Gene targeting using randomly inserted group II introns (targetrons) recovered from an *Escherichia coli* gene disruption library

Jun Yao, Jin Zhong and Alan M. Lambowitz*

Institute for Cellular and Molecular Biology, Department of Chemistry and Biochemistry, and Section of Molecular Genetics and Microbiology, School of Biological Sciences, University of Texas at Austin, Austin, TX 78712, USA

ABSTRACT

The *Lactococcus lactis* Ll.LtrB group II intron retro-homes by reverse-splicing into one strand of a double-stranded DNA target site, while the intron-encoded protein cleaves the opposite strand and uses it to prime reverse transcription of the inserted intron RNA. The protein and intron RNA function in a ribonucleoprotein particle, with much of the DNA target sequence recognized by base-pairing of the intron RNA. Consequently, group II introns can be reprogrammed to insert into specific or random DNA sites by substituting specific or random nucleotide residues in the intron RNA. Here, we show that an *Escherichia coli* gene disruption library obtained using such randomly inserting Ll.LtrB introns contains most viable *E.coli* gene disruptions. Further, each inserted intron is targeted to a specific site by its unique base-pairing regions, and in most cases, could be recovered by PCR and used unmodified to obtain the desired single disruptant. Additionally, we identified a subset of introns that insert at sites lacking T+5, a nucleotide residue critical for second-strand cleavage. All such introns tested individually gave the desired specific disruption, some by switching to an alternate retrohoming mechanism targeting single-stranded DNA and using a nascent lagging DNA strand to prime reverse transcription.

INTRODUCTION

Mobile group II introns use a remarkable mobility mechanism, termed retrohoming, in which the excised intron RNA uses its ribozyme activity to insert directly into a DNA target site by reverse-splicing and is then reverse-transcribed by the intron-encoded protein (IEP) (1,2). Retrohoming is mediated by a ribonucleoprotein (RNP) complex that contains the IEP and the excised intron RNA, with both used for DNA target site recognition (3,4). Because the IEP recognizes only a small number of fixed positions and most of the target specificity comes from base-pairing of the intron RNA to the DNA target sequence, it is possible to reprogram group II introns to insert into desired sites simply by modifying the intron RNA (5,6). This feature combined with their very high insertion frequencies and specificity have made it possible to develop mobile group II introns into gene targeting vectors, dubbed 'targetrons' [reviewed in (7)]. A targetron derived from the *Lactococcus lactis* Ll.LtrB intron has been used for chromosomal gene disruption and site-specific DNA insertion in a variety of Gram-negative and Gram-positive bacteria (8–12).

The RNPs that mediate retrohoming are formed when the IEP binds to the intron in unspliced precursor RNA and promotes its splicing by stabilizing the catalytically active RNA structure; it then remains bound to the excised intron lariat RNA in a stable RNP complex (13–15). To initiate mobility, group II intron RNPs bind DNA and recognize target sites (16). The DNA target site for the *Lactococcus lactis* Ll.LtrB intron used in the present work is shown in Figure 1A (4,10). The initial recognition event is thought to involve major groove interactions between the IEP and a small number of specific bases in the distal 5′-exon region, including T–23, G–21 and A–20. These base interactions, bolstered by neighboring phosphate-backbone interactions and possibly minor groove interactions between positions −18 and −14, lead to local DNA unwinding, enabling the intron RNA to base-pair to the adjacent 14–16 nt DNA sequence for reverse-splicing into the intron-insertion site. The intron RNA sequences involved in base-pairing are located in two RNA
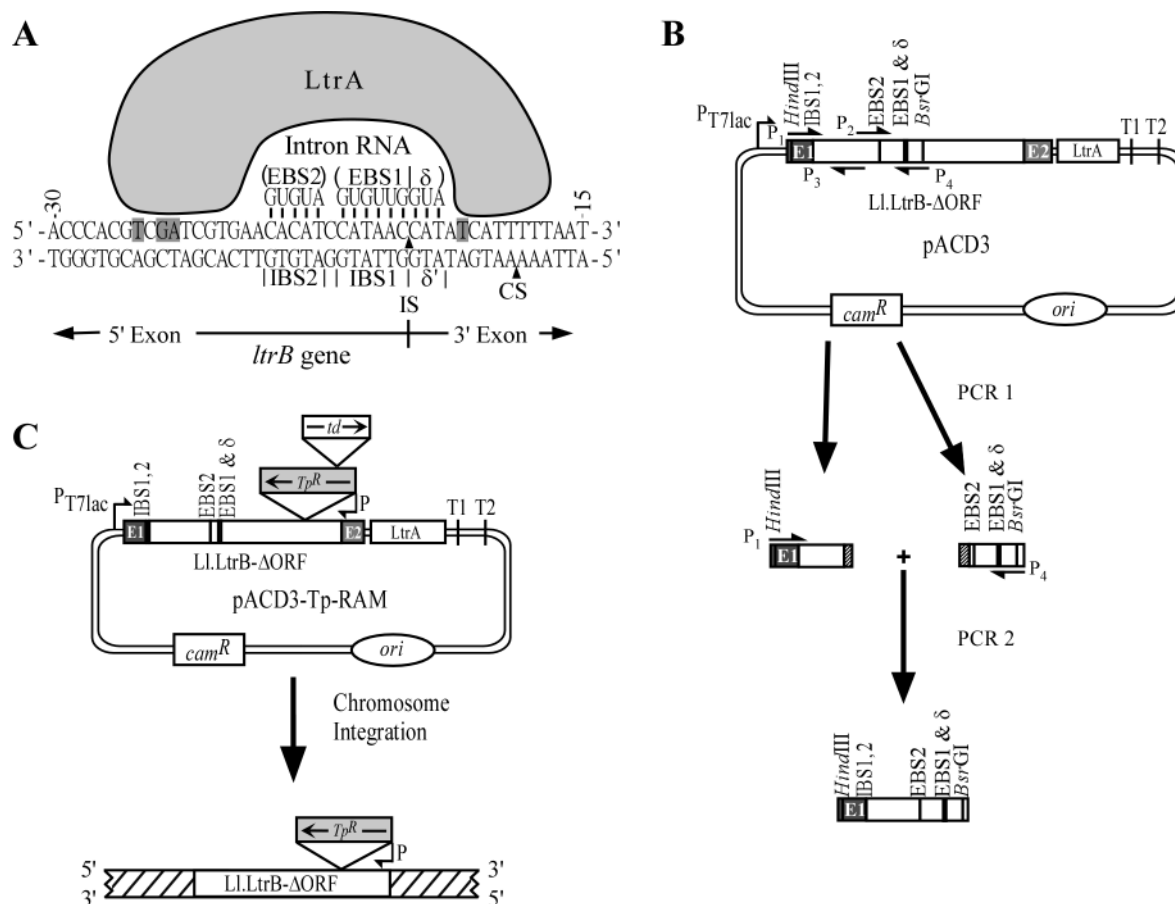
Figure 1. DNA target site interactions of the *L.lactis* Ll.LtrB intron and donor plasmids for expression of retargeted introns. (**A**) The DNA target site for the Ll.LtrB intron is recognized by an RNP containing the IEP (LtrA protein) and excised intron lariat RNA, with both the protein and base-pairing of the intron RNA used to recognize the target sequence. Key bases recognized by the IEP (gray shading) include T−23, G−21 and A−20 in the 5′ exon and T+5 in the 3′ exon; T+5 is required only for second-strand cleavage (5,6,10,12). The intron RNA's EBS2, EBS1 and δ sequences base-pair with IBS2, IBS1 and δ′ sequences located between DNA target site positions −12 and +3. IS and CS (arrowheads) indicate the intron-insertion site and second-strand cleavage site, respectively. (**B**) Intron-donor plasmid pACD3 contains a 0.9 kb Ll.LtrB-ΔORF intron and flanking exons cloned downstream of a T7lac promoter, with the LtrA protein expressed from a position just downstream of the 3′ exon (9). The intron is retargeted by modifying its EBS2, EBS1 and δ sequences to base-pair to the IBS2, IBS1 and δ′ sequences in the DNA target site. The IBS1 and IBS2 sequences in the donor plasmid are also modified to base-pair to the intron's retargeted EBS1 and EBS2 sequences for efficient RNA splicing. The required modifications are introduced into the donor plasmid via a two-step PCR using primers P1 to P4, as diagrammed in the figure (5,9). (**C**) The intron-donor plasmid pACD3-Tp-RAM contains a retrotransposition-activated *Tp*R gene (*Tp*R-RAM) (12). The *Tp*R gene is inserted in group II intron DIV in the reverse orientation, but interrupted by an efficiently self-splicing group I intron (the phage T4 *td* intron) in the forward orientation. During retrotransposition via an RNA intermediate, the group I intron is spliced reconstituting the marker, which can then be selected after integration into a DNA target site.

stem–loops and are denoted exon-binding sites 1 and 2 (EBS1 and EBS2) and δ, while the complementary DNA target-site sequences are denoted intron-binding sites 1 and 2 (IBS1 and IBS2) in the 5′ exon and δ′ in the 3′ exon. Second-strand cleavage to generate the primer for reverse transcription requires additional interactions between the IEP and the 3′ exon, the most critical being recognition of T+5 (4,6). In the absence of second-strand cleavage, the Ll.LtrB intron has been shown to retrohome at low frequency by using nascent leading or lagging strands at DNA replication forks to prime reverse-transcription (17), and such mechanisms are also used for retrohoming by group II introns that encode proteins lacking the DNA endonuclease (En) domain (referred to as En⁻ introns) (18). In all cases, the IEP is thought to synthesize a full-length intron cDNA, which is then integrated into the recipient DNA by a DNA repair mechanism independent of homologous recombination (19–21).

For use in gene targeting, the Ll.LtrB intron (targetron) is expressed from a donor plasmid, such as pACD3 (Figure 1B) (5,9). This plasmid uses an inducible T7lac promoter to express a ΔORF-derivative of the intron and short flanking exons, with the LtrA ORF cloned just downstream of the 3′ exon. The LtrA protein expressed from this position still binds to the intron to promote RNA splicing and mobility, but when the ΔORF intron integrates at a new location, it is unable to splice in the absence of the IEP, yielding a gene disruption. Currently, the targetron is programmed to insert into different sites with the help of a computer algorithm, which scans the target sequence for the best matches to the positions recognized by the IEP and then designs primers from modifying the intron's EBS1, EBS2 and δ sequences to insert into those sites (10). The positions recognized by the IEP are sufficiently few and flexible that the algorithm readily identifies multiple rank-ordered target sites in any gene. The IBS1 and IBS2

sequences in the 5′ exon of the donor plasmid are also modified to be complementary to the retargeted EBS1 and EBS2 sequences for efficient RNA splicing. The necessary modifications are introduced into the donor plasmid by a two-step PCR, using three unique primers (P_1, P_2 and P_4) and one 'fixed' primer (P_3; Figure 1B).

Targetrons that insert at the desired site can be detected by a resulting phenotypic change, by colony PCR or by using a genetic marker inserted in intron domain IV, a non-essential region that extends outside of the intron's catalytic core. Particularly useful for gene targeting is a Retrotransposition Activated Marker (RAM) (12); also referred to as a Retrotransposition-Indicator Gene or RIG (22). The RAM is a selectable gene, such as trimethoprim-resistance (Tp^R), inserted in the group II intron in the reverse orientation, but interrupted by an efficiently self-splicing group I intron in the forward orientation (Figure 1C). During retrotransposition via an RNA intermediate, the group I intron is spliced allowing direct selection of the marker after DNA integration (12).

In addition to targeted disruption, the incorporation of a RAM makes it possible to use a targetron with randomized target-site recognition (EBS1, EBS2 and δ) sequences to obtain a gene disruption library in which group II introns integrate at sites distributed throughout a genome, analogous to global transposon mutagenesis (12). Here, we show that an *E.coli* gene disruption library obtained using this approach contains a high proportion of all viable disruptions. Further, unlike a conventional transposon, each of the inserted targetrons in the library is potentially targeted to a specific location by virtue of its unique EBS1, EBS2 and δ sequences, and in most cases, could be recovered ('fished') by PCR and inserted into a donor plasmid to give the desired single disruptant. Additionally, we identified a subset of introns that could be targeted to sites that lack T+5, including two that retrohome at relatively high frequency, apparently by targeting single-stranded DNA and using a nascent lagging DNA strand to prime reverse-transcription.

MATERIALS AND METHODS

Bacterial strains and growth condition

E.coli HMS174 (DE3) (F^- *recA hsdR Rif*^{R}) (Novagen, Madison, WI), which contains an isopropyl-β-D-thiogalactoside (IPTG)-inducible phage T7 RNA polymerase, was used for generating the chromosomal gene disruption library and for individual gene disruption experiments. DH5α was used for cloning. LB medium was used for most experiments, with antibiotics added as required at the following concentrations: ampicillin, 100 μg/ml; chloramphenicol, 25 μg/ml; and tetracycline, 25 μg/ml. Trimethoprim selection was done in Mueller–Hinton or M9 medium supplemented with 10 μg/ml trimethoprim and 1 μg/ml thymine, which was found necessary for selection of the chromosomally integrated Tp^R marker (12).

Recombinant plasmids

pACD3 contains a 0.9-kb Ll.LtrB-ΔORF intron and flanking exons cloned behind a T7lac promoter, with the IEP expressed from a position just downstream of the 3′ exon (9). pACD3-Tp-RAM (previously denoted pACD3-RAM) is a derivative of pACD3 in which a retrotransposition-activated Tp^R marker was inserted in intron domain IV (12). pACD3E and pACD3E-Tp-RAM are derivatives of the above plasmids in which 5′ exon-position −18 to intron position +289 (positions numbered from the 5′-splice site) were deleted and replaced with an *Eco*RV site.

The *E.coli* gene disruption library was generated using a pACD3-Tp-RAM donor intron modified to contain randomized nucleotide residues at EBS2 positions −12 to −8, EBS1 positions −6 to −1 and δ positions +1 to +3 (12). The corresponding IBS1 and IBS2 positions in the 5′ exon of the donor plasmid were also randomized to obtain nucleotide combinations complementary to the randomized EBS1 and EBS2 sequences for RNA splicing.

Mobility assays to determine the effect of DNA target site orientation on retrohoming frequency were done with donor plasmid pACD2X, which contains a 940 nt Ll.LtrB-ΔORF with a phage T7 promoter inserted near its 3′ end (23). The recipient plasmids were derivatives of pBRR3A-ltrB and pBRR3B-ltrB, which contain the Ll.LtrB target site upstream of a promoterless tet^R gene in opposite orientations relative to the direction of plasmid DNA replication (17). DNA target sites to be tested (positions −30 to +15 from the intron-insertion site) were synthesized as double-stranded DNA oligonucleotides with appended *Aat*II and *Eco*RI sites and swapped for the wild-type Ll.LtrB target site between the corresponding sites of pBRR3A-ltrB and pBRR3B-ltrB.

E.coli chromosomal gene disruption library

The *E.coli* gene disruption library was generated by using the pACD3-Tp-RAM-based donor intron with randomized EBS2, EBS1 and δ sequences, as described previously (12). Briefly, 5 μg of the donor plasmid was electroporated into *E.coli* HMS174(DE3), and transformants were grown overnight at 37°C in LB medium containing chloramphenicol: 10 ml of the overnight culture was then inoculated in 250 ml of fresh LB medium, and cells were grown to log phase for 3 h at 37°C. To induce intron expression, a 15 ml sample of these cells was added to 250 ml of LB containing 500 μM IPTG and incubated at 30°C for 18 h. A 10 ml sample of these cells was then washed and grown to saturation in 250 ml of Mueller–Hinton medium containing trimethoprim (10 μg/ml) plus thymine (1 μg/ml). Chromosomal DNA was isolated from 30 ml of the culture using a Genomic-tip and Genomic DNA Buffer set (Qiagen, Valencia, CA).

Targetron fishing

Targetrons were 'fished' from the *E.coli* gene disruption library DNA by PCR. The PCR for targetrons inserted into the sense and antisense strands used target gene-specific primers (P_s and P_a), which base-pair to 5′- and 3′-flanking sequences, respectively, together with a fixed intron primer (P_i) 5′-TCAGATTCTCGGCATCGCTTTCGTTC, which base-pairs to a sequence downstream of EBS1 (Figure 3). The PCRs were carried out in 40 μl of reaction medium containing 4 U of PlatinumTaq (Invitrogen, Carlsbad, CA), 1 μg of *E.coli* chromosomal DNA and 1 μM of each primer for 30 cycles, with the annealing temperatures and extension times optimized for the size of each target gene and the T_m for each

primer. PCR products were gel-purified and sequenced. For cloning into the donor plasmid, the PCR products were digested with *Bsr*GI plus a second restriction enzyme that generates a blunt end within the 5′-exon of the target gene, and then cloned into the vector backbones of pACD3 or pACD3-Tp-RAM by ligating to *Eco*RV + *Bsr*GI-digested pACD3E or pACD3E-Tp-RAM, respectively. Target gene-specific primers used in this study are summarized in Table 1.

Use of targetrons for chromosomal gene disruption

For chromosomal gene disruptions, pACD3-based plasmids containing fished targetrons were transformed into *E.coli* HMS174(DE3) and grown overnight in LB medium containing chloramphenicol: 50 μl of the overnight culture was inoculated into 5 ml of fresh LB medium containing chloramphenicol and grown at 37°C until OD_{598} was 0.2, then induced with 500 μM IPTG for 18 h at 30°C (12). The cells were washed once with fresh LB by centrifugation at 3750 *g* for 5 min at 4°C, resuspended in 5 ml of fresh LB, and then plated on LB and incubated overnight at 37°C. Disruptants were identified by colony PCR using primers flanking the gene and confirmed by DNA sequencing. Gene disruptions with

Table 1. Primers used for targetron fishing

Primer	Sequence
AraDP$_s$	5′-CGGAATTCAACATGTTAGAAGATCTCAAA-3′
AraDP$_a$	5′-CGGAATTCATTACTGCCCGTAATATGC-3′
CorAP$_s$	5′-CGTGGAATTCATGCTGAGCGCATTTCAACTGGAA-3′
CorAP$_a$	5′-GGCGCTGCAGTTACAACCAGTTCTTCCGCTTAAAG-3′
LacZP$_s$	5′-ATCCTGCAGCGGATAACAATTTCACACAG-3′
LacZP$_a$	5′-ATCCTGCAGACATGGCCTGCCCGG-3′
LhrP$_s$	5′-TCCCCCCCCCCGGGCAGATAATCCAGACCCT-TCATCGC-3′
LhrP$_a$	5′-TCCCCCCCCCGGGCGAGTAATTCCTCTTTTTCCTG-3′
MgtAP$_s$	5′-CGTGGAATTCATGTTTAAAGAAATTTTTACCCGGCT-3′
MgtAP$_a$	5′-GGCGCGGCAGTTATTGCCAGCCGTAACGACGGC-3′
NarQP$_s$	5′-GAAGATATCTTTCAGCGAACTGGAACATTAATG-3′
NarQP$_a$	5′-GAAGATATCTATTCTTTGGTCAGTAGGAGGCATTAC-3′
RecDP$_s$	5′-GAACCCGGGAGGAGGCGTAATGAAATTGC-3′
RecDP$_a$	5′-GAACCCGGGCGTTTATTCCCGTGAACTAAACAACG-3′
RraAP$_a$	5′-GAATGCGCAGCGTAATTTCTTTACTGTGTAGGCC-AACAGG-3′
RrmJP$_s$	5′-CATCCCGGGCCTGCGGAAAACCCCTCGTTTTACAC-3′
RuvAP$_s$	5′-AGAGATATCCTGCTTCAATCATCCTTTACCTC-ATAACGC-3′
RuvAP$_a$	5′-AGAGATATCCTGCTTCAATCATCCTTTACCT-CATAACGC-3′
RuvBP$_s$	5′-GAGATATCACCACTTTGCCGGAAGATGTA-GCAGATC-3′
RuvBP$_a$	5′-AGAGATATCACCTACGGTTCGGTATCTCTG-GTAGTCC-3′
TrpEP$_s$	5′-ATCCTGCAGCCCGCCTAATGAGCG-3′
UvrBP$_s$	5′-GAACCCGGGATGAGTAAACCGTTCAAACTGAATTC-3′
UvrBP$_a$	5′-GAACCCGGGTGTTACGATGCCGCGATAAAC-3′
YbcJP$_s$	5′-GAAGATATCCCACGCTGATTGAAAACACGCT-ACAGG-3′
YcdVP$_a$	5′-AGAGATATCTTTTCAAAAAGTTTGTTGACCT-CAGGTC-3′
YgjHP$_a$	5′-AGAGATATCGTAAGGCGCTGGCGGCCATGCG-AGTAAG-3′
YhbYP$_s$	5′-CATCCCGGGGACCCGTTTTCAATCCCAACGTA-AGCAAA-3′
YhbYP$_a$	5′-CATCCCGGGCCTGCGGAAAACCCCTCGTTTTACAC-3′

Primers are indicated by the gene name followed by s or a, for those used to fish targetrons inserted into the sense or antisense strand, respectively.

pACD3-Tp-RAM-based donor plasmids were done similarly, except that after induction, the cells were grown in 5 ml of M9 medium without antibiotics for 2 h at 30°C, and then plated on M9 medium containing trimethoprim plus thymine, and incubated for 2–3 days at 30°C.

PCR for direct insertion of fished targetrons into donor plasmids without cloning

Targetrons were fished from the chromosomal gene disruption library as above, by using the intron primer P_i, together with a gene-specific primer, which is complementary to 5′- or 3′-flanking sequences, but with an additional 27-nt vector sequence (5′-CCATTCCCCTCTAGAAAAAAGCTTCGT) appended to its 5′ end. The vector sequence is located upstream of the *Eco*RV site of pACD3-Tp-RAM-PCR. The PCR products were purified by using a PCR cleanup kit (Qiagen) and resuspended in 50 μl of distilled water. The first-round PCR products (2 μl) were then combined with *Eco*RV-digested pACD3-Tp-RAM-PCR by an additional PCR, yielding a non-covalently closed donor plasmid, which was transformed directly into *E.coli* for gene disruption.

Southern hybridization

Disruptants were transformed with pACYC177 and grown in LB medium containing ampicillin to kickout the intron-donor plasmid (confirmed by acquisition of Amp^R and loss of Cam^R). The disruptants were then grown up in liquid culture, and chromosomal DNA was isolated by using a Qiagen Genomic-tip and Genomic DNA Buffer set. Southern hybridization was done as described in (10), using a ^{32}P-labeled probe for the retrotransposed intron generated by PCR of pACD3 with primers 5′-TCTTGCAAGGGTACGGAGTA and 5′-GTAGGGAGGTACCGCCTTGTTC. The probe was labeled using a High Prime DNA Labeling kit (Roche Diagnostics, Indianopolis, IN).

Plasmid-based mobility assays to determine the effect of target-site orientation on retrohoming frequency

Intron-donor plasmid pACD2X and recipient plasmids pBRR3A-ltrB, pBRR3B-ltrB or their derivatives containing different target sites were cotransformed into *E.coli* HMS174(DE3) (17). After induction with 100 μM IPTG for 1 h at 37°C, mobility events were detected by plating different dilutions of cells onto LB medium containing ampicillin or ampicillin plus tetracycline. The plates were incubated at 37°C for 24 h and mobility frequencies were calculated from the ratio of (Tet^R + Amp^R)/Amp^R colonies.

RESULTS AND DISCUSSION

Generation of an *E.coli* gene disruption library using randomly inserting targetrons

To explore the efficacy of targetron fishing, we generated an *E.coli* gene disruption library by using a previously constructed targetron in which the EBS2, EBS1 and δ sequences had been randomized (12). The targetron is a 0.9-kb Ll.LtrB-ΔORF intron, containing a Tp^R-RAM gene, and is cloned behind a T7lac promoter in the donor plasmid

pACD3 (Figure 1B). After transformation into *E.coli* HMS174(DE3), which contains an IPTG-inducible phage T7 RNA polymerase, cells were induced with IPTG and grown to saturation in Mueller–Hinton medium containing trimethoprim to select those containing integrated targetrons. This outgrowth step selects against targetrons that have inserted into essential genes or whose disruption results in a decreased growth rate, but in principle such targetrons could be obtained by isolating cells sooner after induction.
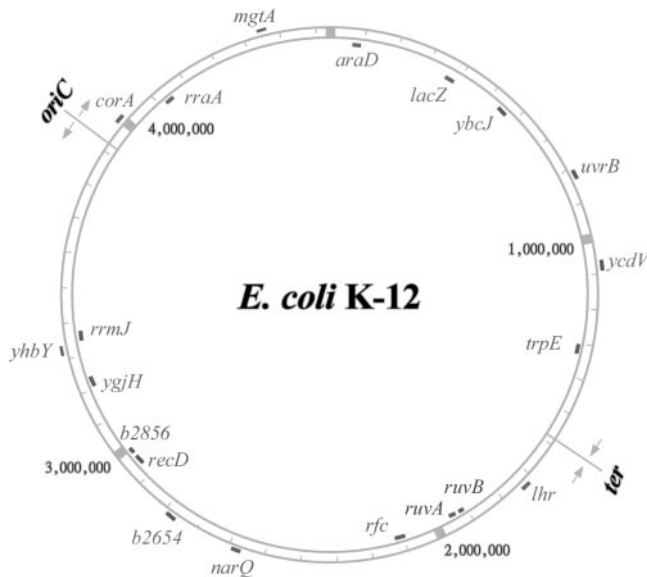

Figure 2. Map of the *E.coli* K-12 genome showing the location of the genes used to test targetron fishing. The map was compiled using the genome sequence of *E.coli* K12 MG1655 (GenBank accession number U00096). The figure shows the bidirectional DNA replication origin (*oriC*) and the terminator region (*ter*), with the arrows indicating the direction of replication away from the origin. Genes located on the outside or inside strands are indicated on the corresponding side of the circle.

Genomic DNA was isolated from the trimethoprim-enriched cell culture and used as a template for PCR to fish targetrons that inserted into specific genes.

Targetron fishing

We selected an initial set of 20 genes, including 11 genes more or less evenly distributed around the *E.coli* chromosome (*araD*, *corA*, *lacZ*, *lhr*, *mgtA*, *narQ*, *recD*, *rfc*, *trpE*, *uvrB* and *yhbY*), two DNA helicase genes (*ruvA* and *ruvB*), the 23S rRNA methyltransferase gene *rrmJ* and six genes encoding potential RNA-binding proteins of unknown function (*b2654*, *b2856*, *rraA*, *ybcJ*, *ycdV* and *ygjH*). In some cases, the genes were selected because of their relevance to other projects in the laboratory. Fifteen of these genes were known to be non-essential, while the remaining five (*b2654*, *b2856*, *rraA*, *ybcJ* and *yhbY*) had not been characterized as essential or non-essential (24) (http://shigen.lab.nig.ac.jp/ecoli/pec/index. jsp). The location of the genes on the *E.coli* chromosome is shown in Figure 2.

Targetrons were fished from the *E.coli* gene disruption library by PCR. For each gene, we carried out PCRs using a 'fixed' primer (P_i) complementary to an intron sequence just downstream of EBS1, together with a gene-specific primer in the 5′- or 3′-flanking region to amplify targetrons that had inserted into the sense or antisense strands, respectively (P_s and P_a, respectively; Figure 3). Both PCRs yield products that span the 5′-integration junction and include the targetron's EBS2, EBS1 and δ sequences, as well as the target gene's complementary IBS2 and IBS1 sequences. If more than one targetron has inserted into the gene, the PCR yields multiple products of different sizes. The PCR products were sequenced and cloned into the vector backbones of the donor plasmids pACD3 or pACD3-Tp-RAM (see Materials and Methods).

We successfully fished targetrons that inserted into 17 of the 20 genes tested (8 both strands, 4 sense strand only and 5 antisense strand only; in some cases, only one strand was attempted; see Table 1). Target sequences and base-pairing
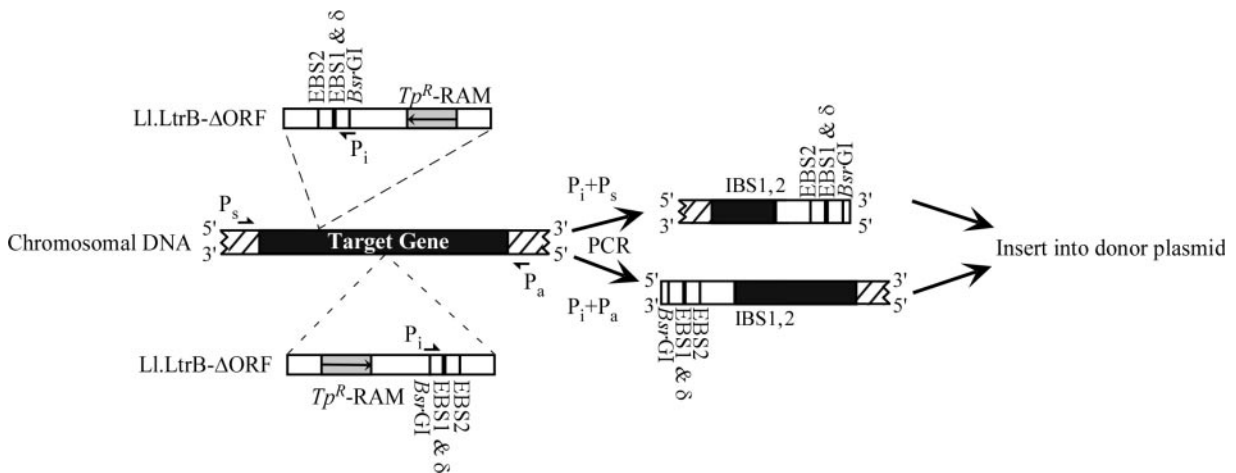

Figure 3. Scheme for targetron fishing. Chromosomal DNA from an *E.coli* gene disruption library containing randomly inserted Ll.LtrB introns is used as a template for PCR with a 'fixed' primer within the intron (P_i) and target gene-specific primers upstream (sense-strand insertions) or downstream (antisense-strand insertions) of the target gene (P_s and P_a, respectively). The resulting PCR products span the targetron integration junction and contain the 5′-exon sequences IBS1 and IBS2 in the target gene and the target-site recognition sequences EBS2, EBS1 and δ in the intron. The PCR product is then ligated or PCRed (see Figure 8) into a donor plasmid and transformed into an *E.coli* strain to obtain the desired single disruptant. Tp^R-RAM is the trimethoprim-resistance retrotransposition-activated marker gene inserted in the Ll.LtrB-ΔORF intron (see Figure 1C).
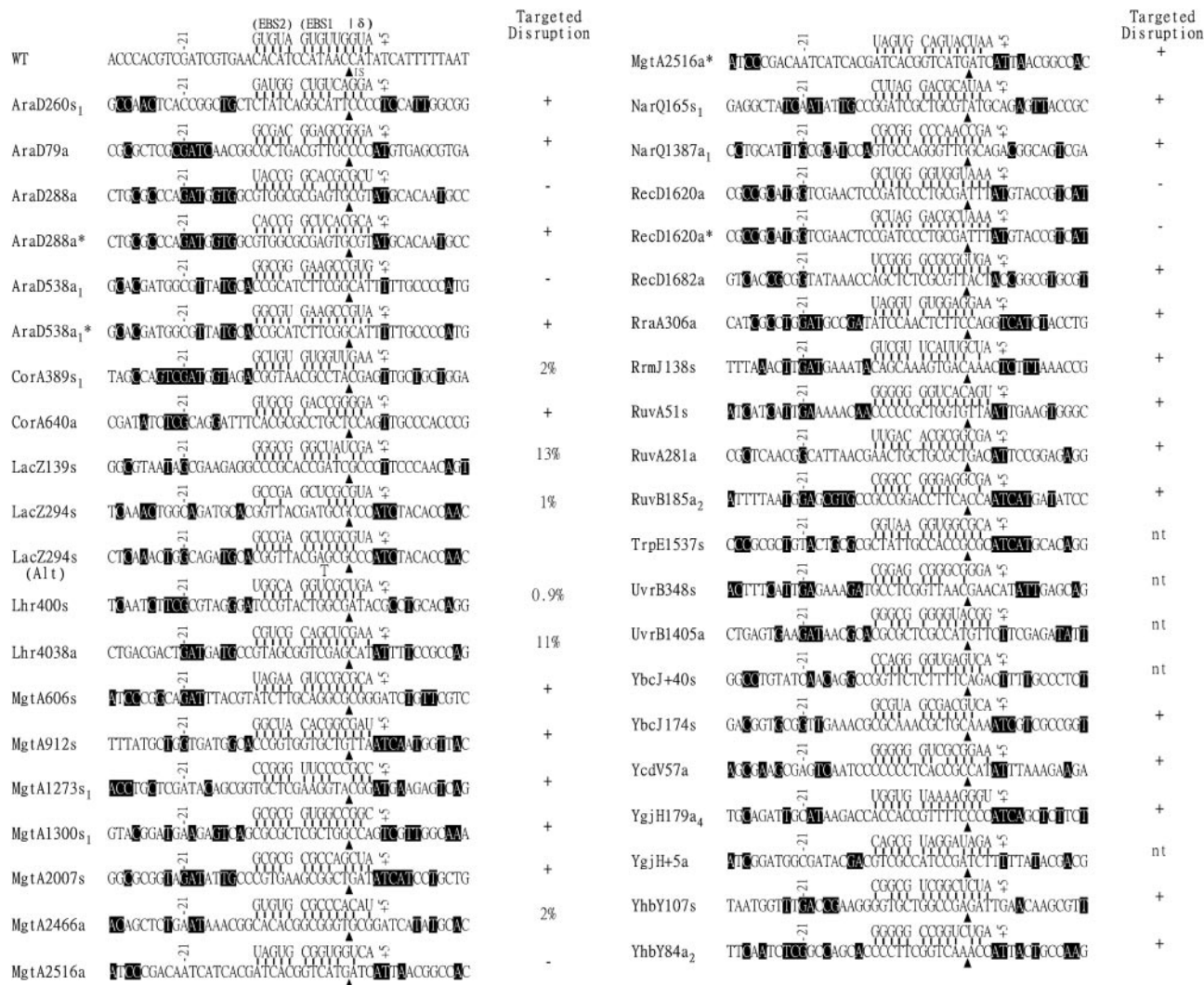
Figure 4. Insertion sites and base-pairing interactions of fished targetrons. The figure shows target site positions −30 to +15 and EBS/IBS and δ/δ′ base-pairing interactions for the wild-type Ll.LtrB intron (top left) and targetrons fished from the *E.coli* gene disruption library. Targetrons are named by nucleotide position 5′ to their insertion site in the target gene's coding sequence, followed by 's' or 'a' indicating sense or antisense strands, respectively. Nucleotide residues that match the wild-type Ll.LtrB target site are highlighted in black with white letters; arrowheads indicate the targetron-insertion site. Asterisks indicate targetrons that were modified to form Watson–Crick base pairs with the target sequence at previously mispaired positions and to have a favorable A residue at the δ+3 position [see (10)]. Different targetrons that inserted at the same site are distinguished by subscripts (see Figure 5 for complete list). LacZ294s (Alt) shows an alternate RNA–DNA base-pairing for this targetron in which T−4 in the DNA target site is flipped out. The ability of the fished targetron cloned in pACD3 or pACD3-Tp-RAM to give the desired single disruption is indicated under the column 'Targeted Disruption'. Numbers indicate the insertion frequency based on PCR screening of 24–108 colonies for targetrons cloned in pACD3, and + or − indicate successful or unsuccessful disruption, respectively, for targetrons cloned in pACD3-Tp-RAM; nt, not tested.

interactions for the fished targetrons are summarized in Figures 4 and 5. Multiple targetrons were isolated for *araD*, *lacZ*, *mgtA*, *recD*, *ruvB* and *ygjH*, while only one targetron was isolated for each of other genes. Two targetrons that inserted into 3′-untranslated regions were also isolated, one 40 nt downstream of *ybcJ* (YbcJ+40s) and the other 5 nt downstream of *ygjH* (YgjH+5a) (Figure 4). In a number of cases, two or more targetrons that differ slightly in their EBS1, EBS2 or δ sequences were found inserted at the same site (four in AraD260s, three in MgtA1273s, two each in AraD538a, CorA389s, MgtA1300s, NarQ165s, NarQ1387a, RuvB186a and YhbY84a; different targetrons that insert at the same site are distinguished by subscripts). The frequency of targetron insertion at different sites could be influenced by many factors, including the fitness of the target sequence, the activity

of the target gene, partial or complete occlusion of sites by DNA-binding proteins or the frequency of suitable targetrons in the donor-plasmid library. The distance of the target site from the DNA replication origin could also be a factor, although highly efficient targetrons can be obtained for sites distributed throughout the *E.coli* genome (12).

The three genes for which we were unable to fish targetrons were *b2654* and *b2856*, both of which are very small (333 and 165 bp, respectively) and of unknown essentiality, and *rfc* (aka *wbbH*, 1167 bp), which encodes O antigen polymerase (25). Neither the *rfc* gene nor other genes in the same operon are essential in *E.coli* K12 MG1655 (24), and we were able to disrupt *rfc* at low frequency in HMS174(DE3) by using a computationally designed targetron (Rfc706a) with a Tp^R-RAM gene (data not shown). Thus, the failure to recover
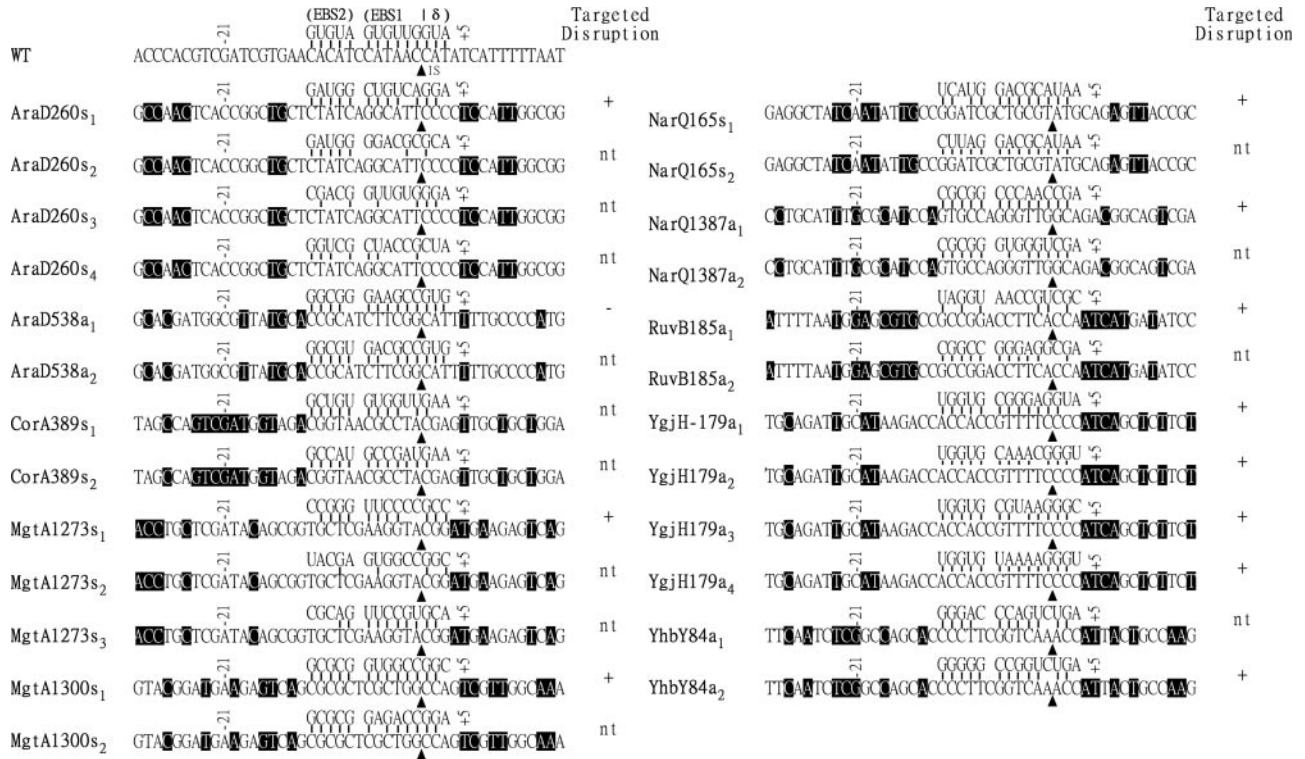
Figure 5. Cases in which two or more different targetrons inserted at the same target site. The different targetrons are distinguished by their unique EBS and δ sequences and are denoted by subscripts. Target sequences, base-pairing interactions and the ability of the fished targetron to give the desired single disruptant when retransformed into *E.coli* HMS174(DE3) are indicated as in Figure 4.
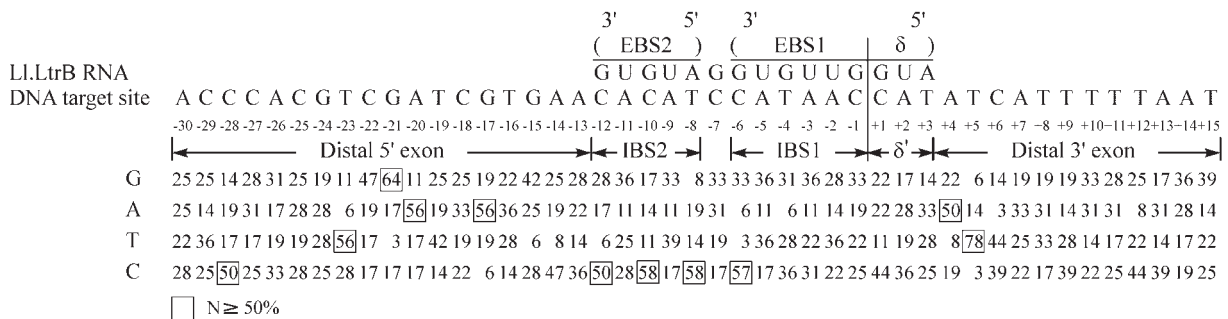
Figure 6. Nucleotide frequencies at insertion sites of fished targetrons. The target sequence and base-pairing interactions for the wild-type Ll.LtrB intron are shown at the top. Numbers indicate nucleotide frequencies (%) at positions −30 to +15 for 36 fished-targetron insertion sites in the *E.coli* genome. Nucleotide residues conserved in ≥50% of the inserted targetrons are boxed.

Rfc targetrons from the library could reflect either that the disruptants were underrepresented initially or that they grow slowly and were lost in the outgrowth step.

Characteristics of fished targetrons and their insertion sites

Figures 6 and 7 show compilations of nucleotide frequencies and base-pairing interactions at the 36 insertion sites of targetrons fished from the library. These data are generally in good agreement with those obtained previously for randomly integrated targetrons isolated from an *E.coli* gene disruption library by inverse PCR (12). Most of the isolated targetrons have good matches for the critical positions recognized by the IEP, including G−21 (64%) and T+5 (78%), as well as good but not perfect EBS/IBS and δ/δ′ pairings (see also Figures 4 and 5). The compilations show that the requirement for base-pairing is most stringent for positions −12 to +1 (81–100% base-pairing), and less stringent at position +2 (72%), while position +3 shows a previously noted strong preference for an A-residue in the intron RNA (73%) regardless of its potential base-pairing partner in the DNA target site (10,12). The apparent preference for the wild-type C−28 (50%) was not seen previously (12) and likely reflects a statistical anomaly in the smaller data set in the present work.

Notably, several target sites deviate substantially from the canonical sequence recognized by the IEP, including some that were sites of multiple insertions by different targetrons

RNA/DNA	3′ (—— EBS2 ——) 5′					(—— EBS1 ——)							δ		5′
Ll.LtrB RNA	G	U	G	U	A	G	G	U	G	U	U	G	G	U	A
DNA target site	C	A	C	A	T	C	C	A	T	A	A	C	C	A	T
position	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3
	IBS2						IBS1						δ′		
A / A	0	0	0	0	6	0	0	0	3	0	0	0	0	8	25
A / C	0	0	0	0	0	0	0	3	0	0	3	0	0	3	22
A / G	0	0	0	0	0	0	0	0	0	0	3	0	0	3	8
A / T	0	17	3	17	[11]	3	3	17	11	11	17	11	8	8	[19]
C / A	0	0	0	0	0	0	0	3	0	0	0	3	6	3	3
C / C	0	0	0	0	0	0	0	0	3	0	0	0	3	0	0
C / G	22	28	14	31	8	25	25	22	22	33	22	31	19	14	3
C / T	3	0	3	0	0	0	0	0	3	0	6	0	0	3	0
G / A	0	0	0	0	6	0	0	0	0	3	0	3	0	6	0
G / C	[50]	28	[58]	17	58	56	[56]	14	33	31	19	[22]	[42]	31	0
G / G	3	6	0	0	0	3	3	6	0	0	3	3	3	0	0
G / T	0	8	6	22	3	6	0	17	[8]	11	11	11	3	8	6
U / A	17	[11]	14	[11]	8	8	8	[8]	3	[8]	[11]	11	19	[11]	6
U / C	0	0	0	0	0	0	0	0	0	0	0	3	0	3	3
U / G	3	3	3	3	0	3	6	8	8	0	3	0	0	0	3
U / T	3	0	0	0	0	0	0	3	6	0	0	0	0	0	3
Base paired (%)	92	94	97	100	89	97	97	86	86	97	81	86	83	72	31

Figure 7. Base-pairing interactions at insertion sites of fished targetrons. Base-pairing interactions between the wild-type Ll.LtrB intron and its *ltrB* DNA target site are shown at the top. Numbers indicate the frequency (%) of each RNA/DNA base pair at each position for 36 fished targetrons and their insertion sites in the *E.coli* genome. The wild-type base pair at each position is boxed, and the percentage of Watson–Crick plus G/T or U/G wobble base pairs at each position is tabulated at the bottom.

(Figure 5). The latter include the target sites for AraD260s, AraD538a and MgtA1273s, all of which lack T–23, G–21, A–20 and G–17, and those for NarQ165s and NarQ1387a, both of which lack T+5, a critical nucleotide residue for second-strand cleavage. LacZ294s also appears to have a highly deviant target site with an unfavorable G at position −23 and C at position −21, and six mispairings between positions −12 and +1 (Figure 4). However, we noticed that target site could conform to that for the wild-type Ll.LtrB intron if a single T residue at position −4 was flipped-out to change the register by one nucleotide residue [see alternate pairing for LacZ294s (Alt) in Figure 4], and we confirmed that a target site in which T–4 was deleted supported efficient retrohoming in a plasmid assay (data not shown). We show below that all of the targetrons that recognize deviant target sites still gave the expected disruption when tested individually.

Use of fished targetrons to obtain single disruptants

Thirty-one of the fished targetrons were recloned into the pACD3 and/or pACD3-Tp-RAM donor plasmid and transformed into *E.coli* HMS174(DE3) to see if they would give the desired single disruptant. We note that the target gene's 5′-flanking sequences, which are cloned with the targetron into the donor plasmid (see Figure 3), automatically contain IBS1 and IBS2 sequences complementary to the targetron's unique EBS1 and EBS2 sequences, since these base-pairing interactions were necessary for targetron integration by reverse-splicing (Figures 4, 5 and 7). In the donor construct, base-pairing between the targetron's EBS sequences and the complementary IBS sequences in the 5′ exon is necessary for RNA splicing to generate active RNPs (26).

Of the 31 targetrons tested, 27 gave the desired single disruptant without further modification (results summarized under 'Targeted Disruption in Figures 4 and 5). In each case, insertion of the targetron at the correct site was confirmed by colony PCR and DNA sequencing (data not shown). The four targetrons that did not give the desired single disruptants, AraD288a, AraD538a$_1$, MgtA2516a and RecD1620a, each contains mispairings in the EBS/IBS interactions. When retested after modifying their EBS sequences to form Watson–Crick base pairs at the mispaired positions and to insert a preferred A residue at the δ+3 position (targetrons indicated by asterisks in Figure 4), three gave the desired single disruptant. The RecD1620a targetron, which was confirmed to have the correct sequence, did not give the desired single disruption. It appears to have a favorable target site for IEP recognition, but its EBS2 sequence, 5′-GGCGC, could mispair with a neighboring sequence just upstream in the intron RNA to disrupt the Id1 stem–loop, which displays EBS2. Such mispairings have been found to result in low targeting efficiencies (10).

In general, more efficient targetrons had better matches to the wild-type target site and higher numbers of EBS/IBS and δ/δ' base pairs than did the less efficient targetrons. However, even targetrons with deviant target sites or significant numbers of mispairings gave specific disruptions. The LacZ294s targetron, mentioned above, gave disruptions at a frequency of 1% by blue-white screening. YhbY84a$_2$ and MgtA1273s$_1$, which have two or three mispairs at key positions, respectively, had targeting efficiencies sufficient to detect using the Tp^R-RAM, as did all four YgjH179a targetrons, despite mispairing at different positions in three of the targetrons (Figures 4 and 5). Finally, seven targetrons that insert at sites lacking T+5 were also tested, and all gave the desired specific disruptions (Lhr400s; MgtA606s, MgtA2466a, NarQ165s$_1$, NarQ1387a$_1$, RecD1682a, and YhbY107s; Figure 4).

A PCR method for inserting fished targetrons into donor plasmids without cloning

For high-throughput approaches, it is desirable to generate donor plasmids containing the fished targetrons without cloning. For this purpose, we developed the procedure diagrammed in Figure 8 in which a 27 nt vector sequence (v) was added to the 5′ end of the target gene-specific primers used for targetron fishing. The resulting PCR product was then used in a second PCR with *Eco*RV-digested donor plasmid pACD3E-Tp-RAM-PCR (12), yielding a non-covalently closed circular PCR product, which can be transformed directly into *E.coli*. This method avoids the use of a gene-specific restriction site flanking the target gene for blunt-end cloning into the vector. A test of this method in which the AraD260s$_1$ targetron was amplified directly into the pACD3E-Tp-RAM plasmid backbone gave the correct disruption at roughly the same frequency as the conventional cloning method (data not shown).

Specificity of targetron insertion

Because some of the 'fished' targetrons have suboptimal protein recognition and base-pairing interactions, there was a danger that they would be less specific than designed targetrons. To test the specificity of targetron insertion, we carried out Southern hybridizations for a randomly chosen disruptant obtained with each of the fished targetrons that was active in
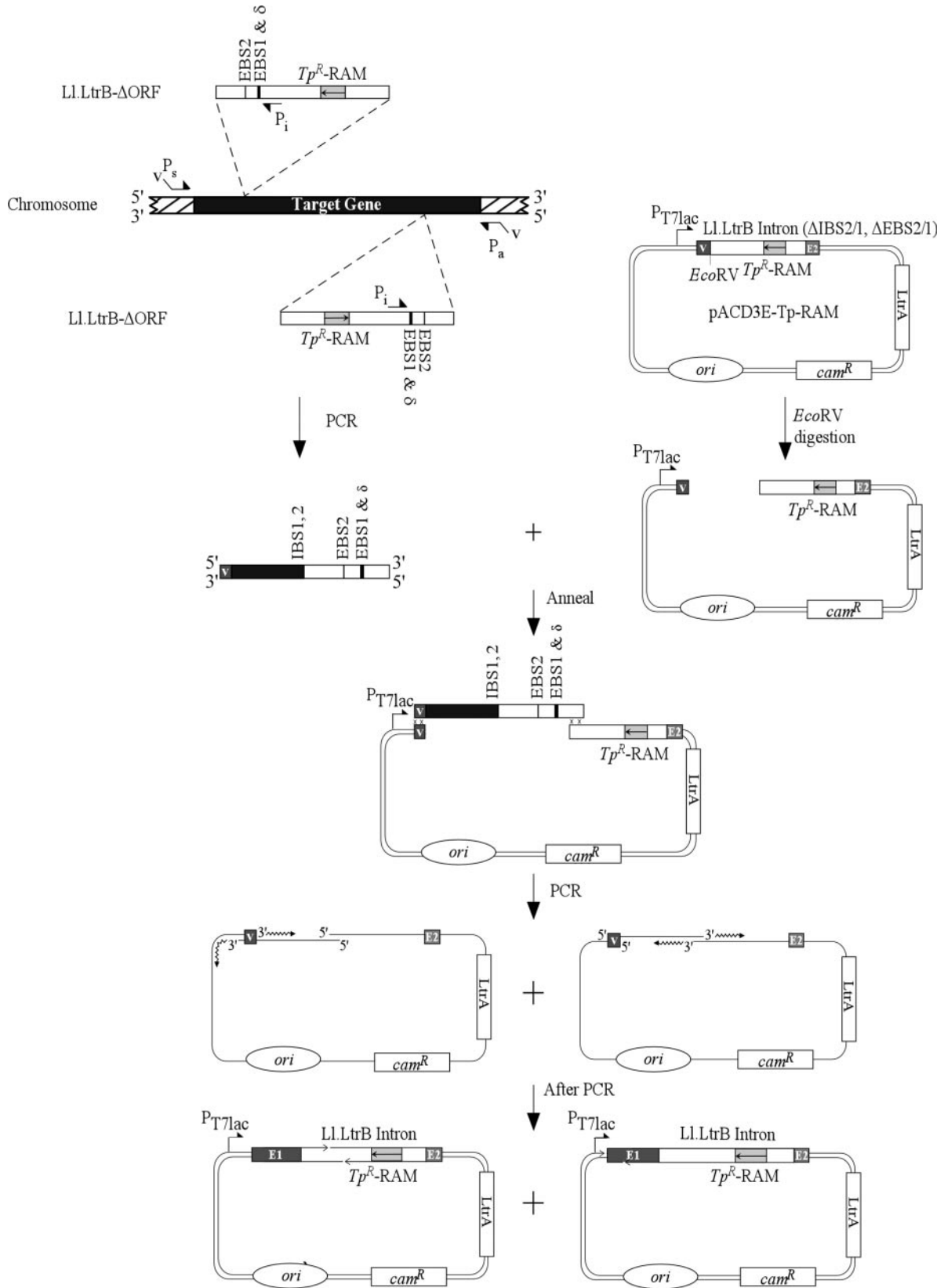
Figure 8. PCR method for inserting fished targetrons directly into the donor plasmid without cloning. A 361 bp linear DNA fragment (top left) spanning the 5′ integration junction of the targetron was generated by PCR using primers P_i and P_s or P_a, the latter two with an additional 5′ sequence complementary to the vector (v). The PCR product was then used as a primer for a second PCR with the 6.7 kb *Eco*RV-digested pACD3-Tp-RAM-PCR plasmid as the template. The final circular PCR product contains gaps (arrowheads) at different positions, depending on the strand of the 361 bp linear DNA from which priming occurred. The PCR product can be transformed directly into *E.coli* or after ligation to seal nicks.
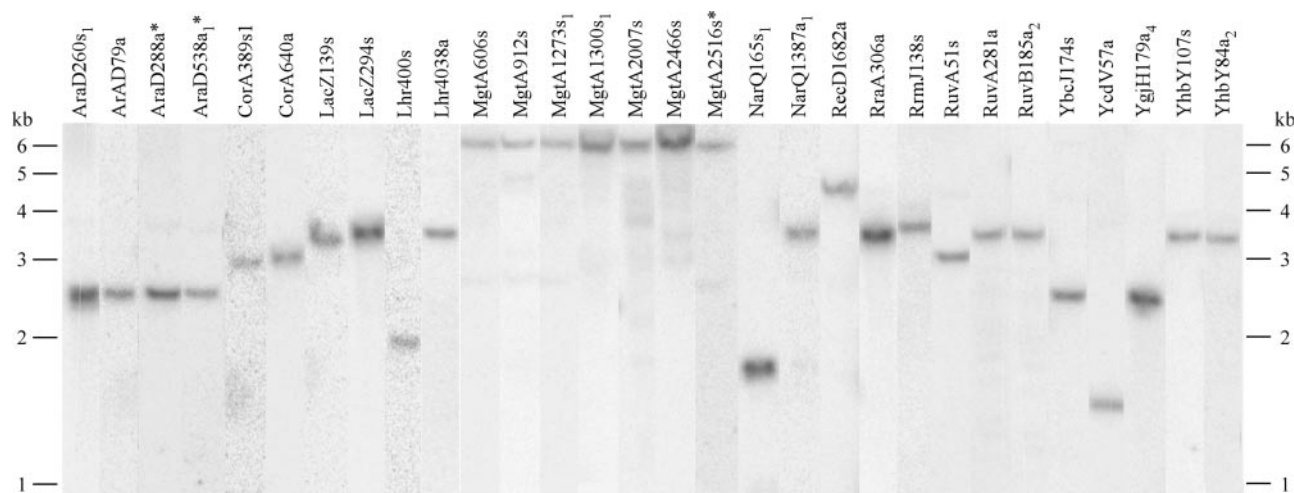
Figure 9. Southern hybridizations showing the integration specificity of fished targetrons. Disruptants were obtained using the indicated fished targetrons in *E.coli* HMS174(DE3), and their DNA was isolated, digested with restriction enzymes, blotted to a nylon membrane and hybridized with a ^{32}P-labeled probe specific for the Ll.LtrB intron (see Materials and Methods). Restriction enzymes were *Bgl*I (AraD260s$_1$, AraD79a, AraD288a*, AraD538a$_1$*, LacZ139s, LacZ294s, Lhr400s, Lhr4038a, RecD1682a, RraA306a, RrmJ138a, YbcJ174s, YcdV57a, YhbY107s, YhbY84a$_2$); *Bpm*I (CorA389s$_1$, CorA640a, YgjH179a$_4$); and *Ssp*I (MgtA606s, MgtA912s, MgtA1273s$_1$, MgtA1300s$_1$, MgtA2007s, MgtA2466a, MgtA2516a*, NarQ165s$_1$, NarQ1387a$_1$, RuvA51s, RuvA281a, RuvB185a$_2$). The blot was dried and scanned with a PhosphorImager. Numbers at the left and right indicate positions of molecular weight markers (1 kb plus ladder; Invitrogen, Carlsbad, CA).

giving disruptants. In 22 cases, the Southern hybridizations showed a single prominent band of the size expected for site-specific insertion of the targetron into the target gene, while in the remaining eight cases (AraD288a*, AraD538a$_1$*, Lhr4038a, MgtA606s, MgtA1300s$_1$, MgtA2516a*, RuvA51s and YhbY84a$_2$), the initial disruptant showed multiple bands. In these eight cases, we isolated 5–10 additional transformants using a shorter induction period (2 h instead of 18 h; see Materials and Methods), and in each case, we obtained at least one transformant that gave a single integration at the desired site. Figure 9 shows Southern blots for the complete set of single disruptions for each targetron. We note that all seven of the tested targetrons that insert at sites lacking T+5 gave the desired single disruptant, with only one (MgtA606s) being in the group that initially gave multiple bands.

Analysis of retrohoming into target sites that lack T+5

Those targetrons that insert at sites lacking T+5 could do so either by an En-dependent pathway with inefficient second-strand cleavage or by En-independent pathways using nascent leading or lagging DNA strands to prime reverse-transcription (17,18,22,27). To help distinguish these possibilities, we tested the ability of each targetron to retrohome into its target site cloned in recipient plasmids in opposite orientations relative to the direction of DNA replication. The assay used is diagrammed in Figure 10A. In this assay, the Amp^R recipient plasmids pBRR3A and pBRR3B contain the target site cloned immediately upstream of a promoterless *tet^R* in orientations denoted LAG or LEAD, according to whether a nascent lagging or leading DNA strand could be used to prime reverse-transcription of the inserted intron RNA (17). The donor intron contains a phage T7 promoter inserted near its 3′ end, so that its insertion into the target site activates the expression of the *tet^R* gene. Mobility frequencies are then determined from the ratio of (Tet^R + Amp^R)/Amp^R colonies.

In previous work using the above assay, we found that retrohoming of the Ll.LtrB intron under conditions where second-strand cleavage is inhibited by mutations in either the IEP or DNA target site shows a pronounced bias for the LEAD orientation (17). This bias is thought to reflect that when the intron reverse-splices into double-stranded DNA prior to passage of a replication fork, it is positioned to directly use a nascent leading-strand primer, while use of a lagging-strand primer requires the potentially disruptive passage of the replication fork through the inserted intron. In contrast, some En$^-$ group II introns are thought to insert preferentially into single-stranded DNA at replication forks, enabling the efficient use of lagging-strand primers, and such introns show the opposite orientation bias in this type of assay (18).

The results of the assays for the wild-type Ll.LtrB intron and those targetrons that insert into sites lacking T+5 are summarized in Figure 10B. As expected, the wild-type Ll.LtrB intron, which uses an En-dependent retrohoming pathway, showed no significant orientation bias. Five of the seven target sites lacking T+5 are located on the leading template strand of the *E.coli* chromosome. In the plasmid assay, all of these sites were used at either low or undetectable frequency. Three of these sites (Lhr400s, MgtA2466a and RecD1682a) showed a clear bias for insertion into the leading template strand, the pattern expected for the reverse-splicing into double-stranded DNA and use of a nascent leading DNA strand to prime reverse-transcription, while the other two sites (NarQ165s and YhbY107s) showed no significant strand bias, the pattern expected for an En-dependent pathway with inefficient second-strand cleavage.

The remaining two target sites lacking T+5 are located on the lagging template strand of the *E.coli* chromosome. In the plasmid assay, one of these sites (MgtA606s) showed a very pronounced bias for the lagging template strand, with a relatively high insertion frequency (7.4%) in that orientation, and no detectable insertion in the opposite orientation (<0.0001%).
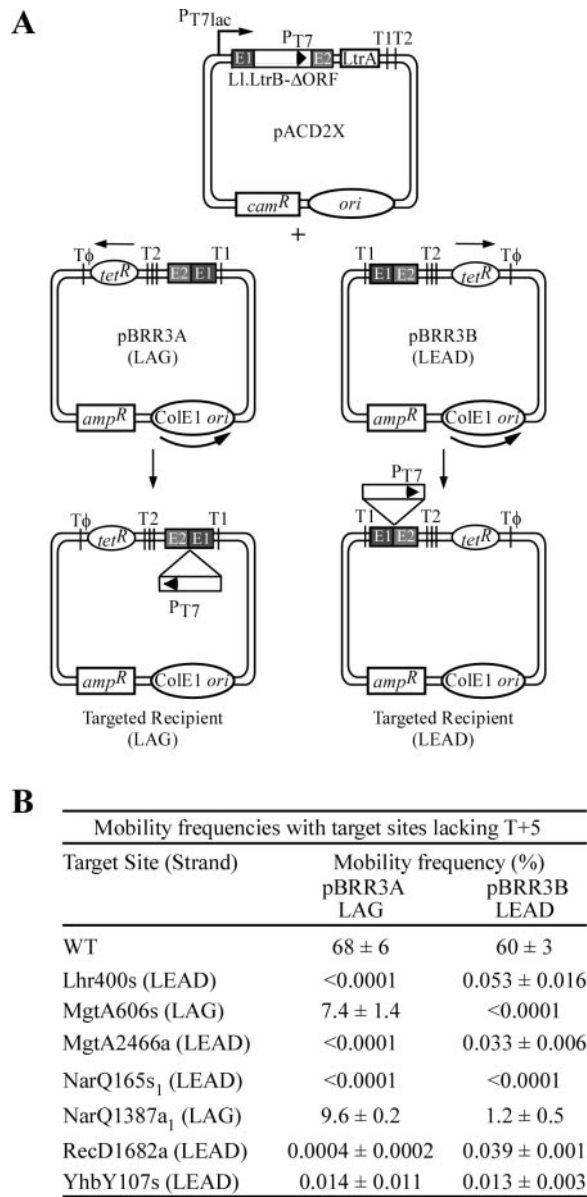
Figure 10. Mobility assays with target sites lacking T+5 cloned in opposite orientations relative to the direction of plasmid replication. (**A**) Mobility assay. The Cap^R donor plasmid pACD2X expresses a 940 nt Ll.LtrB-ΔORF intron with a phage T7 promoter inserted near its 3′ end. The Amp^R recipient plasmids pBRR3A and pBRR3B contain a target site/*tet^R* cassette cloned in opposite orientations denoted LAG or LEAD, depending on whether a nascent lagging or leading DNA strand could be used to prime reverse-transcription of the reverse-spliced intron RNA. Insertion of the intron into the target site activates the expression of the *tet^R* gene, and mobility frequencies are measured as the ratio of (Tet^R + Amp^R)/Amp^R colonies. T1 and T2 are *E.coli rrnB* transcription terminators, which terminate read-through transcription by *E.coli* RNA polymerase, but not phage T7 RNA polymerase. Tφ is a phage T7 transcription terminator. (**B**) Mobility frequencies. The target site is indicated to the left with (LEAD) or (LAG) indicating whether the site is located on the leading- or lagging-template strand of the *E.coli* chromosome. Mobility frequencies (%) are the mean ± SD for at least three replicate experiments for each targetron/target site combination.

This bias is expected for reverse-splicing into single-stranded DNA and use a nascent lagging strand to prime reverse-transcription. The inability of MgtA606s to reverse-splice into double-stranded DNA may reflect suboptimal IEP

recognition in the 5′ exon and/or that it has relatively weak base pairs at EBS2/IBS2 positions −12 and −11 (U–A and A–T), which are thought to be critical for nucleating local DNA unwinding (4,10).

The other site found on the lagging template strand of the *E.coli* chromosome, NarQ1387a, showed a substantial bias for the LAG orientation (9.6%), but also supports relatively efficient retrohoming in the LEAD orientation (1.2%). This pattern suggests preferential targeting of single-stranded DNA, but substantial reverse-splicing into double-stranded DNA. The ability of the NarQ1387a$_1$ targetron to efficiently reverse-splice into double-stranded DNA may reflect that its EBS2/IBS2 pairing consists largely of stable GC or CG base pairs (Figure 4), thereby favoring displacement of the opposite strand for reverse-splicing into double-stranded DNA.

SUMMARY AND PROSPECTS

The targetron fishing approach selects targetrons that have successfully inserted into desired target genes and uses them as reagents to obtain specific single disruptions. Despite the previously observed preference for targetron insertion near the chromosome replication origin under similar conditions (12), the library obtained here is sufficiently complex to contain most viable *E.coli* gene disruptants. The fished targetrons can in principle be used to obtain disruptions in any bacterial strain, so long as the target sequence does not deviate at significant positions. The remarkable finding is that a high proportion of the fished targetrons gave the desired single disruptant without further modification, even if there were mispairings in the EBS/IBS and δ–δ′ interactions. By selecting only targetrons that have inserted successfully, the fishing approach avoids factors that could negatively impact intron-insertion frequency and have not yet been incorporated into computer algorithms for target site selection and intron design. These include the effect of higher-order DNA structure and thermal stability profiles of the target sites, protein-binding at DNA target sites, and the deleterious effects of different combinations of nucleotide residues on intron RNA structure. Targetron fishing also has the advantage of requiring only two target gene-specific primers and a simple single-step PCR, compared to a more complex PCR for retargeting computationally designed group II introns (10).

The finding that some targetrons retain specificity even in the absence of perfect EBS/IBS and δ–δ′ interactions implies a significant degree of redundancy in the RNA and protein interactions used for target site selection. Additionally, we were surprised by the number of the fished targetrons that gave specific disruptions despite substantial deviations from current target site-selection rules (10). One of these, LacZ294s, appears to recognize its target site by flipping out one target site nucleotide residue in the base-paired region, and seven targetrons insert specifically at sites lacking T+5, which is critical for second-strand cleavage. Two of these targetrons appear to insert at relatively high efficiency by targeting single-stranded DNA and using a nascent lagging strand to prime reverse-transcription, a mechanism used by En$^-$ group II introns (18,27). Thus, in addition to providing a rapid method for obtaining validated targetrons, our results show that Ll.LtrB can be targeted to a much greater range of sites

than predicted by the current computer algorithm (10), including sites used preferentially in single-stranded DNA and without second-strand cleavage. Finally, our findings indicate that En$^+$-group II introns can switch surprisingly readily to En-independent retrohoming pathways, thereby facilitating their dispersal to new target sites. Such a switch may in turn lead to dispensability of the En domain, possibly contributing to its loss multiple times during evolution in different group II intron lineages (28).

ACKNOWLEDGEMENTS

Conflict of interest statement. Group II intron gene targeting technology is subject to patents licensed by The Ohio State University and the University of Texas at Austin to InGex, LLC, and rights to sell research tools based on the technology are sublicensed to Sigma-Aldrich. A.M.L. is a minority equity holder in InGex, LLC, and all of the authors are potential recipients of royalties paid to Ohio State and the University of Texas at Austin.

REFERENCES

1. Belfort,M., Derbyshire,V., Parker,M.M., Cousineau,B. and Lambowitz,A.M. (2002) Mobile introns: pathways and proteins. In Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), *Mobile DNA II*. ASM Press, Washington DC, pp. 761–783.
2. Lambowitz,A.M. and Zimmerly,S. (2004) Mobile group II introns. *Annu. Rev. Genet.*, **38**, 1–35.
3. Guo,H., Zimmerly,S., Perlman,P.S. and Lambowitz,A.M. (1997) Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. *EMBO J.*, **16**, 6835–6848.
4. Singh,N.N. and Lambowitz,A.M. (2001) Interaction of a group II intron ribonucleoprotein endonuclease with its DNA target site investigated by DNA footprinting and modification interference. *J. Mol. Biol.*, **309**, 361–386.
5. Guo,H., Karberg,M., Long,M., Jones,J.P.,III, Sullenger,B. and Lambowitz,A.M. (2000) Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science*, **289**, 452–457.
6. Mohr,G., Smith,D., Belfort,M. and Lambowitz,A.M. (2000) Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. *Genes Dev.*, **14**, 559–573.
7. Lambowitz,A.M., Mohr,G. and Zimmerly,S. (2005) Group II intron homing endonucleases: ribonucleoprotein complexes with programmable target specificity. In Belfort,M., Derbyshire,V., Stoddard,B. and Wood,D. (eds), *Homing Endonucleases and Inteins*. Springer-Verlag, Heidelberg, in press.
8. Frazier,C.L., San Filippo,J., Lambowitz,A.M. and Mills,D.A. (2003) Genetic manipulation of *Lactococcus lactis* by using targeted group II introns: generation of stable insertions without selection. *Appl. Environ. Microbiol.*, **69**, 1121–1128.
9. Karberg,M., Guo,H., Zhong,J., Coon,R., Perutka,J. and Lambowitz,A.M. (2001) Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nat. Biotechnol.*, **19**, 1162–1167.
10. Perutka,J., Wang,W., Goerlitz,D. and Lambowitz,A.M. (2004) Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. *J. Mol. Biol.*, **336**, 421–439.
11. Runyen-Janecky,L.J., Reeves,S.A., Gonzales,E.G. and Payne,S.M. (2003) Contribution of the *Shigella flexneri* Sit, Iuc, and Feo iron acquisition systems to iron acquisition *in vitro* and in cultured cells. *Infect. Immun.*, **71**, 1919–1928.
12. Zhong,J., Karberg,M. and Lambowitz,A.M. (2003) Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucleic Acids Res.*, **31**, 1656–1664.
13. Matsuura,M., Noah,J.W. and Lambowitz,A.M. (2001) Mechanism of maturase-promoted group II intron splicing. *EMBO J.*, **20**, 7259–7270.
14. Noah,J.W. and Lambowitz,A.M. (2003) Effects of maturase binding and Mg^{2+} concentration on group II intron RNA folding investigated by UV cross-linking. *Biochemistry*, **42**, 12466–12480.
15. Saldanha,R., Chen,B., Wank,H., Matsuura,M., Edwards,J. and Lambowitz,A.M. (1999) RNA and protein catalysis in group II intron splicing and mobility reactions using purified components. *Biochemistry*, **38**, 9069–9083.
16. Aizawa,Y., Xiang,Q., Lambowitz,A.M. and Pyle,A.M. (2003) The pathway for DNA recognition and RNA integration by a group II intron retrotransposon. *Mol. Cell*, **11**, 795–805.
17. Zhong,J. and Lambowitz,A.M. (2003) Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription. *EMBO J.*, **22**, 4555–4565.
18. Martínez-Abarca,F., Barrientos-Durán,A., Fernández-López,M. and Toro,N. (2004) 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.*, **32**, 2880–2888.
19. Cousineau,B., Smith,D., Lawrence-Cavanagh,S., Mueller,J.E., Yang,J., Mills,D., Manias,D., Dunny,G., Lambowitz,A.M. and Belfort,M. (1998) Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell*, **94**, 451–462.
20. Martínez-Abarca,F., García-Rodríguez,F.M. and Toro,N. (2000) Homing of a bacterial group II intron with an intron-encoded protein lacking a recognizable endonuclease domain. *Mol. Microbiol.*, **35**, 1405–1412.
21. Mills,D.A., Manias,D.A., McKay,L.L. and Dunny,G.M. (1997) Homing of a group II intron from *Lactococcus lactis* subsp. *lactis* ML3. *J. Bacteriol.*, **179**, 6107–6111.
22. Ichiyanagi,K., Beauregard,A., Lawrence,S., Smith,D., Cousineau,B. and Belfort,M. (2002) Retrotransposition of the Ll.LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. *Mol. Microbiol.*, **46**, 1259–1272.
23. San Filippo,J. and Lambowitz,A.M. (2002) Characterization of the C-terminal DNA-binding/DNA endonuclease region of a group II intron-encoded protein. *J. Mol. Biol.*, **324**, 933–951.
24. Hashimoto,M., Ichimura,T., Mizoguchi,H., Tanaka,K., Fujimitsu,K., Keyamura,K., Ote,T., Yamakawa,T., Yamazaki,Y., Mori,H. *et al.* (2005) Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol. Microbiol.*, **55**, 137–149.
25. Lukomski,S., Hull,R.A. and Hull,S.I. (1996) Identification of the O antigen polymerase (*rfc*) gene in *Escherichia coli* O4 by insertional mutagenesis using a nonpolar chloramphenicol resistance cassette. *J. Bacteriol.*, **178**, 240–247.
26. Michel,F. and Ferat,J.L. (1995) Structure and activities of group II introns. *Annu. Rev. Biochem.*, **64**, 435–461.
27. Coros,C.J., Landthaler,M., Piazza,C.L., Beauregard,A., Esposito,D., Perutka,J., Lambowitz,A.M. and Belfort,M. (2005) Retrotransposition strategies of the *Lactococcus lactis* Ll.LtrB group II intron are dictated by host identity and cellular environment. *Mol. Microbiol.*, **56**, 509–524.
28. Zimmerly,S., Hausner,G. and Wu,X. (2001) Phylogenetic relationships among group II intron ORFs. *Nucleic Acids Res.*, **29**, 1238–1250.