVI [2]. Unlike organellar introns, most bacterial group II introns have an internally encoded (ORF within DIV) reverse transcriptase-maturase (RT). This intron-encoded protein (IEP) is required for folding the intron RNA into a catalytically active structure in vivo [3-6]. Mobility of these group II introns occurs by means of a target DNA-primed reverse transcription mechanism involving a ribonucleoprotein (RNP) complex containing both the intron RNA and the IEP [7-9].

The group II IEPs have an N-terminal RT domain homologous to retroviral RTs, followed by a putative RNA-binding domain associated with RNA splicing or maturase activity (domain X), and C-terminal DNA binding (D)/ DNA endonuclease (En) region [10,11]. Biochemical analyses of LtrA mutants (the IEP of the Ll.ltrB intron of Lactococcus lactis) have suggested that the N-terminus of the RT domain is required for protein interactions with the high-affinity binding site in subdomain DIVa of the intron, whereas other regions of the RT and domain X interact with conserved catalytic core regions [13]. Domain X is located in the position corresponding to the “thumb” and part of the connection domains of retroviral RTs, and appears to have a similar structure to these enzymes [10,12]. The RT domain and domain X are required for RNA splicing [13]. The En domain, which carries out second-strand cleavage to generate the primer for reverse transcription of the inserted intron RNA, contains sequence motifs characteristic of the H-N-H family of endonucleases, interspersed with two pairs of cysteine residues [11,14,15]. Deletion of the conserved En domain abolishes bottom-strand cleavage, but the truncated protein retains RNA splicing activity and
can carry out reverse splicing of the intron RNA into double-stranded DNA target sites. Further deletions of the upstream variable region abolish stable DNA binding and reverse splicing into double-stranded DNA target sites, but the protein retains its ability to splice RNA and to carry out reverse splicing into single-stranded DNA target sites [16-19], albeit at a lower rate than the wild-type protein (~10% wild-type levels).

Three main classes (IIA, IIB and IIC) of group II introns have been described based on the conserved intron RNA structures [2, 20-25]. The *Lactococcus lactis* L1.ltrB intron and the yeast aI1 and aI2 introns, which are the best studied mobile introns and serve as a paradigm for group II intron mobility, all belong to the IIA class. The *Sinorhizobium meliloti* group II intron RmInt1 is a mobile intron that belongs to subclass IIB3 [24] (although some have classified it as belonging to bacterial class ORF D, [21]) showing a IIB-like RNA structure with some IIA features (for further details see [21]). Moreover, unlike lactococcal and yeast introns, the RmInt1 IEP and the members of this subclass lack the C-terminal D/En region [11,14,21,24,27]. *In vitro* assays show that RmInt1 RNPs are thus unable to carry out second-strand cleavage, but do perform reverse splicing into the target site, in both single- and double-stranded DNA substrates [28]. RmInt1 is an efficient mobile element, with two retrohoming pathways for mobility, the preferred pathway involving reverse splicing of the intron RNA into single-stranded DNA at a replication fork, using the nascent lagging DNA strand as the primer for reverse transcription [29]. Like lactococcal and yeast introns, RmInt1 retrohoming also requires base-pairing interactions between the intron RNA and the DNA target [30,31]. A previous report [11] suggested that the IEP of RmInt1 differs from other IEPs in having only a short (20 aa) C-terminal extension (hereafter referred to as the C-tail) after a typical domain X, which appears to be unrelated to the C-terminal regions of other group II IEPs. It has also been suggested that this C-tail may be a primordial or remnant DNA-binding region, an extension of domain X, or simply a non functional extension.
However, the contribution, if any, of the RmInt1 IEP or its C-tail to DNA binding and target-site recognition has remained unclear.

We studied the C-terminal region of the RmInt1 IEP along to the maturase domain and the C-tail. We found that the C-tail and upstream amino-acid residues, located in a putative $\alpha$-helical region, form a functionally important region of the IEP maturase domain conserved throughout the evolution of subclass IIB3 group II IEPs. We show here that conserved residues in the former putative $\alpha$-helical region are specifically required for the insertion of the intron into DNA targets using the leading strand pathway.
Results and Discussion

Multiple Sequence-Structure Alignments

Multiple sequence alignments with Group II IEPs of Subclass IIB3 suggested that the C-tail of the RmInt1 IEP may extend from amino acid residues 400 to 419 ([11], Fig. 1). Fig. 1A shows multiple sequence alignments of the C-terminal region (domain X and downstream residues) of subclass IIB3 group II IEPs (see Material and Methods). The C-terminal region includes the two most highly conserved sequence motifs in domain X of group II IEPs: RGWXNYY (RmInt1 residues 349-355) and R(K/R)XK (RmInt1 residues 380-383). The predicted secondary structure of the RmInt1 domain X includes four putative $\alpha$-helices, as in most group II IEPs [12], flanked by two putative short $\beta$-strands. The two conserved domain X motifs are found at or near the C-termini of $\alpha_2$ and $\alpha_3$, respectively. The $\alpha$-helices $\alpha_1$, $\alpha_2$ and $\alpha_3$ potentially correspond to $\alpha$-helices $\alpha_H$, $\alpha_I$ and $\alpha_J$ in the thumb of HIV-RT [12]. However, $\alpha$-helix $\alpha_4$ of RmInt1 IEP has no equivalent predicted structure in HIV-1 RT or LtrA protein.

The domain X region of group II intron RTs extends downstream from $\alpha_J$ into the region corresponding to the connection domain of HIV-1 RT [10], which is characterized by three adjoining $\beta$-strands involved in protein dimerization [12]. This downstream region contains a conserved lysine residue in domain X (K483 in LtrA; [12]), whose mutation reduces maturase activity [13]. Interestingly, the amino acid residue in the equivalent position of IIB3 introns is a highly conserved leucine residue (L396 in RmInt1). Upstream of this conserved leucine residue, the domain X contains the conserved amino acid residues HKXRA (RmInt1 residues 388 to 392) carrying a stretch of basic amino acids at the N-terminus of a predicted $\alpha$-helix ($\alpha_4$). Some of the residues of the HKXRA motif are also conserved in some other group II IEPs at similar positions, together with the predicted $\alpha$-helix [12]. In addition, an
idiosyncratic conserved sequence motif of subclass IIB3 group II IEPs, LF(V/A)HW (RmInt1 residues 406-410), lies within a predicted β-strand (β2) in the C-tail.

To summarize, the information content (Fig. 1B) of each position in domain X suggests that the C-terminal portion of the subclass IIB3 maturase (Fig. 2A) is characterized by a well conserved sequence motif (hereafter referred to as IIB3 IEP motif), LX3AX3PXLF(V/A)HW (RmInt1 residues 396-410). Thus, these IEPs appears to diverge from the C-terminal regions of other group II IEPs from amino acid residues 396 to 419, corresponding to a larger C-tail than previously thought. Our findings indicate that the C-terminal portion of the IEP (including part of the maturase and the C-tail), has been conserved throughout evolution of subclass IIB3 group II IEPs, suggesting a subgroup-specific function.

Effect of mutations in the C-terminal region of the RmInt1 IEP on RNA splicing in vivo.

We constructed a series of mutants to identify the features of the C-terminal region of the RmInt1 IEP. Three of these mutants had C-terminal truncations of different sizes, whereas others had amino acid substitutions in various positions (Fig. 2A). Intron RNA excision was analyzed by primer extension analysis in both total RNA (Fig. 2B) and RNP particles preparations (Fig. 2C) using a primer P (see Material and Methods) complementary to a sequence located 80-97 nt from the 5’end of the intron [28]. The previously reported domain X mutant K381A [32], in which the last conserved lysine residue of the conserved R(K/R)XK motif was replaced by an alanine residue, retained RNA splicing activity (~30% wild-type), measured in both RNA and RNP particle preparations. This data suggests that the mutant K381A IEP remain bound to the spliced lariat intron RNA. By contrast the mutant (ΔC29) in which the IEP was truncated such that the last 29 amino acid residues were missing showed no detectable RNA splicing activity, consistent with the truncation affecting part of domain X. Interestingly, mutants with shorter C-terminal truncations (ΔC14 and ΔC21)
lacking part or all of the C-tail displayed no detectable splicing activity in vivo. A similar result was obtained with the previously reported domain X double mutation YY→AA [32] in the conserved RGWXNYY motif, in which the Y354 and Y355 amino acids were replaced by alanine residues and the pKG2.5X mutant [26], in which the IEP was truncated in the RT domain. Thus, the pattern of inhibition for truncations and YY→AA mutations was consistent with gross structural alterations to the IEP. We conclude that the C-tail is structurally and functionally important for these RT proteins.

Despite the very different structures and properties of the changed amino acid within the predicted α4-helix (Fig. 2A), these mutants retained substantial RNA splicing activity, reaching values of 50-102% wild-type levels. Similarly, the P404T mutant retained a high level of RNA splicing activity in both RNA (80% wild-type) and RNP particles preparations, (70% wild-type). The data for these mutants indicate that the conserved residues in the predicted α4-helix and the downstream P404 residue are not required for RNA splicing. By contrast, mutations in the conserved residues of the downstream LF(V/A)HW motif within a predicted β-strand (β2) strongly inhibited (R406, R407, D410, F410 and P410) or abolished the splicing reaction. The most critical residue is the conserved H409, which is invariant in multiple sequence alignments. Our results suggest that the conserved residues in the predicted β2 strand are required for efficient intron RNA splicing and hence they contribute to the maturase function of these RT proteins. Together, these results are consistent with the possibility that the conserved amino acid residues in the predicted α4 helix rather than being part of the maturase domain represent a distinct functional region.

Effect of mutations in the C-terminal region of the RmInt1 IEP on intron mobility

In order to perform mobility assays for RmInt1 a donor and recipient plasmid assay was conducted as reported previously [29]. S. meliloti strain RMO17 harboring the intron donor
plasmid was transformed with recipient plasmids in which the target site was cloned in both orientations with respect to the direction of DNA replication. As expected all the mutants in the C-terminal region of the RmInt1 IEP that showed no detectable splicing activity in vivo were unable to insert site-specifically in DNA targets cloned in both orientations with respect to the replication fork (Fig. 3). Interestingly, none of the mutations in the conserved residues in the α4 helix displayed retrohoming on pJB0.6LEAD containing the target cloned in the leading-strand orientation, but some of them (M391, R400 and V400) retained retrohoming activity into the target DNA site when inserted into the lagging strand (pJB0.6LAG). This phenotype depends on the amino acid used for the replacement (compare E400 with mutations R400 and V400). By contrast, the P404T mutant retains mobility efficiencies close to 80% on pJB0.6LEAD or similar to those of the wild-type on pJB0.6LAG. The C-tail R407 mutant, as occurs with point mutation K381A in domain X, displays no detectable mobility, which might be explained by the strongly reduced splicing activity of both mutants. However, the R406 and F410 mutants retained substantial levels of retrohoming activity with DNA targets cloned in both leading and lagging strand orientation. Again, this ability depends on the residue used for the replacement (compare F410 with mutations D410 or P410). Furthermore, similar mutations carried out in more efficient constructs for retrohoming (ΔORF and IEP expressed in cis) showed similar invasion levels in both target orientations with respect to the direction of the plasmid replication (data not shown). Therefore, mobility can not be directly predicted from the extent of splicing: mutants that splice poorly, such as R406 and F410, are nevertheless mobile, whereas mutants E388 and D389, which are highly active in splicing, do not support any detectable intron mobility.

To summarize mutations of conserved residues within the predicted α4 helical region led to substantial level (≥50% wild-type) of splicing activity, but highly impaired intron mobility. Such mobility was detected only with target sites cloned to serve as templates for the nascent
lagging strand in DNA replication fork, the preferred retrohoming pathway of RmInt1, and it
depends on the residue used for the replacement. Thus, these residues seem to contribute to
intron mobility and might be required for intron insertion into DNA targets on the template
for the leading strand DNA replication. Additional data further support the above conclusion,
i.e. the F410 mutation in the C-tail showed similar reduction of retrohoming (47% of the wild
type) in a target in the lagging strand orientation, but still has retrohoming in the leading
strand template (55% of the wild type). It has been suggested that this minor retrohoming
pathway [29] may involve reverse splicing into either double-stranded DNA or transiently
single-stranded DNA target sites, and that priming may include random non specific
opposite-strand nicks, a nascent leading strand or *de novo* initiation of cDNA synthesis. As
most of the mutants were able to cleave both single- and double-stranded DNA substrates
(not shown), the impairment of mobility may reflect the requirement of these residues for
interactions involved in the priming reaction after reverse splicing of the intron RNA into
these target sites. Our results support the hypothesis that these group II intron IEPs may have
adapted to function in mobility by different mechanisms to make use of either leading or
lagging-oriented targets in the absence of DNA endonuclease domain.
Material and Methods

Bacterial strains, media and growth conditions

*Sinorhizobium meliloti* RMO17 was cultured at 28°C on TY medium for RNA extraction and RNP particle isolation. *Escherichia coli* DH5α was used for the construction of mutants and cloning. *E. coli* was grown in LB medium at 37°C. For plasmid maintenance the antibiotic kanamycin was added at a concentration of 200 μg/ml for rhizobia and 50 μg/ml for *E. coli*; ampicillin was added at a concentration of 200 μg/ml for both and the medium was supplemented with tetracycline at a concentration of 10 μg/ml for mobility assays.

Sequence alignments and secondary-structure prediction

We searched the NCBI database for subclass IIB3 group II IEPs, using BlastP with the amino-acid sequence (127 residues) of the RmInt1 IEP domain X (positions 293-419) as the protein query sequence. The first 66 Blast hits obtained with this query protein were complete or fragmented group II IEPs of the IIB3 subclass. For sequence alignments, we chose 35 complete IEPs harbored by different bacterial species including 18 out of 22 currently (updated March 11, 2008) classified as group II intron bacterial class ORF D in the group II intron data base (www.fp.ucalgary.ca/group2introns/). ClustalW was used to generate sequence alignments, and secondary structure was predicted with the JPred server (http://www.compbio.dundee.ac.uk/~www-jpred/). The logo sequence [33,34] was obtained based on this alignment (http://weblogo.berkeley.edu/).

RmInt1 and mutant derivative

The pKG2.5, pKG2.5X, pKG2.5-YAH, pKG2.5D5-CGA, pKG2.5-ΔC29, pKG2.5-A354A355, pKG2.5-A381 constructs have been described in previous studies [26,28,32].
Most of the RmInt1 IEP maturase mutants (Supplementary Table 1) were generated by site-directed mutagenesis, using the Altered Sites II \textit{in vitro} Mutagenesis pAlter-1 System (Promega), with changes introduced in the pAL2.5 plasmid. This plasmid contains RmInt1 flanked by exons -175/+466 inserted into pALTER-1 as a \textit{Sph}I fragment [28]. The changes were introduced through the use of DNA oligonucleotides hybridizing around the position of the intended mutation, abolishing antibiotic resistance. The final constructs were generated by inserting the RmInt1-containing fragment resulting from \textit{Bam}HI/\textit{Spe}I digestion of pAL2.5 into pKG0. The primers used for mutagenesis are shown in supplementary Table 1. The pKG2.5-V400 mutant was constructed by a two-step PCR procedure using the Triple Master™ PCR System (Eppendorf). Two pairs of primers were designed to amplify the 5’ and 3’ sections of the IEP, respectively: a 5’ end primer mut UP (5’-GTCAGCGGTGCTGGAAGTATG-3’) and a 3’ end primer A400V/DN (5’-ATTTTCCGCACACAGCTTTGCAAGA-3’) were used to generate the upstream 824 bp fragment; a 5’ end primer A400V/UP (5’-GAAAGCTGGTGCGGGAAAATCCGGG-3’) and a 3’ end primer mut DN (5’-GCGCGCGTAATACGACTCAC-3’) were used to generate the downstream 689 bp fragment. The mutagenic primers contained a 20 bp region of overlap and introduced a valine (V) residue in place of the moderately conserved alanine (A) in position 400 of the IEP, by changing a C to a T in intron position 1745. The final 1492 bp fragment was amplified, digested with \textit{Eco}RI and \textit{Spe}I and used to replace the corresponding wild-type fragment in pKG2.5.

\textbf{RNA isolation and RNP particle preparation}

RNA and RNPs were extracted from free-living cultures of \textit{S. meliloti} strain RMO17 containing plasmids encoding the wild-type or mutant RmInt1, as described in previous studies [28,32]. For RNA isolation, we collected the cells present in 10 ml of TY medium,
supplemented with kanamycin and with an OD$_{600}$ close to 0.6 units. The cells were lysed and
their DNA was eliminated by incubation with 50 units of RNase-free DNase I. RNP-enriched
fractions were obtained from 200 ml of TY medium plus kanamycin with an OD$_{600}$ of around
0.8 units. A clarified lysate of the bacterial cells was layered onto a 1.85 M sucrose cushion
and subjected to 20 h of centrifugation in a Beckman Ti50 rotor at 50,000 x g. The resulting
pellet was resuspended in 10 mM Tris-HCl (pH 8.0), 1 mM DTT.

Primer extension assays

These assays were carried out on both total RNA and RNP particle preparations. Primer
extension reactions were carried out essentially as previously described [28]. The annealing
mixture had a volume of 10 µl and contained either 15 µg of total RNA or 0.125 A$_{260}$ units
(equivalent of 5µg of ssRNA) of RNP particles and 0.2 pmol (300,000 CPM) of (5'-32P)-
labeled P primer (5'-TGA AAG CCG ATC CCG GAG) in 10mM PIPES (pH 7.5), 400 mM
NaCl. This mixture was first heated at 85ºC for 5 min, and was then rapidly cooled to 60ºC
and allowed to cool more slowly to 45ºC. Extension reactions were initiated by adding 40 µl
of 50 mM Tris-HCl (pH 8.0), 60 mM NaCl, 10 mM DTT, 6 mM MgOAc, 1 mM each of all
four dNTPs, 60 µg/ml actinomycin D (SIGMA), 15 units of RNAguard™ RNase inhibitor
(GE Healthcare) and 7 units of AMV RT (Roche). Reaction mixtures were incubated at 42ºC
for 60 min. The reaction was stopped by adding 15 µl of 3M NaAc (pH 5.2) and 150 µl of
cold ethanol. Samples were resolved by electrophoresis in a denaturing 6% polyacrylamide
gel. Primer extension products were quantified with the Quantity One software package (Bio-
Rad Laboratories) and excision efficiency was measured as 100[S/(S+Pr)].

In vivo retrohoming assays
The mobility of RmInt1 was revealed by a two-plasmid assay and further Southern hybridization [26,29]. A donor plasmid (pKG2.5 or IEP mutant derivatives) containing the full-length intron flanked by a 640 bp fragment (-174/+466) of ISRm2011-2 was transferred from *E. coli* DH5α to *S. meliloti* RMO17, an RmInt1-less strain. The rhizobial host contained a recipient plasmid bearing a 640 bp fragment with the intron insertion site in the same (pJB0.6LAG) or opposite (pJB0.6LEAD) orientation to the replication fork. The recipient plasmid pJBΔ129, which lacks the RmInt1 target, was used as a negative control in these assays. Plasmids isolated from transconjugants were analyzed by *SalI* digestion, agarose gel electrophoresis and Southern blotting with an ISRm2011-2 probe. Generally, we could obtain three hybridization bands: the linearized donor plasmid (7859 bp), a fragment of the recipient plasmid containing the intron DNA target (2017 bp) and an extra band when the recipient plasmid has been invaded (3901 bp). Retrohoming *in vivo* efficiency was determined as the ratio of intron-invaded recipient plasmids to the homing products plus non invaded recipient plasmids, and was expressed as a percentage. The data shown are the means of at least eight independent determinations, with the corresponding standard deviation.
Acknowledgments

We would like to thank Ascensión Martos Tejera and Vicenta Millán Casamayor for technical assistance. This work was supported by research grant BIO2008-00740 from the Ministerio de Ciencia y Tecnología and grant CVI-01522 from Junta de Andalucía. M.D.M.-S. was supported by a predoctoral fellowship from Junta de Andalucía.
References


Figure legends

Fig. 1. Multiple sequence alignments. (A) The C-terminal region of the RmInt1 IEP (Sr.me.I1) was aligned with other group II IEPs of subclass IIB3, using ClustalW. Conserved amino acid residues are highlighted: black, >50% identity; gray, >50% similarity; shading was achieved with Boxshade (bioweb.pasteur.fr/seqanal/interfaces/boxshade.html). Residue numbers according to RmInt1 sequence. The predicted secondary structure of the RmInt1 IEP X domains, based on JPred folding prediction, is shown above the alignments, and a consensus sequence (indicated by dots) is shown below. Residues identical in all sequences are indicated by asterisks. Highly conserved motifs in the X domain of group II IEPs RGWXNYY (RmInt1 residues 349-355) and R(K/R)XK (RmInt1 residues 380-383) are indicated by a line above the secondary structure prediction. The putative boundaries of domain X and the C-tail are indicated by arrows showing the RmInt1 IEP sequence. The bacterial species and the corresponding accession numbers of the IEPs are: *Sinorhizobium meliloti* (Sr.me.I1, NP_437164); *Ensifer adhaerens* (E.a.I1, AAP83798); *Sinorhizobium medicae* (Sr.med., YP_001313619); *Sinorhizobium terangae* (Sr.t.I1, AAU95643); *Escherichia coli* (E.c.I2, CAAS4637); *Shewanella putrefaciens* (Sh.p., YP_001181807); *Azoarcus* sp. EbN1 (Az.sp., YP_159836); *Legionella pneumophila* (L.p., YP_001251128); *Escherichia coli* B (E.c., ZP_01698243); *Prosthecocloris aestuarii* (Pr ae.I3, ZP_00592895); *Prosthecocloris vibrioformis* (Pr.vi.I1, YP_001129678); *Pelodyction phaeocllathratiforme* (Pe.ph.I1, ZP_00589124); *Chlorobium phaeobacteroides* (Ch.ph., YP_911931); *Syntrophus aciditrophicus* (Sy.a., YP_460783); *Methanosarcina acetivorans* (M.a.I5, NP_619481); uncultured archaeon GzFos32G12 (UA.I3,, AAU83697); *Bacillus thuringiensis* (B.thu., ZP_00738953); *Paracoccus denitrificans* (Pa.de.I1, ZP_00628808); *Photorhabdus luminescens* (Ph.l.I2, NP_928428); *Magnetococcus* sp. (Ma.sp.I3,
YP_864580); *Pseudomonas aeruginosa* (P.ae., ABR13526); *Pseudomonas stutzeri* (P.st. I3
YP_001172226); *Burkholderia phymatum* (Bu.ph., ZP_01505671); *Frankia* sp. (Fr.sp.,
YP_482811); *Saccharopolyspora erythraea* (S.ery., YP_001104541); *Pelobacter acetylenicus* (Pe.a., AAQ08377); deltaproteobacterium MLMS-1 (delta, ZP_01288325);
*Bradyrhizobium japonicum* (B.j.I1, NP_768692); *Shigella dysenteriae* (S.dy.I1, YP_406035);
*Alkaliphilus metalliredigens* (Al.me.I4, YP_001321146); *Bacteroides thetaiotaomicron*
(B.t.I4, NP_811528); uncultured marine bacterium 18874410 (UMB.I3, AAL78690);
uncultured marine bacterium 18874275 (UMB.I1, AAL78688); *Psychroflexus torquis* (Pch.t.,
ZP_01254488); uncultured marine bacterium 18874408 (UMB.I2, AAL78689). The introns
are named according to the Zimmerly nomenclature (www.fp.ucalgary.ca/group2introns/).

(B) Sequence logo for group II IEPs of subclass IIB3. The sequence logo
(http://weblogo.berkeley.edu/) shows the information content (4 bits = no degeneracy) for
each position in domain X, and is based on the multiple sequence alignment shown in Fig.
2A. Amino acids are colored according to properties: basic, blue (K, R, and H); acidic, red (D
and E); hydrophobic, green (P, L, I, V, M, F, W, Y, and A); polar, purple (N, Q, S, and T);
and black (G and C). Highly conserved motifs in the X domain are indicated by a line below
the logo.

Fig. 2. Effect of RmInt1 IEP C-terminal mutations on intron RNA splicing. (A) Detailed
sequence of the C-terminal region of the RmInt1 IEP. Conserved amino acids are shown in
bold. Changes are indicated below each position; deletions are shown with arrows below the
sequence. The boxed residues correspond to the IIB3 IEP motif at the C-terminal region. The
predicted secondary structure is indicated above the sequence; α-helices are represented with
cylinders and the β-strand is shown as an arrow. Amino acid positions are indicated. (B)
Splicing measured in total RNA samples, or (C) in RNP particles preparations. A
representative gel of the primer extension analysis is shown for each mutant. The molecular
sizes of the cDNAs extension products, spliced intron RNA (S) and unspliced precursors (Pr), are
indicated. cDNA bands corresponding to the resolved extension products were quantified with
the Quantity One software package (Bio-Rad Laboratories) and intron splicing was measured
as 100[S/(S+Pr)]. Splicing efficiency was plotted as percentage of wild-type values in
pKG2.5. In addition to the C-terminal mutants, other mutants were used as negative controls:
2.5X, which has a frame-shift at the beginning of the IEP sequence; YAHH, which has a
mutation affecting the active site for RT activity (RT domain 5); and, D5-CGA, which has a
mutation in the catalytic triad of the ribozyme catalytic core (RNA domain V).

Fig. 3. Retrohoming in vivo of wild-type RmInt1 and mutant derivatives on DNA target sites
cloned in opposite orientations relative to the direction of plasmid replication. Plasmid pools
from *S. meliloti* RMO17 harboring donor (pKG2.5) and target plasmids (pJB0.6LEAD or
pJB0.6LAG) were analyzed by digestion and Southern hybridization with an exon-specific
probe. Recipient plasmid without the DNA target (pJBΔ129) was used as a negative control
in the assays. Schematic diagrams of the mobility assays are shown at the top (not drawn to
scale). Here, we indicate the *Sal*I restriction sites (S) in the plasmids and the probe used (P)
as well as the orientation of the target respect to the replication fork (arrows). The recipient
plasmids contain the intron DNA target cloned in the same (LAG) or in opposite (LEAD)
orientation depending on whether the nascent lagging or leading DNA strand could be used
as a primer for reverse transcription of the inserted intron RNA. The Southern blots are
shown below and the hybridization signal corresponding to the target recipient plasmid (T)
and the homing product (H) are indicated. Donor plasmid is omitted.
Fig. 2
Fig. 3