

Site-specific reverse splicing of a HEG-containing group I intron in ribosomal RNA

Ása B. Birgisdóttir¹ and Steinar Johansen^{1,2,*}

¹Department of Molecular Biotechnology, Institute of Medical Biology, University of Tromsø, N-9037 Tromsø, Norway and ²Faculty of Fisheries and Natural Sciences, Bodø Regional University, N-8049 Bodø, Norway

Received January 18, 2005; Revised and Accepted March 18, 2005

ABSTRACT

The wide, but scattered distribution of group I introns in nature is a result of two processes; the vertical inheritance of introns with or without losses, and the occasional transfer of introns across species barriers. Reversal of the group I intron self-splicing reaction, termed reverse splicing, coupled with reverse transcription and genomic integration potentially mediate an RNA-based intron mobility pathway. Compared to the well characterized endonuclease-mediated intron homing, reverse splicing is less specific and represents a likely explanation for many intron transpositions into new genomic sites. However, the frequency and general role of an RNA-based mobility pathway in the spread of natural group I introns is still unclear. We have used the twin-ribozyme intron (Dir.S956-1) from the myxomycete *Didymium iridis* to test how a mobile group I intron containing a homing endonuclease gene (HEG) selects between potential insertion sites in the small subunit (SSU) rRNA *in vitro*, in *Escherichia coli* and in yeast. Surprisingly, the results show a site-specific RNA-based targeting of Dir.S956-1 into its natural (S956) SSU rRNA site. Our results suggest that reverse splicing, in addition to the established endonuclease-mediated homing mechanism, potentially accounts for group I intron spread into the homologous sites of different strains and species.

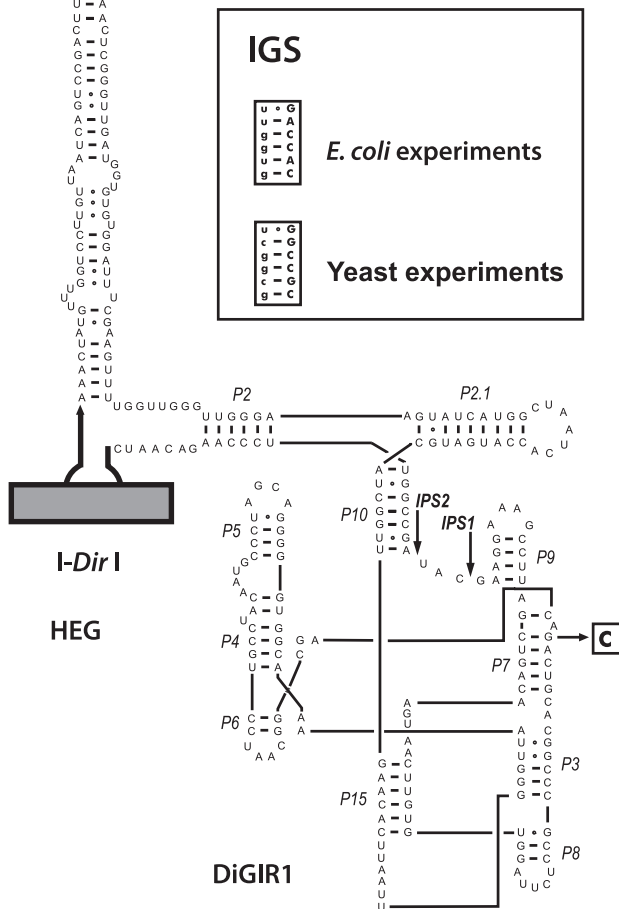
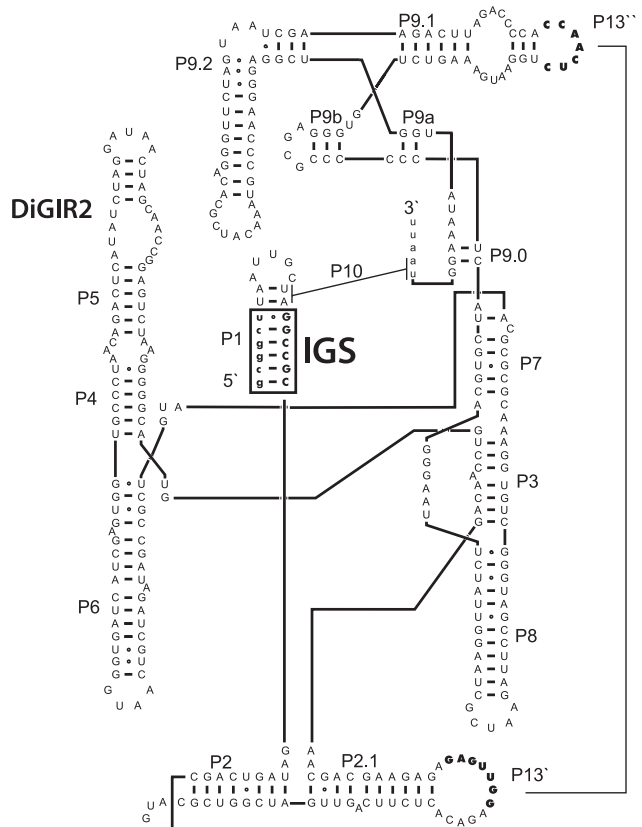
INTRODUCTION

Group I introns are self-splicing, mobile introns that are relatively common in ribosomal DNA (rDNA) of fungi and protists (1–3). There are two different outcomes of group I intron mobility; i.e. introns can insert themselves into the homologous site in an intron-less allele (intron homing), or

less frequently, they may spread into a new site in the same or a different gene. Endonuclease-mediated intron homing is an efficient process and involves a homology-dependent gene conversion event at the DNA level. Homing is initiated by an intron-encoded homing endonuclease that recognizes and generates a double-stranded DNA break close to the site of intron insertion (4). Group I introns may also insert into the natural intron insertion site at the RNA level, and subsequently become stably integrated into the host genome. The reaction pathway involves reversal of the forward self-splicing reaction of the intron. During the process, the short internal guide sequence (IGS; 4–6 nt) near the 5' end of the free intron RNA recognizes the insertion site, and the intron becomes integrated into the non-interrupted exon RNA. Reverse transcription (RT) of the recombined RNA and genomic integration of the resulting cDNA is necessary for stable genome integration. Experimental evidence for reverse splicing both *in vitro* and in a bacterial system comes from studies with the *Tetrahymena* prototype intron (5–7). The intron was targeted against position L1926 in the *Escherichia coli* large subunit (LSU) rRNA to preserve the U-G pair at the 5' splice site (position L1925 is homologous to the *Tetrahymena* splice junction, but would result in a C-G pair at the 5' splice site). Expression of the intron in *E.coli* resulted in complete integration of the intron into the targeted rRNA site. The experiments showed that reverse splicing might also result in intron integration into heterologous rRNA sites (7). Furthermore, variations in the guide sequence resulted in partial intron integration at numerous sites in the *E.coli* LSU rRNA and complete integration at one novel site (8). In conclusion, these results indicate that the frequency of reverse splicing is modulated by the structure of the rRNA and that reverse splicing *in vivo* is only moderately site specific. Therefore, reverse splicing could contribute to transposition of group I introns among rRNA genes.

The small subunit (SSU) rDNA of the *Didymium iridis* myxomycete is interrupted by the mobile group I intron Dir.S956-1 [inserted after position 956, see (9) for rDNA intron nomenclature]. Dir.S956-1 has an unusual twin-ribozyme organization (Figure 1) (10,11) with a large

*To whom correspondence should be addressed. Tel: +47 77 64 53 67; Fax: +47 77 64 53 50; Email: Steinar.Johansen@fagmed.uit.no



sequence insertion in the P2 paired element of a canonical group IE splicing ribozyme (DiGIR2). The unique P2 insertion encodes a small group I-like RNA processing ribozyme (DiGIR1) followed by the I-DirI homing endonuclease gene (HEG). DiGIR2 catalyzes intron splicing, 3'-splice site hydrolysis and the formation of full-length intron RNA circles (10,12–14). Removal of DiGIR1 and the HEG allows DiGIR2 to function as a group IE ribozyme of only 343 nt. Unlike DiGIR2, DiGIR1 is not involved in intron splicing. Instead DiGIR1 is involved in the maturation of the HEG mRNA (11,15–18). The Dir.S956-1 intron is mobile in genetic crosses between intron-containing and intron-lacking *Didymium* strains, and efficiently invades the intron-less alleles in a highly specific manner typical of endonuclease (in this case I-DirI) mediated homing events (19).

Comparative sequence data, in general, support intron spread into homologous sites as the most common outcome of horizontal intron transfers in nature (20–22). Here, we investigate if the natural Dir.S956-1 twin-ribozyme intron, which is capable of invading intron-less rDNA alleles through endonuclease-mediated homing, can reverse splice into the SSU rRNA. We demonstrate integration of Dir.S956-1 into its natural rRNA site in *Escherichia coli* and *Saccharomyces cerevisiae*. Our results indicate that an RNA-based intron mobility pathway can contribute to group I intron spread into homologous sites.

MATERIALS AND METHODS

Plasmid construction

The plasmid pMAL-Dir.S956-1 (EC), used for expression in *E.coli*, was prepared by ligating an EcoRI–PstI fragment containing the *Didymium* twin-ribozyme intron with a mutated IGS (see Figure 1) into the expression vector pMAL-c2X (New England Biolabs) leaving the lacZ' translation frame restored after intron splicing. The mutated intron-containing fragment was prepared by PCR amplification of pDiSSU1 (12) using primers OP490 and OP801 (see Table 1 for primers used). The plasmids pMAL-Dir.S956-1 ΔORF (EC), pMAL-Dir.S956-1 Δ353 (EC), pMAL-Dir.S-956-1 Δ353G1 (EC) and pMAL-DiGIR2 (EC) were constructed by a similar approach using pDiSSU1-ΔORF, pDiSSU1-Δ353, pGI and pDiGIR2 (10,12), respectively as a template. The plasmid used for yeast expression (pYGal-Dir.S956-1) was constructed by ligating an EcoRI–SphI cut PCR product containing the wild-type sequence of Dir.S956-1 into the yeast expression vector pYESM (23). The wild-type sequence of Dir.S956-1 was obtained by PCR amplification of pDiSSU1 with primers OP906 and OP951. pGEM3-DiGIR2 (EC) was prepared by subjecting pMAL-DiGIR2 (EC) to PCR using vector primers OP604 and OP164. The obtained PCR product was digested

Figure 1. Secondary structure model of the twin-ribozyme group I intron Dir.S956-1. Paired segments (P) are indicated in both the processing ribozyme (DiGIR1) and the splicing ribozyme (DiGIR2). The internal guide sequence (IGS) that base pairs to the 5' exon sequence in the yeast and *E.coli* experiments is boxed. The location and orientation of the 786 nt I-DirI homing endonuclease gene (HEG), within the distal region of the DiGIR2 P2 helix and the 3' region of DiGIR1, are indicated. The G to C inactivation mutation in the catalytic core segment P7 in DiGIR1 is shown. IPS1 and IPS2 are the internal processing sites of DiGIR1.

Table 1. Primers used in this work

OP8	TGGGTTGGGAAGTATCAT	Intron primer
OP85	AAGGTCGAGCTCACCCAGATTC	Intron primer
OP164	GTAAAACGACGGCCAGT (Forward M13 primer)	pMAL primer
OP236	CAGTTACCCACTAGTA	Intron primer
OP298	CTAGAACGCTGGCAGTTAG	Intron primer
OP315	CTTGGGATACCCGATGCTTCCTTT	Intron primer
OP377	ATACCAATCTGTGCACTTCA	Intron primer
OP490	GGACGAATTCAGCATGTGGTT- TAATTGCTAGACCACGATAGTC- AGCATG	Intron primer
OP604	GGTCGTCAGACTGTCGATGAAGCC	pMAL primer
OP619	GGATGCTTCCTTTCCGGAA	Intron primer
OP621	GTAGTCCACGCCGTAACGATG	<i>E.coli</i> SSU primer
OP622	AGGGTTGCGCTCGTTGCGGGACTTAA	<i>E.coli</i> SSU primer
OP776	GCTCACCTAGGCGACGATCCCTA	<i>E.coli</i> SSU primer
OP777	CGATTAACGCTTGCACCCTCCGT	<i>E.coli</i> SSU primer
OP778	GCGAAGAACCTTACCTGGTCTTGA	<i>E.coli</i> SSU primer
OP780	GCTCATTGACGTTACCCGCAGAA	<i>E.coli</i> SSU primer
OP782	TGAAGAGTTTGTATCATGGCTCAGA	<i>E.coli</i> SSU primer
OP784	GCCCTTACGACCAGGGTACACA	<i>E.coli</i> SSU primer
OP785	AGGAGGTGATCCAACCGCAGGT	<i>E.coli</i> SSU primer
OP801	CCAATCTGCAGATCAAATTACCT- TTATACCAG	Intron primer
OP906	GGACGCATGCCGTCCGGCGTA- GAGGATCTG	pGEM2 primer
OP951	GGTAGAATTCAGACCGGAAGCT- TGGGCTGCAG	PGEM2 primer
OP1009	GGATAACCAGCAAATGCTAGCAC	Yeast SSU primer

with EcoRI and PstI, and ligated into corresponding sites in pGEM-3Zf(+) (Promega). The plasmid containing the *E.coli* SSU rRNA substrate sequence from positions 800 to 1115 (pGEM3-rRNA) was constructed by ligating a PCR product containing the rRNA sequence into pUC18 using the Sure-Clone ligation kit (Amersham Pharmacia Biotech Inc.) and then subcloning it into the pGEM-3Zf(+).

Expression of group I intron ribozymes in *E.coli* and *S.cerevisiae*

Overnight cultures of JM 109 *E.coli* cells transformed with pMAL-Dir.S956-1 (EC) or one of its variants were inoculated to $A_{600} = 0.1$ in 50 ml LB+ medium with 100 µg/ml ampicillin. The cultures were grown at 37°C until $A_{600} = 0.4$, before addition of 1 mM isopropyl β-D-thiogalactosidase (IPTG). After the addition of IPTG, samples were obtained 1 and 2 h [and 4 h for the DiGIR2 (EC) transformants] after the induction, and the A_{600} value noted. The samples were diluted with LB+ medium in order to give appropriate cell concentration for total RNA isolation. Prior to IPTG-induction, 10 ml of the culture were removed and transferred to another tube and incubated for 2 h. A negative-control sample was then collected. pYGal-Dir.S956-1 was transformed into *S.cerevisiae* strain INVSc2 (*MATα*, *his3-Δ200*, *ura3-167*) from Invitrogen as described previously (24). Transformed cells were grown at 30°C in synthetic minimal medium with glucose as a carbon source and supplemented with amino acids. The culture was then plated onto plates containing galactose as a carbon source. Furthermore, cells from the galactose plates were grown in the same media until $A_{600} = 1.6$ in order to give an appropriate cell concentration for total RNA isolation.

Isolation of cellular RNA, purification of RNA and northern analysis

Total RNAs from *E.coli* and *S.cerevisiae* were isolated using the RNeasy Mini or Midi Kit (Qiagen) according to the manufacturer's protocols. Total RNA for RT-PCR analyses was treated with RQ1 RNase-free DNase (Promega), and then further purified using the RNeasy RNA Cleanup protocol provided with the RNeasy kit. Furthermore, RNA was run through a Chromaspin TE-1000 gel filtration column (CLONTECH) in order to eliminate unreacted free DiGIR2 intron or twin-ribozyme intron processing products. Northern blotting was performed using the NorthernMax™ Kit (Ambion). Five micrograms of total RNA isolated from intron-containing *E.coli* or *S.cerevisiae* was separated according to size by electrophoresis through a denaturing formaldehyde agarose gel (1%), and then transferred to a positively charged Nylon blotting membrane (BrightStar-Plus; Ambion). The RNA of interest was located by hybridization with ³²P-labelled intron-specific RNA (ribozymes) followed by autoradiography. The HEG and DiGIR1 probes were amplified from the Dir.S956-1 intron by PCR using primers OP377 and OP298, and OP8 and OP315, respectively. The PCR products were subsequently cloned into the pGEM-T vector using the pGEM-T ligation kit (Promega). The resulting probe-plasmids were then linearized with NcoI, phenol/chloroform extracted and transcribed with SP6 RNA polymerase (Promega) in the presence of [α -³²-P]UTP (3000 Ci/mmol, Amersham). The DiGIR2 probe was made from the DiGIR2 intron using a HincII linearized pGEM3-DiGIR2 plasmid as a template during transcription.

Reverse transcription, PCR, product isolation and sequencing

cDNA synthesis was performed on purified total RNA from *E.coli*, *S.cerevisiae* or on *in vitro* transcribed reverse spliced RNA using the 'First-Strand cDNA Synthesis Kit' (Amersham Pharmacia Biotech). A 2 µl aliquot from reverse transcription (RT) was amplified by a standard 50 µl PCR reaction. Cycle parameters: initial denaturation of 95°C for 2 min, then 28 cycles of 90°C for 40 s, 50 or 60°C for 1 min and 72°C for 40 s and finally a terminal elongation at 72°C for 5 min. For amplification of 5' intron integration junctions in *E.coli*, six different upstream SSU rDNA annealing primers were used, OP621, OP776, OP778, OP780, OP782 and OP784, together with the intron-specific downstream primer OP619 (also used in the RT reaction) or alternatively with intron primers OP165 or OP814 that are closer to the 5' end of the intron than OP619. 3' intron integration junctions were amplified using three different downstream *E.coli* SSU rDNA annealing primers OP622, OP777 and OP785 (also used in the RT reaction) in combination with the upstream intron-specific primer OP85. RT-PCR amplification of 3' intron integration junctions in *S.cerevisiae* was performed with the intron-specific primer OP85, and the yeast SSU rDNA-specific primer OP1009 (same as used in the RT reaction). The amplified RT-PCR products were extracted and purified from 2% agarose gels using the QIAquick Gel Extraction Kit (Qiagen). All plasmids and PCR products were sequenced manually using the dideoxy DNA sequencing method applying the Thermo Sequenase Radiolabeled Terminator Cycle Sequencing Kit (Amersham Life Science) or automatically using the BigDyeTerminator

Cycle Sequencing Ready Reaction Kit (PerkinElmer) running on an ABI Prism 377 system (PerkinElmer).

***In vitro* transcription and *in vitro* intron integration**

In vitro transcription was performed on linearized plasmids with a T7 RNA polymerase (Stratagene) using a modified version of the transcription buffer from Stratagene with only 2 mM MgCl₂ (low salt buffer). The RNA was uniformly labelled using [α -³⁵S]CTP (10 μ Ci/ μ l; Amersham Pharmacia Biotech) during the transcription, phenol-extracted, and ethanol precipitated. Prior to PAGE-purification, intron-containing transcripts were subjected to self-splicing conditions (40 mM Tris-HCl pH 7.5, 200 mM KCl, 2 mM Spermidine, 5 mM DTT, 10 mM MgCl₂, 0.2 mM GTP) at 50°C for 1 h and precipitated. The RNAs of interest were isolated from 5% polyacrylamide gels and incubated in 400 μ l elution buffer (300 mM NH₄Ac, 0.1% SDS, 10 mM Tris-HCl pH 8 and 2.5 mM EDTA pH 8) on a rotating wheel at 4°C overnight. The elution mixture was then sterile filtrated and the RNA ethanol precipitated. Five picomoles of PAGE-purified linear DiGIR2 (EC) intron RNA or DiGIR2 (EC) full-length circles (transcribed from the pGEM3-DiGIR2 (EC) plasmid) were incubated with 5 pmol PAGE-purified rRNA substrate (transcribed from the pGEM3-rRNA plasmid) in a total reaction volume of 39 μ l. The incubation was performed at 50°C in a pre-warmed 1.5 \times reverse splicing buffer containing 40 mM Tris-HCl pH 7.5, 200 mM KCl, 2 mM Spermidine, 5 mM DTT and 50 mM MgCl₂. Samples of 3 μ l were collected after 120 min incubation time and subjected to reverse transcription and PCR (see above). The primers used for amplification of 5' intron integration junctions were OP621 and OP236 (intron-specific) while the 3' intron integration junctions were amplified using OP85 and OP622.

RESULTS

Reverse splicing in *E.coli*

We investigated the ability of the mobile Dir.S956-1 twin-ribozyme intron to reverse splice in *E.coli*. Dir.S956-1 was targeted against position 956 in the bacterial SSU rRNA (S956). This position is homologous to the natural intron insertion site in *Didymium* even though the sequences surrounding the sites are not identical (5'-GUGGUU⁹⁵⁶UAAUUC in *E.coli* and 5'-GCGGCU⁹⁵⁶UAAUUU in *Didymium*). During the first step of reverse splicing, a stretch of usually 4–6 nt upstream of the integration site is recognized by the intron internal guide sequence (IGS) (5). The IGS of Dir.S956-1 was therefore changed from 5'-GGCCGC to 5'-GACCAC (Figure 1) to allow correct base pairing between the IGS and the *E.coli* SSU target sequence. This intron construct was designated Dir.S956-1 (EC). The highly expressed endogenous SSU rRNA provides an abundant substrate for reverse splicing. Dir.S956-1 (EC) with short flanking exon sequences (11 nt of 5' exon and 10 nt of 3' exon) was expressed in bacteria from the pMAL-c2X expression vector, under the control of the IPTG-inducible P_{tac} promoter (Figure 2A). Forward intron splicing from the vector transcripts is a prerequisite for reverse splicing in *E.coli*. Two hours after IPTG-induction, total RNA was isolated from intron-containing cells and the

formation of free intron RNAs in *E.coli* was indirectly monitored by RT-PCR amplification of ligated exon sequences (data not shown). In addition, free intron RNAs were directly visualized by northern blot analyses (see below). After verifying successful intron splicing, we searched for reverse splicing products. DNase-treated isolated total RNA was subjected to the RT-PCR approach presented in Figure 2B. Different primer pairs (i.e. one intron-specific and one SSU rDNA-specific) were used to amplify across 5' and 3' intron-exon junctions in the *E.coli* SSU rRNA (see Materials and Methods for details). We expected the intron to potentially target SSU rRNA sequences with only limited match to the intron IGS sequence, and therefore designed primers that would cover the entire *E.coli* SSU rRNA. A number of distinct RT-PCR products, amplified with the different primer pairs, were cut from gels, purified and directly DNA sequenced. However, intron integration was detected at S956 exclusively. Sequencing of the RT-PCR products of 238 and 390 bp, from the primer pairs covering S956, revealed 3' and 5' intron integration junctions, respectively to site 956 (Figure 2C). The other analysed RT-PCR products also revealed integration at S956 or were caused by non-specific primer annealing (data not shown). PCR reactions, with the respective primer pairs covering site 956, on the purified total RNA (without the RT reaction) and on isolated total DNA from the same bacteria did not amplify the products indicative of intron integration at site 956 (data not shown). Thus, we can conclude that the amplified RT-PCR products arise from RNA and that the integration event we observe is limited to reverse intron splicing at the RNA level. In summary, our results show that Dir.S956-1 intron RNAs reverse splice into site 956 in *E.coli* SSU rRNA. This position is homologous to the natural splice junction in *Didymium*.

Reverse splicing is dependent on the DiGIR2 ribozyme only

The *Didymium* intron consists of three distinct intron domains: the DiGIR2 splicing ribozyme, the DiGIR1 processing ribozyme and the HEG encoding the I-Dir1 homing endonuclease (Figure 1). We wanted to define the roles of DiGIR2, DiGIR1 and the HEG in reverse splicing, and expressed a series of deletion variants (Figure 3A) of the *Didymium* intron from the pMAL-vector in *E.coli*. The constructs contain the IGS that is specific for S956 in *E.coli* and lack the entire DiGIR1-HEG insertion (DiGIR2 construct), lack the HEG (Dir.S956-1 Δ HEG), or lack parts of the HEG (Dir.S956-1 Δ 353). We also included an intron variant with an inactive DiGIR1 [i.e. G to C mutation at the P7 guanosine binding site; see (10); Figure 1], and lacking a part of the HEG (Dir.S956-1 Δ 353G1-). By using the RT-PCR approach described in Figure 2B on isolated total RNA (isolated 2 h after IPTG induction), we identified intron integration to S956 for all constructs. Both the 5' and 3' integration junctions were readily amplified and provide evidence for the complete integration of the different intron RNAs into site 956 in *E.coli* SSU rRNA (data not shown).

Reverse splicing was further analyzed by northern blotting. Total RNAs isolated 0–2 h after IPTG induction were separated on formaldehyde-agarose gels, transferred to nylon membranes and hybridized to different intron-specific RNA probes.

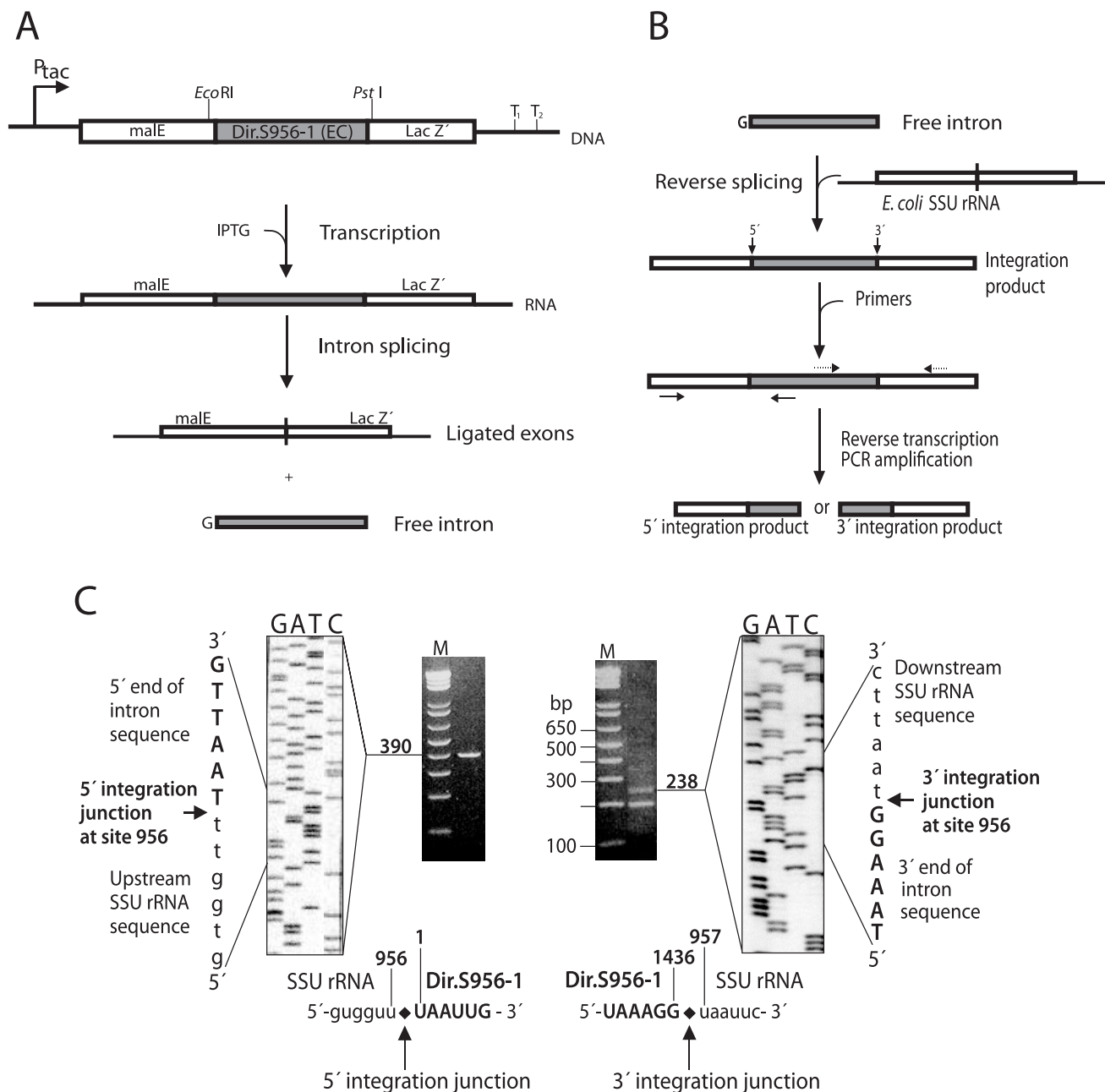


Figure 2. Reverse splicing of *Dir.S956-1* (EC) in *E. coli*. (A) Diagram of the construct pMAL-*Dir.S956-1* (EC) used for intron expression in *E. coli*. An *EcoRI*-*Pst*I fragment containing the *Dir.S956-1* (EC) intron was cloned into the pMAL-c2X expression vector under the control of the IPTG-inducible P_{tac} promoter. Self-splicing of the intron (shown schematically) results in ligated exon sequences and a free intron RNA. (B) Schematic presentation of RT-PCR amplification of reverse splicing products. After reverse splicing, integration products were reverse transcribed and amplified by PCR. For 5' integration junctions, the upstream primer anneals to the 5' exon and the downstream primer to the intron. For 3' integration junctions, the upstream primer is targeted to the intron and the downstream primer is complementary to the 3' exon. (C) RT-PCR amplification and sequence analysis of intron-*E. coli* SSU rRNA junctions. The RNA was from bacterial cells transformed with pMAL-*Dir.S956-1* (EC) and induced with IPTG for 2 h. RT-PCR with primers OP621 and OP619 for 5' integration junctions results in a product of 390 bp, which by sequencing analysis reveals the 5' end of the intron ligated to U956 of *E. coli* SSU rRNA. Amplification of the 3' integration junctions is shown with primers OP85 and OP622. The product of 238 bp represents the 3' intron integration junction at position S956. Non-specific annealing of primer OP622 to the pMAL-vector sequence during the RT reaction gave rise to the smaller product of ~200 bp. The RNA sequences flanking the observed integration junctions (marked with a diamond) at S956 are given with the intron sequence marked in bold capital letters and the rRNA sequence in lower case letters. M, size marker: 1 kb Plus DNA Ladder (Gibco BRL).

The RNA probes were complementary to the DiGIR1, the HEG or the DiGIR2 RNA sequences (Figure 3A). Several important observations were made. First, all the intron variants are able to excise themselves from precursor RNAs in *E. coli*.

This conclusion is based on the fact that signals corresponding to free intron RNAs were readily detected in all cases (Figure 3B). Second, all intron variants are able to completely reverse splice and integrate into *E. coli* SSU rRNA. A signal

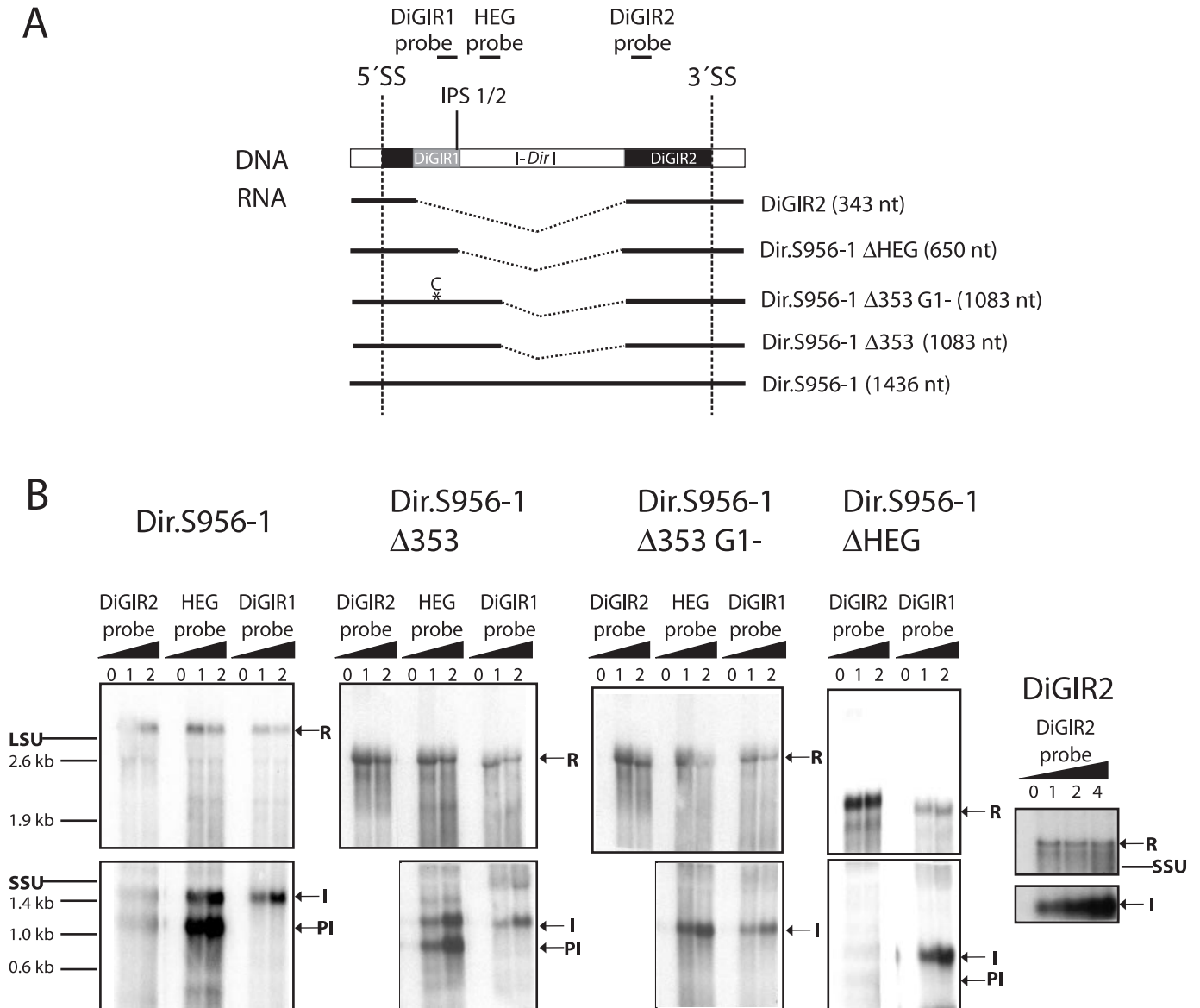


Figure 3. Reverse splicing of Dir.S956-1 (EC) and derived deletion variants in *E. coli*. (A) Schematic representation of the different intron constructs used in reverse splicing experiments. The processing sites are indicated above the illustrated DNA construct containing the Dir.S956-1 (EC) intron: 5' SS, 5' splice site; IPS1/2, internal processing sites; 3' SS, 3' splice site. The positions of the intron ribonucleotide probes used in northern blotting (B) are indicated as bars at the top of the diagram. The group I-like ribozyme DiGIR1 and the group I splicing ribozyme DiGIR2 as well as the I-DirI homing endonuclease gene (HEG) are indicated. A broken line indicates a deletion of DiGIR1 or a specific part of the HEG. The asterisk with a C above indicates a location where an engineered mutation changes the wild-type G to a C resulting in an inactive DiGIR1. The length (in nucleotides) of the different intron variants is given. (B) Northern blot analysis revealing reverse splicing (denoted by R) of Dir.S956-1 (EC) and its variants using three different ³²P-UTP-labelled RNA probes specific for DiGIR1, DiGIR2 and the HEG, respectively. Also, signals for free intron RNAs (denoted by I) and processed free introns (denoted by PI) for variants with an active DiGIR1 are detected on the blots. Lanes 0–4, 5 μg total RNA isolated from intron containing *E. coli* cells grown 0, 1, 2 or 4 h, respectively, in the presence of 1 mM IPTG. The lower parts of the blots assumed to reveal the signal for free intron RNAs and the intron processing products were hybridized separately. Approximate sizes were determined with an RNA size marker (0.28–6.58 kb; Promega) and the marker positions on the blots are indicated on the left. The positions of *E. coli* LSU (2904 nt) and SSU (1542 nt) rRNAs on the blots are also marked.

correlating with the size of SSU rRNA with an integrated intron was detected for each intron construct (Figure 3B) and indicates complete intron integration. Finally, the DiGIR1 ribozyme is active in *E. coli*. Internally processed free introns (i.e. DiGIR1 catalyses cleavage at an internal processing site) emit signals except for the intron variant with an inactive DiGIR1. These results demonstrate that the

Dir.S956-1 intron is dependent on DiGIR2, but not DiGIR1 nor the HEG, during reverse splicing.

Reverse splicing in yeast

We wanted to test for reverse splicing in eukaryote cells and expressed the *Didymium* intron in the yeast *Saccharomyces*

cerevisiae. These experiments followed an approach similar to that described for reverse splicing in *E.coli* (above). The wild-type intron, flanked by 67 bp of 5' exon and 205 bp of 3' exon, was inserted into the yeast expression vector pYESM (23) behind the *GALI* promoter. The plasmid denoted pYGAL-Dir.S956-1 was transformed into the *S.cerevisiae* strain INVSc2. With purified total RNA from galactose-induced yeast cells as template, RT-PCR reactions amplified products consistent with intron integration to site 956 (*E.coli* SSU rRNA numbering). DNA sequencing of the 3' intron-exon junction implied integration at site 956 (Figure 4A). In contrast, we were unable to amplify the corresponding 5' exon-intron integration junction (site 956). The observation that 3' integration junctions are obtained more readily than 5' integration junctions has been noted in reverse splicing experiments with the *Tetrahymena* intron (7,8). Perhaps the first step of reverse splicing is more efficient than the second in yeast. Alternatively, many reverse splicing products resplice in yeast. A northern blot analysis on yeast total RNA, using the HEG-specific probe (Figure 3A), revealed a signal for an RNA of ~2 kb (Figure 4B) The northern blot analysis was also conducted with the DiGIR2-specific probe giving the same signal pattern (data not shown). The size expected for a reverse splicing intermediate where the 3' end of Dir.S956-1 is ligated to U957 of the yeast SSU rRNA (the first step of reverse splicing) is 2052 nt. A signal for such an intermediate was not detected in the reverse splicing experiments in *E.coli*, and indicates a reduced efficiency of the second step of reverse splicing in yeast compared to bacteria. A weak signal (also detected with the DiGIR2 probe) correlating with the size of yeast SSU rRNA harbouring the *Didymium* intron implies complete integration of Dir.S956-1 into yeast SSU rRNA (Figure 4B). However, due to failure to

amplify the 5' integration junction by RT-PCR, we can only conclude partial reverse splicing in yeast.

In vitro integration of circular intron RNAs

It has previously been speculated that full-length intron RNA circles may have a role in intron horizontal transfers (13,25). The DiGIR2 ribozyme catalyses the formation of RNA circles that contain the entire intron sequence (12,13). We investigated the potential for these circles to integrate *in vitro* into a substrate derived from the *E.coli* SSU rRNA (positions 800–1115). For simplicity, we used the minimal DiGIR2 intron construct (i.e. DiGIR1 and HEG deleted, see Figure 3A) that efficiently catalyses intron full-length circle formation as well as reverse splicing (above). As a control, we tested in parallel *in vitro* reverse splicing of linear DiGIR2 intron RNAs. Equal amounts (5 pmol) of PAGE-purified substrate RNA (part of *E.coli* SSU rRNA containing S956), and PAGE-purified circular or linear DiGIR2 (EC) RNAs (with the IGS specific for S956) were incubated under conditions that favour reverse splicing [see (5); Materials and Methods]. After 120 min of incubation, the RNAs were used as templates in RT-PCR reactions as described above. Products for 5' and 3' integration junctions (318 and 238 bp, respectively) were amplified for both the circular and linear intron RNAs (Figure 5), and subsequent sequencing analysis implied intron integration after nucleotide U956 in the *E.coli* SSU rRNA transcript (Figure 5). An additional RT-PCR product of 289 bp (Figure 5) indicated partial intron integration after position 905 in the SSU rRNA substrate (i.e. only the 3' intron-exon junction was found). To summarize, these results suggest that DiGIR2 (EC) circles as well as the linear form of intron are able to integrate *in vitro* into *E.coli* SSU rRNA substrate.

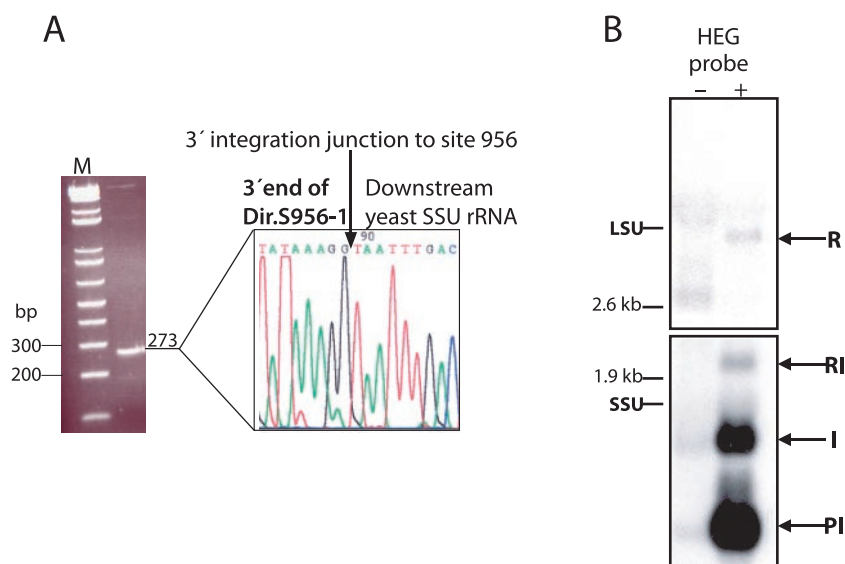


Figure 4. Detection of Dir.S956-1 (WT) reverse splicing in *S.cerevisiae*. (A) RT-PCR amplification on yeast total RNA using primers OP85 and OP1009 resulting in a product of 273 bp. Sequencing of this product confirms the 3' integration junction of Dir.S956-1 to site 956 of yeast SSU rRNA. (B) Northern blot analysis on 5 µg total RNA isolated from Dir.S956-1 containing yeast cells. The blot with a ^{32}P -UTP labelled HEG riboprobe (indicated in Figure 3A) reveals several intron containing signals: R, reverse spliced product; RI, reverse spliced intron intermediate where the 3' end of the intron is ligated to the yeast SSU rRNA 3' exon; I, free intron RNA; PI, DiGIR1 processed free intron. The lower part of the blot assumed to reveal the signal for free intron RNA and intron processing products was hybridized separately. Lanes denoted with a plus indicate that the total RNA is isolated from galactose-induced Dir.S956-1 containing yeast cells and lanes denoted with a minus indicate lack of galactose induction. The positions of LSU (3393 nt) and SSU (1789) yeast rRNAs on the blot as well as the positions of two molecular marker signals (from the 0.28 to 6.58 kb RNA size marker; Promega) are indicated on the left.

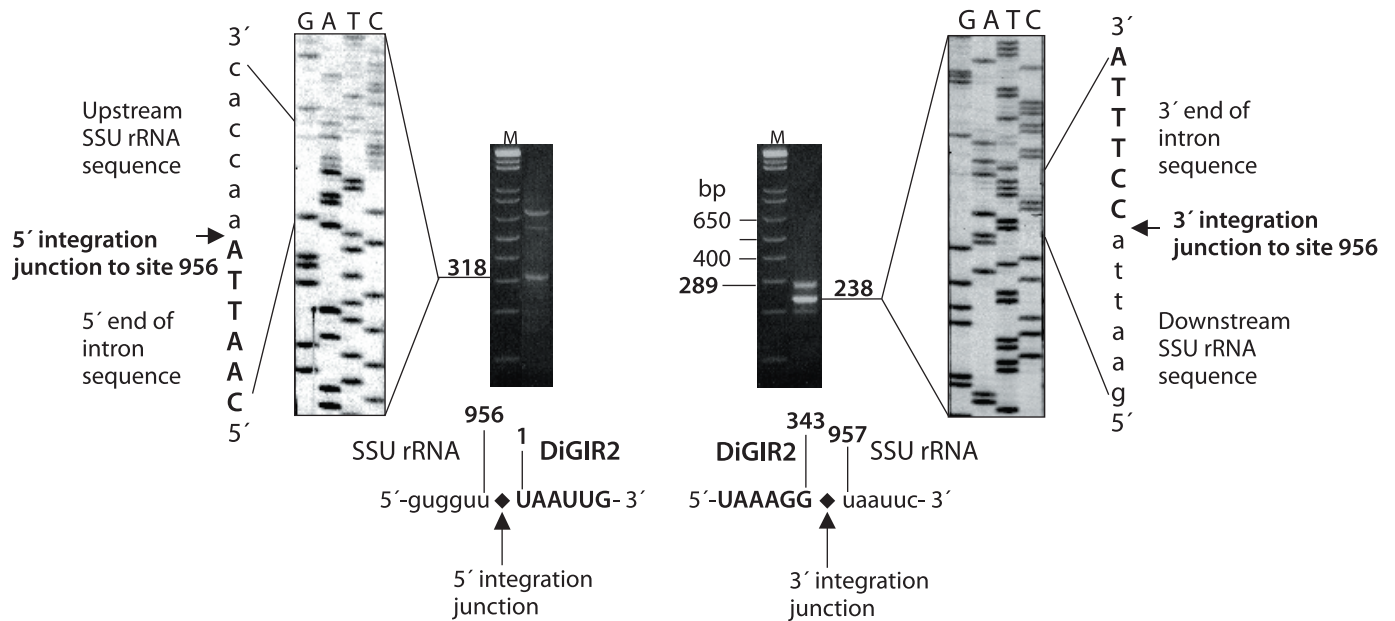


Figure 5. Detection of *in vitro* integration of DiGIR2 (EC) circular or linear intron RNAs into an *E. coli* rRNA substrate by RT-PCR and sequencing: 5 pmol PAGE-purified linear or circular DiGIR2 (EC) and 5 pmol PAGE-purified substrate RNA containing S956 were incubated under reverse splicing conditions [40 mM Tris-HCl (pH 7.5), 50 mM MgCl₂, 200 mM KCl, 2 mM spermidine, 5 mM DTT, 50°C]. After 120 min incubation time, products for 5' and 3' intron integration junctions (318 and 238 bp, respectively) were amplified by RT-PCR for both linear and circular DiGIR2 RNAs. The RT-PCR products presented here are from the experiment with the linear intron, but a similar band pattern was obtained for the intron circle. The primer pairs used for amplification of intron integration junctions were OP621 and OP236 for the 5' junction, and OP622 and OP85 for the 3' junction. The RNA sequences flanking the observed integration junctions (denoted with a diamond) at site 956 are given with the intron sequence marked in bold capital letters and the rRNA sequence in lower case letters. The product of 289 bp represents partial reverse splicing to site 905 in the rRNA. Other visible RT-PCR products are caused by non-specific primer annealing during the RT reaction. M, size marker: 1 kb Plus DNA Ladder (Gibco BRL).

DISCUSSION

We investigated reverse splicing of the mobile twin-ribozyme group I intron Dir.S956-1 into *E. coli* and *S. cerevisiae* SSU rRNA. The intron was expressed from plasmids in bacteria and in yeast. Although we searched for intron integration products along the entire bacterial SSU rRNA, reverse splicing appeared to be restricted to site 956. This site is homologous to the natural intron location in the *Didymium* genome. Reverse splicing of the *Didymium* intron was dependent on the DiGIR2 splicing ribozyme, but not the DiGIR1 ribozyme nor the I-DiI1 homing endonuclease gene. Interestingly, results from *in vitro* analyses indicate that full-length intron RNA circles also promote intron integration into *E. coli* rRNA targets.

Site-specific reverse splicing

Our results provide evidence for site-specific reverse splicing of the *Didymium* intron into *E. coli* SSU rRNA. These findings are contrasted by studies in which the *Tetrahymena* intron was expressed in *E. coli* (7,8). The *Tetrahymena* intron was targeted to site 1926 in the *E. coli* LSU rRNA (L1926) to accommodate for the 'U' requirement at the site of insertion (position L1925 is homologous to the natural insertion site in *Tetrahymena*, but this nucleotide is a 'C' in *E. coli*). In addition to complete integration of the *Tetrahymena* intron into L1926, partial reverse splicing was detected at several novel rRNA sites (7). The integration to these sites was explained by a 4–13 nt match between the intron IGS [including P10 (4 bp in the natural ribozyme)] and the sequences that surround each

of the novel integration sites. In our study with the *Didymium* intron, the *in vivo* data showed that reverse splicing of the *Didymium* intron was restricted to S956.

How can we explain the differences between the *Tetrahymena* and *Didymium* reverse splicing data? One explanation is that the LSU rRNA provides a target RNA that is almost twice as large as the SSU rRNA (2904 nt compared to 1542 nt of SSU rRNA), and therefore contains more sites that potentially can base pair with the intron IGS. The *E. coli* LSU rRNA contains five sites that are fully complementary to the mutated *Tetrahymena* IGS (including the targeted site) whereas the bacterial SSU rRNA contains only one site (S956) with complete match to the altered *Didymium* IGS. It is unlikely however that this alone can explain the observed differences. Reverse splicing of the *Tetrahymena* intron was detected at just three of the five sites with full complementarity to the IGS and the intron also targets sites with as little as four nucleotide match (7). Furthermore, the bacterial SSU rRNA contains 11 sites that are able to make four or more base pairs with the mutated *Didymium* IGS. Alternatively, secondary structures in LSU rRNA have been reported to compete with the reverse splicing reaction of the *Tetrahymena* intron into its natural cognate L1925 site (26,27). With L1926 hidden in the LSU rRNA structure *in vivo*, the *Tetrahymena* intron is blocked from integrating into this site and could account for lower integration efficiency at that site. Finally, the *Tetrahymena* and *Didymium* ribozymes represent the distantly related group IC1 and group IE, respectively, with significant differences in domain organization, hydrolysis, RNA circularization, and *in vitro* splicing (10,12–14). Exactly how

these differences can contribute to discrepancy in the observed reverse splicing outcome is unclear. Whatever the reasons are, our findings of reverse splicing into a single site that corresponds to the natural intron homing site correlate very well with numerous phylogenetic analyses that support horizontal group I intron transfers almost exclusively into the homologous site (3,21,22,28). We conclude that the observed distribution of rDNA group I introns among fungi and protists could be explained by a combination of two mobility mechanisms: ribozyme-dependent reverse splicing that results in intron integration into a homologous site in a new organism (horizontal intron transfer) and the homing endonuclease-dependent intron mobility mediating efficient unidirectional spread throughout the population.

Perspectives

Reverse splicing has been ascribed a role in horizontal transfer of group I introns, and full-length circular introns have been suggested in group I intron mobility (13). Full-length circles (FLCs) are common among nuclear group I introns, and they carry all the genetic information of the intron (12,13,25). Our data suggest a link between FLCs and intron integration at the RNA level. The *in vitro* experiments with DiGIR2 imply integration of circular intron RNAs into the *E. coli* SSU rRNA substrate similar to reverse splicing with the linear version of the intron. Assuming an unknown mechanism by which the circle performs the integration reaction, one could depict the following scenario for the *Didymium* intron: full-length intron RNA circles are formed at the expense of the host genome in order to preserve the intron, e.g. during crisis of the host. In fact, previous studies in *Didymium* conclude that Dir.S956-1 intron FLCs accumulate during amoebae starvation and encystment, probably due to an increased stability of the circular RNA molecule (17). The intron RNA circles will then be presented to a new host by, for example, feeding conditions, symbiosis or viral infections. Integration of circular intron RNAs into foreign SSU rRNA, followed by genomic integration, positions the intron at a site homologous to its natural splice junction. Finally, the intron encoded homing endonuclease will ensure efficient spread of the intron within the population. In this manner, the biological role of intron FLCs would be intron horizontal transfer mediated by intron integration at the RNA level.

ACKNOWLEDGEMENTS

We thank Kari Haugli and Anna Vader for valuable advice during the initial work, and Henrik Nielsen and Peik Haugen for critical reading of the manuscript. This work was supported by grants from The Norwegian Research Council and Norwegian Institute of Gene Ecology. Funding to pay the Open Access publication charges for this article was provided by University of Tromsø, IMB.

Conflict of interest statement. None declared.

REFERENCES

- Johansen,S., Muscarella,D.E. and Vogt,V.M. (1996) Insertion elements in ribosomal DNA. In Zimmermann,R.A. and Dahlberg,A.E. (eds), *Ribosomal RNA: Structure, Evolution, Processing, and Function in Biosynthesis*. CRC Press, Boca Raton, pp. 89–108.

- Cannone,J.J., Subramanian,S., Schnare,M.N., Collett,J.R., D'Souza,L.M., Du,Y., Feng,B., Lin,N., Madabusi,L.V., Muller,K.M., Pande,N., Shang,Z., Yu,N. and Gutell,R.R. (2002) The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BioMed. Central Bioinformatics*, **3**, 1–31.
- Jackson,S.A., Cannone,J.J., Lee,J.C., Gutell,R.R. and Woodson,S.A. (2002) Distribution of rRNA introns in the three-dimensional structure of the ribosome. *J. Mol. Biol.*, **323**, 35–52.
- Belfort,M. and Roberts,R.J. (1997) Homing endonucleases: keeping the house in order. *Nucleic Acids Res.*, **25**, 3379–3388.
- Woodson,S.A. and Cech,T.R. (1989) Reverse self-splicing of the *Tetrahymena* group I intron: implication for the directionality of splicing and for intron transposition. *Cell*, **57**, 335–345.
- Roman,J. and Woodson,S.A. (1995) Reverse splicing of the *Tetrahymena* IVS: evidence for multiple reaction sites in the 23S rRNA. *RNA*, **1**, 478–490.
- Roman,J. and Woodson,S.A. (1998) Integration of the *Tetrahymena* group I intron into bacterial rRNA by reverse splicing *in vivo*. *Proc. Natl Acad. Sci. USA.*, **95**, 2134–2139.
- Roman,J., Rubin,M.N. and Woodson,S.A. (1999) Sequence specificity of *in vivo* reverse splicing of the *Tetrahymena* group I intron. *RNA*, **5**, 1–13.
- Johansen,S. and Haugen,P. (2001) A new nomenclature of group I introns in ribosomal DNA. *RNA*, **7**, 935–936.
- Decatur,W.A., Einvik,C., Johansen,S. and Vogt,V.M. (1995) Two group I ribozymes with different functions in a nuclear rDNA intron. *EMBO J.*, **14**, 4558–4568.
- Einvik,C., Elde,M. and Johansen,S. (1998) Group I twintrons: genetic elements in myxomycete and schizopyrenid amoeboflagellate ribosomal DNAs. *J. Biotechnol.*, **64**, 63–74.
- Johansen,S. and Vogt,V.M. (1994) An intron in the nuclear ribosomal DNA of *Didymium iridis* codes for a group I ribozyme and a novel ribozyme that cooperate in self-splicing. *Cell*, **76**, 725–734.
- Nielsen,H., Fiskaa,T., Birgisdottir,A.B., Haugen,P., Einvik,C. and Johansen,S. (2003) The ability to form full-length intron RNA circles is a general property of nuclear group I introns. *RNA*, **9**, 1464–1475.
- Haugen,P., Andreassen,M., Birgisdottir,A.B. and Johansen,S. (2004) Hydrolytic cleavage by a group I intron ribozyme is dependent on RNA structures not important for splicing. *Eur. J. Biochem.*, **271**, 1015–1024.
- Einvik,C., Nielsen,H., Westhof,E., Michel,F. and Johansen,S. (1998) Group I-like ribozymes with a novel core organization perform obligate sequential hydrolytic cleavages at two processing sites. *RNA*, **4**, 530–541.
- Vader,A., Nielsen,H. and Johansen,S. (1999) *In vivo* expression of the nucleolar group I intron-encoded I-DirI homing endonuclease involves the removal of a spliceosomal intron. *EMBO J.*, **18**, 1003–1013.
- Vader,A., Johansen,S. and Nielsen,H. (2002) The group I-like ribozyme DiGIR1 mediates alternative processing of pre-rRNA transcripts in *Didymium iridis*. *Eur. J. Biochem.*, **269**, 5804–5812.
- Johansen,S., Einvik,C. and Nielsen,H. (2002) DiGIR1 and NaGIR1: naturally occurring group I-like ribozymes with unique core organization and evolved biological role. *Biochimie*, **84**, 905–912.
- Johansen,S., Elde,M., Vader,A., Haugen,P., Haugli,K. and Haugli,F. (1997) *In vivo* mobility of a group I twintrons in nuclear ribosomal DNA of the myxomycete *Didymium iridis*. *Mol. Microbiol.*, **24**, 737–745.
- Nikoh,N. and Fukatsu,T. (2001) Evolutionary dynamics of multiple group I introns in nuclear ribosomal RNA genes of endoparasitic fungi of the genus *Cordyceps*. *Mol. Biol. Evol.*, **18**, 1631–1642.
- Haugen,P., Runge,H.J. and Bhattacharya,D. (2004) Long-term evolution of the S788 fungal nuclear small subunit rRNA group I introns. *RNA*, **10**, 1084–1096.
- Lundblad,E.W., Einvik,C., Ronning,S., Haugli,K. and Johansen,S. (2004) Twelve group I introns in the same pre-rRNA transcript of the myxomycete *Fulligo septica*: RNA processing and evolution. *Mol. Biol. Evol.*, **21**, 1283–1293.
- Decatur,W.A., Johansen,S. and Vogt,V.M. (2000) Expression of the *Naegleria* intron endonuclease is dependent on a functional group I self-cleaving ribozyme. *RNA*, **6**, 616–627.
- Gietz,R.D. and Woods,R.A. (1994) High efficiency transformation in yeast. In Johnston,J.A. (ed.), *Molecular Genetics of Yeast: Practical Approaches*. Oxford University Press, Oxford, pp. 121–134.

25. Haugen,P., De Jonckheere,J.F. and Johansen,S. (2002) Characterization of the self-splicing products of two complex *Naegleria* LSU rDNA group I introns containing homing endonuclease genes. *Eur. J. Biochem.*, **269**, 1641–1649.
26. Woodson,S.A. and Cech,T.R. (1991) Alternative secondary structures in the 5' exon affect both forward and reverse self-splicing of the *Tetrahymena* intervening sequence RNA. *Biochemistry*, **30**, 2042–2050.
27. Woodson,S.A. and Emerick,V.L. (1993) An alternative helix in the 26S rRNA promotes excision and integration of the *Tetrahymena* intervening sequence. *Mol. Cell. Biol.*, **13**, 1137–1145.
28. Haugen,P., Coucheron,D.H., Ronning,S.B., Haugli,K. and Johansen,S. (2003) The molecular evolution and structural organization of self-splicing group I introns at position 516 in nuclear SSU rDNA of myxomycetes. *J. Euk. Microbiol.*, **50**, 283–292.