

REVIEW

Open Access

# Biotechnological applications of mobile group II introns and their reverse transcriptases: gene targeting, RNA-seq, and non-coding RNA analysis

Peter J Enyeart, Georg Mohr, Andrew D Ellington and Alan M Lambowitz\*

## Abstract

Mobile group II introns are bacterial retrotransposons that combine the activities of an autocatalytic intron RNA (a ribozyme) and an intron-encoded reverse transcriptase to insert site-specifically into DNA. They recognize DNA target sites largely by base pairing of sequences within the intron RNA and achieve high DNA target specificity by using the ribozyme active site to couple correct base pairing to RNA-catalyzed intron integration. Algorithms have been developed to program the DNA target site specificity of several mobile group II introns, allowing them to be made into 'targetrons.' Targetrons function for gene targeting in a wide variety of bacteria and typically integrate at efficiencies high enough to be screened easily by colony PCR, without the need for selectable markers. Targetrons have found wide application in microbiological research, enabling gene targeting and genetic engineering of bacteria that had been intractable to other methods. Recently, a thermostable targetron has been developed for use in bacterial thermophiles, and new methods have been developed for using targetrons to position recombinase recognition sites, enabling large-scale genome-editing operations, such as deletions, inversions, insertions, and 'cut-and-pastes' (that is, translocation of large DNA segments), in a wide range of bacteria at high efficiency. Using targetrons in eukaryotes presents challenges due to the difficulties of nuclear localization and sub-optimal magnesium concentrations, although supplementation with magnesium can increase integration efficiency, and directed evolution is being employed to overcome these barriers. Finally, spurred by new methods for expressing group II intron reverse transcriptases that yield large amounts of highly active protein, thermostable group II intron reverse transcriptases from bacterial thermophiles are being used as research tools for a variety of applications, including qRT-PCR and next-generation RNA sequencing (RNA-seq). The high processivity and fidelity of group II intron reverse transcriptases along with their novel template-switching activity, which can directly link RNA-seq adaptor sequences to cDNAs during reverse transcription, open new approaches for RNA-seq and the identification and profiling of non-coding RNAs, with potentially wide applications in research and biotechnology.

**Keywords:** Genome engineering, Metabolic engineering, Next-generation RNA sequencing, Ribozyme, Synthetic biology, Systems biology, Targetron

## Review

### Introduction

Mobile group II introns are bacterial retrotransposons that perform a remarkable ribozyme-based, site-specific DNA integration reaction ('retrohoming') and encode an equally remarkable reverse transcriptase (RT), both of which have been harnessed for biotechnological applications [1-3].

Retrohoming occurs by a mechanism in which the group II intron RNA uses its ribozyme activity to insert directly into a DNA strand, where it is reverse transcribed by the intron-encoded RT (also referred to as the intron-encoded protein or IEP), yielding a cDNA copy of the intron that is integrated into the genome [4]. Because mobile group II introns recognize DNA target sequences largely by base pairing of sequence motifs within the intron RNA, they can be programmed to insert into desired DNA sites by simply modifying the intron sequences so as to base pair to the new target site. This feature allows mobile group II

\* Correspondence: [lambowitz@austin.utexas.edu](mailto:lambowitz@austin.utexas.edu)  
Departments of Molecular Biosciences and Chemistry, Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX 78712, USA

introns to be made into gene targeting vectors, or 'targetrons', which combine high DNA integration efficiency with readily programmable and reliable DNA target specificity [5-7]. Targetrons are widely used for genetic engineering of bacteria, and efforts continue to adapt them for function in eukaryotes.

Group II intron RTs function in retrohoming by synthesizing a full-length cDNA of the highly structured intron RNA with high processivity and fidelity [8-10], properties that are useful for biotechnological applications involving cDNA synthesis, such as qRT-PCR and next-generation RNA sequencing (RNA-seq). The RTs also have a novel template-switching activity that enables facile attachment of adaptor sequences containing primer-binding sites and barcodes to cDNAs. These properties, combined with the availability of naturally occurring thermostable group II intron RTs [11,12] open new approaches for RNA-seq and the profiling and discovery of miRNAs and other non-coding RNAs [10,13].

Here, we describe how the novel biochemical activities of mobile group II introns and their RTs, which were acquired during the evolution of group II introns as mobile genetic elements, have been adapted for biotechnological applications. We then review how group II intron-derived targetrons have been used for the genetic engineering of diverse bacteria, as well as recent advances in targetron technology. The latter include the development of a thermotargetron for gene targeting in thermophiles, methods for using targetrons to position recombinase recognition sites for large-scale genome

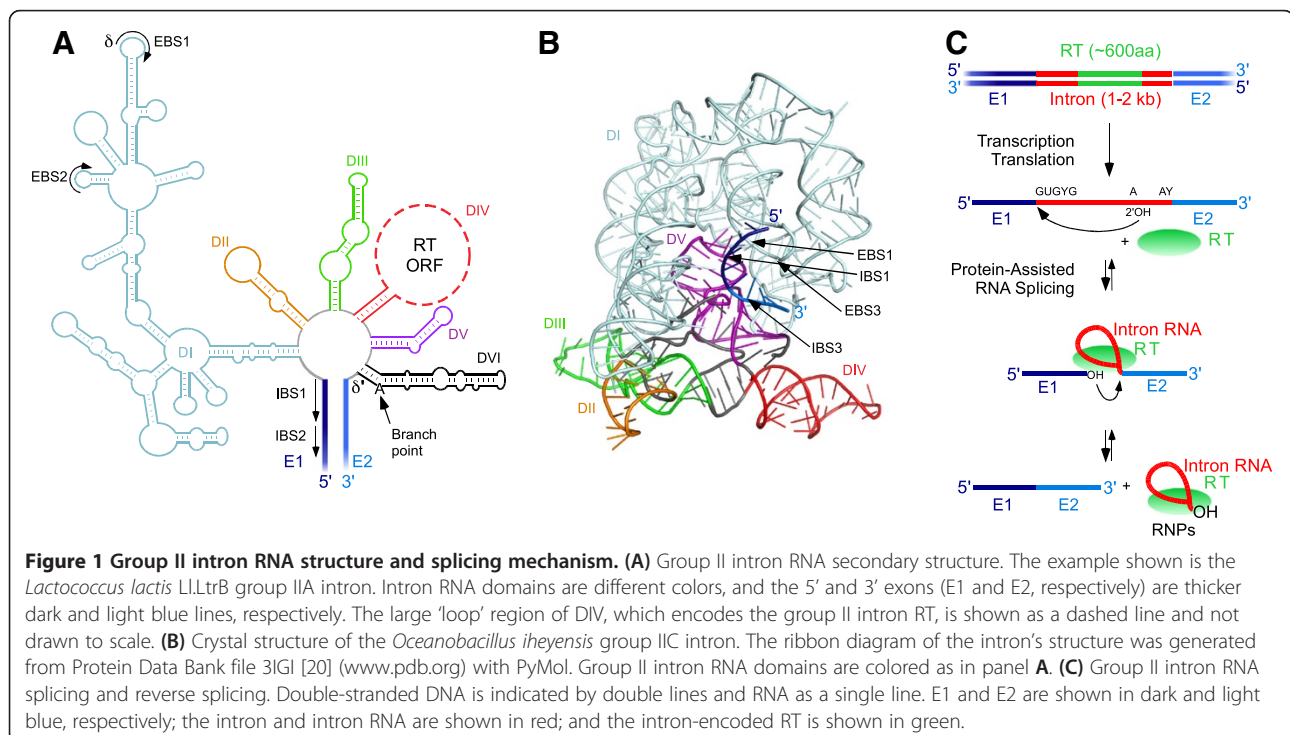
rearrangements, and progress in developing targetrons for gene targeting in eukaryotes. Finally, we discuss the development of thermostable group II intron RTs from bacterial thermophiles as new tools for cDNA synthesis, with potentially wide applications in research and biotechnology.

### Mobile group II introns

Mobile group II introns are found in bacteria, archaea, and the mitochondrial and chloroplast DNAs of some eukaryotes, and are thought to be evolutionary ancestors of spliceosomal introns, the spliceosome, retrotransposons, and retroviruses in higher organisms [3,14,15]. They are especially prevalent and widespread in bacteria, with hundreds of bacterial group II introns having been identified by genome sequencing [16].

Mobile group II introns consist of a catalytically active intron RNA, which encodes an RT (Figure 1) [1-3,17]. Group II intron RNAs have a length of 400 to 800 nts, excluding the ORF encoding the RT [3]. They have little sequence similarity to each other, but fold into a conserved three-dimensional structure consisting of six interacting double helical domains (DI-DVI) (Figure 1A and B) [17-21].

The folded group II intron RNA contains an active site that uses specifically bound  $Mg^{2+}$  ions to catalyze RNA splicing via two sequential transesterification reactions that yield ligated exons and an excised intron lariat RNA, the same reaction mechanism used for the splicing of nuclear spliceosomal introns in eukaryotes (Figure 1C) [1]. Because the transesterification reactions used for



splicing are reversible, the intron RNA can also catalyze reverse splicing of the intron into RNA or DNA sites containing the ligated exon sequence, with reverse splicing into DNA playing a key role in intron mobility. Both steps of reverse splicing (referred to as complete reverse splicing) result in the insertion of the excised intron RNA between the 5' and 3' exons, while the first step (referred to as partial reverse splicing) results in the attachment of the 3' end of the intron RNA to the 5' end of the downstream exon, leaving a strand break.

Some key regions of group II intron RNAs are DI, which contains the motifs that base pair with the DNA target site; DIV, which contains the ORF encoding the RT; DV, a metal-ion-binding domain that comprises most of the active site; and DVI, which contains the branch-point nucleotide [19]. Three subclasses of group II introns, denoted IIA, IIB, and IIC, have been distinguished by variations of the conserved RNA structure [3]. Crystal structures of a group IIC intron at different stages of reaction have been determined, providing insight into the nature of the active site and the mechanisms of RNA splicing and reverse splicing (Figure 1B) [19-21].

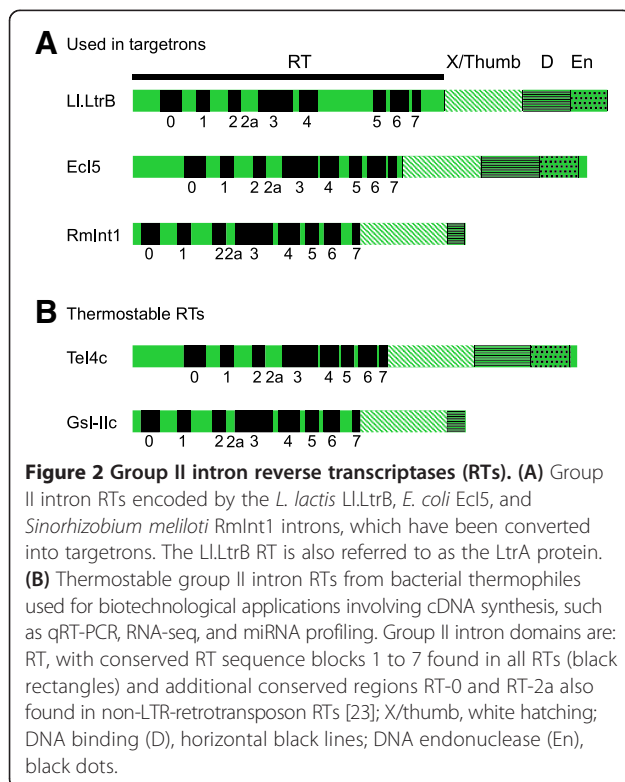
Group II intron RTs typically consist of 400 to 600 amino acids and contain a series of conserved motifs characteristic of retroviral and other RTs [3]. Figure 2 shows schematics of several group II intron RTs that are discussed in this review. Group II intron RTs contain conserved N-terminal RT and X domains, which correspond

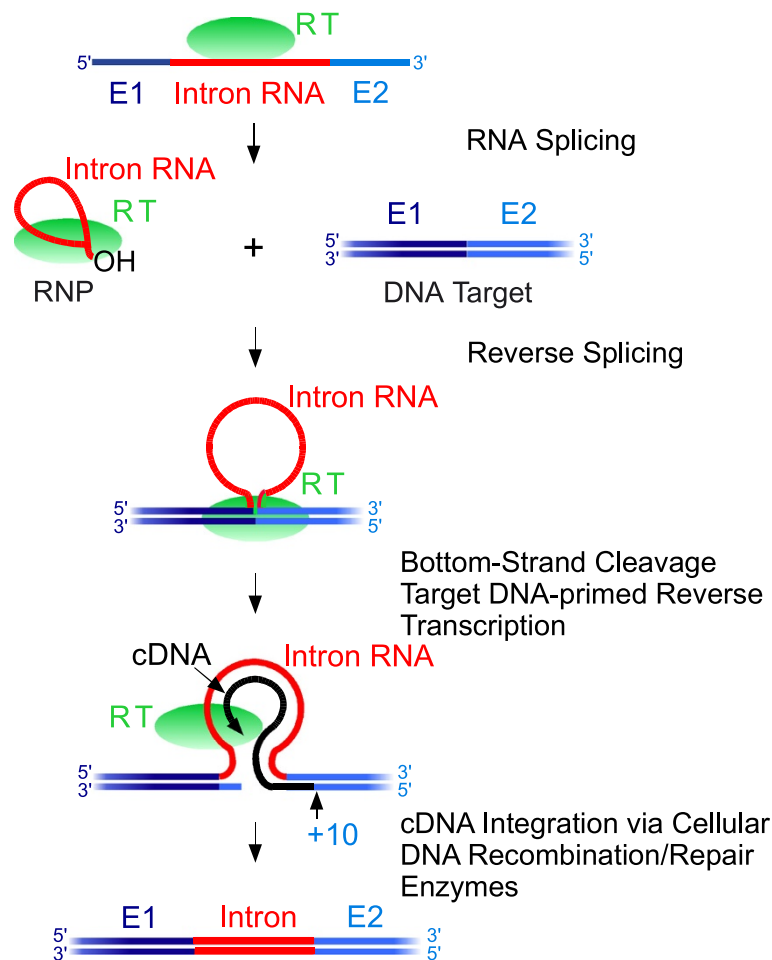
to the fingers/palm and thumb domains of retroviral RTs, respectively. In addition to reverse transcription, the RT and X domains bind specifically to the intron RNA to stabilize the active ribozyme structure for RNA splicing and reverse splicing (referred to as 'maturase' activity, an example of protein-assisted RNA catalysis). Group II intron RTs lack an RNase H domain, but typically have C-terminal DNA-binding (D) and DNA endonuclease (En) domains that interact with DNA target sites during retrohoming. Some IEPs, such as that encoded by RmInt1 from *Sinorhizobium meliloti* [22] (Figure 2A), lack the En domain. Notably, the RT and thumb domains of group II intron RTs are larger than those of retroviral RTs and contain an N-terminal extension and several distinctive 'insertions' between the conserved RT sequence blocks [23]. The larger RT and thumb domains may enable more extensive interactions with RNA templates and thus contribute to the high processivity of group II intron RTs (see below).

#### Group II intron retrohoming

Group II intron retrohoming has been reviewed in detail elsewhere [3,4], and here we describe only the major steps and variations pertinent to the mechanism of gene targeting. As shown in Figure 3, retrohoming starts with the group II intron splicing out of a larger RNA molecule, typically a transcript of the gene in which the group II intron is inserted. Splicing is accomplished via folding of the intron RNA into a catalytic structure, with help of the RT, which binds the intron RNA and stabilizes the active RNA tertiary structure. As discussed above, splicing occurs via two transesterification reactions that yield ligated exons and an excised intron lariat. After splicing, the RT remains tightly bound to the excised intron lariat RNA in a ribonucleoprotein (RNP) complex that initiates retrohoming by recognizing DNA target sequences by a combination of site-specific binding of the RT and base pairing of sequence motifs in the intron RNA, described in detailed below. The intron RNA then integrates directly into the DNA target site by full reverse splicing (see above), while the endonuclease activity of the RT cuts the opposite DNA strand slightly downstream of the insertion site, leaving an overhang with a cleaved 3' end that is used as a primer for synthesis of a cDNA copy of the inserted intron RNA by the RT [24-26]. Introns encoding RTs lacking the endonuclease activity retrohome by using nascent lagging or leading DNA strands at DNA replication forks as primers for reverse transcription [27-29].

The cDNA copy of the reverse-spliced intron RNA is integrated into the host genome by common cellular DNA recombination or repair mechanisms, a feature that contributes to the wide host range of group II introns. Recent findings have further elucidated late steps





**Figure 3 Group II intron retrohoming.** In the first step, the reverse transcriptase (RT) binds to the intron in a larger initial transcript of a gene and promotes RNA splicing, resulting in a ribonucleoprotein (RNP) complex that contains the excised intron lariat RNA and the tightly bound RT. RNPs recognize DNA target sites by using both the RT and base pairing of the intron RNA and then promote reverse splicing of the intron RNA into the top strand of the double-stranded DNA. After reverse splicing, the bottom DNA strand is cleaved by the En domain of the RT, and the 3' end generated at the cleavage site is used as a primer for target DNA-primed reverse transcription of the inserted intron RNA. The resulting intron cDNA (black) is integrated into the host genome by cellular DNA recombination or repair mechanisms.

in group II intron integration in *Escherichia coli*, in which a cellular RNase H degrades the intron RNA, and replication restart proteins then recruit the host replicative DNA polymerase, which synthesizes DNA corresponding to the sense strand of the intron [30]. Host nucleases trim DNA overhangs, and ligases repair remaining nicks [31].

Some group II introns splice via hydrolysis rather than branching and thus excise a linear rather than a lariat intron RNA [32,33]. During retrohoming, linear group II intron RNAs can carry out only the first step of reverse splicing, attaching the 3' end of the linear intron to the downstream DNA exon, which, combined with En cleavage of the opposite strand, yields a double-strand break that can be repaired by homologous recombination with exogenous DNA [34]. This double-strand break-stimulated recombination provides an alternative gene targeting mechanism for group II introns, analogous to that used by

Zn-finger nucleases, TALENs, and CRISPR-based systems [35]. In some hosts, the linear group II intron RNA inserted at a target site is reverse-transcribed to yield a cDNA that can be integrated into the genome by non-homologous end joining [36,37].

#### DNA-target site recognition

The key to using group II introns for gene targeting is their mode of DNA target site recognition. Group II intron RNPs recognize DNA target sequences by using both the RT and base pairing of the intron RNA, with the latter contributing most of the DNA target specificity [5,38]. Group IIA, IIB, and IIC introns differ somewhat in how they recognize DNA target sites, and these differences impact design and performance in the biotechnological context. The major target site interactions for the *Lactococcus lactis* LI.LtrB intron, the most widely



used for gene targeting, and several other group II introns used as targetrons are illustrated in Figure 4.

In group IIA introns, like the Ll.LtrB intron, the intron RNA contains three sequence motifs in DI that recognize DNA target sites by base pairing. These are denoted EBS1, EBS2, and  $\delta$ , and they base pair to complementary sequences in the DNA target site denoted IBS1, IBS2, and  $\delta'$  (where EBS stands for 'exon-binding site' and IBS stands for 'intron-binding site'; these same interactions also occur upon splicing out of a larger RNA molecule). The Ll.LtrB RT (denoted LtrA protein) recognizes nucleotides both upstream and downstream of the IBS/ $\delta'$  sequences (colored purple and blue, respectively, in Figure 4). Binding of the RT promotes DNA melting [39], enabling the intron RNA to base pair to the DNA target sequence, and DNA bending, which positions the target DNA properly for cleavage and priming of reverse transcription [40].

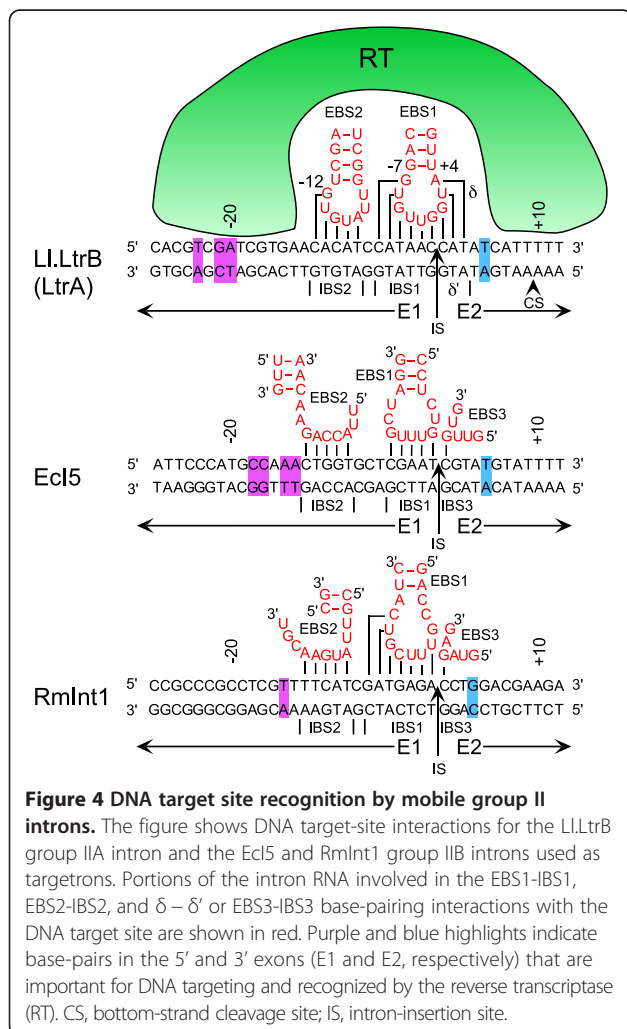
Group IIB introns, like Ecl5 and Rmlnt1, also contain three sequence elements that recognize the DNA target

site by base pairing. Specifically, EBS1, EBS2, and EBS3 base pair to corresponding IBS sequences in the target. The EBS3 sequence is located in a different part of the DI structure than the corresponding  $\delta$  sequence in group IIA introns [41]. The RT again recognizes flanking sequences. In Ecl5, a relatively well-studied example of this class [42], the RT recognizes a similar number of residues as the RT of Ll.LtrB, although the identities and locations of these residues differ. Rmlnt1, whose RT lacks the En domain, inserts into the single-stranded DNA formed during replication and thus has no requirement for DNA melting [28]. The Rmlnt1 RT recognizes only two critical nucleotide residues, but additional sequences may contribute [43].

Group IIC introns recognize short IBS1 and IBS3 sequences. A DNA hairpin, such as those found in gene terminators or phage attachment sites, is also a key recognition determinant and seems to take the place of the IBS2 sequence for these introns, although the mechanism of recognition is as yet unknown [44-46]. Group IIC introns can thus integrate into multiple sites, and specificity is limited.

Group IIA and IIB introns have high DNA-target specificity and integrate only rarely into ectopic sites (for example, retrotransposition of the Ll.LtrB intron into ectopic sites in the *E. coli* chromosome occurs at a frequency of 0.1 to  $30 \times 10^{-6}$ ) [3,47]. This high specificity reflects, in part, the fact that group II introns use both the RT and base pairing of the intron RNA to recognize their DNA target sequences, with the RTs of the Ll.LtrB and Ecl5 introns most stringently recognizing 4 to 5 nts and intron RNA base pairing extending over 11 to 14 nts spanning the intron-insertion site. Additionally, because the heteroduplex between the intron RNA and DNA target strand must bind to the intron RNA's active site for reverse splicing, mismatches in base pairing strongly affect the  $k_{cat}$  as well as  $K_m$  of the targeting reaction, providing greater discrimination against mispairings than can be obtained by binding affinity alone [48].

This intertwining of DNA target binding and catalysis differs from CRISPR-based systems, which have been used in bacteria and eukaryotes and also rely on base-pairing between RNA and DNA to provide specificity [49-55]. CRISPR systems use a guide RNA bound by a protein endonuclease (Cas9 being the canonical example) and can in theory target any stretch of twenty base pairs that is followed by a specific 'protospacer adjacent motif' (PAM), which in currently utilized systems is a stretch of two to five nts recognized by the endonuclease. However, the guide RNA does not play a catalytic role and thus specificity appears to be governed solely by its binding affinity to the DNA target site, with the protein endonuclease cutting anytime the RNA/



protein complex carries long enough at a given site. Indeed, concerns have been raised about the high off-targeting rate of these systems, with off-target sites having up to five mismatches found to be targeted at efficiencies similar to the intended site [56]. A further limitation for wide use in bacteria is that, unlike group II introns, CRISPR-based systems function only to introduce a double-strand break, and integration of exogenous DNA at the break site is dependent upon homologous recombination at a higher efficiency than is found in most bacterial species [53].

### Targetrons

Because mobile group II introns recognize their DNA target sites by a combination of base-pairing interactions and site-specific binding of the RT, the target site recognized by the RNP can be modified by finding other sites compatible with RT recognition and then changing the EBS/ $\delta$  sequences of the intron as necessary to match the new site [5]. Such retargeted mobile group II introns are called 'targetrons.' Group II introns that have been made into targetrons include both group IIA introns (Ll.LtrB [7]) and group IIB introns (Ecl5 [42] and RmInt1 [57]). Group IIC introns are less appealing as candidates for retargeting because they recognize hairpin structures via as yet unknown mechanisms. The Ll.LtrB targetron is commercially available through Sigma-Aldrich, and both the Ll.LtrB and Ecl5 targetrons are available through Targetronics.

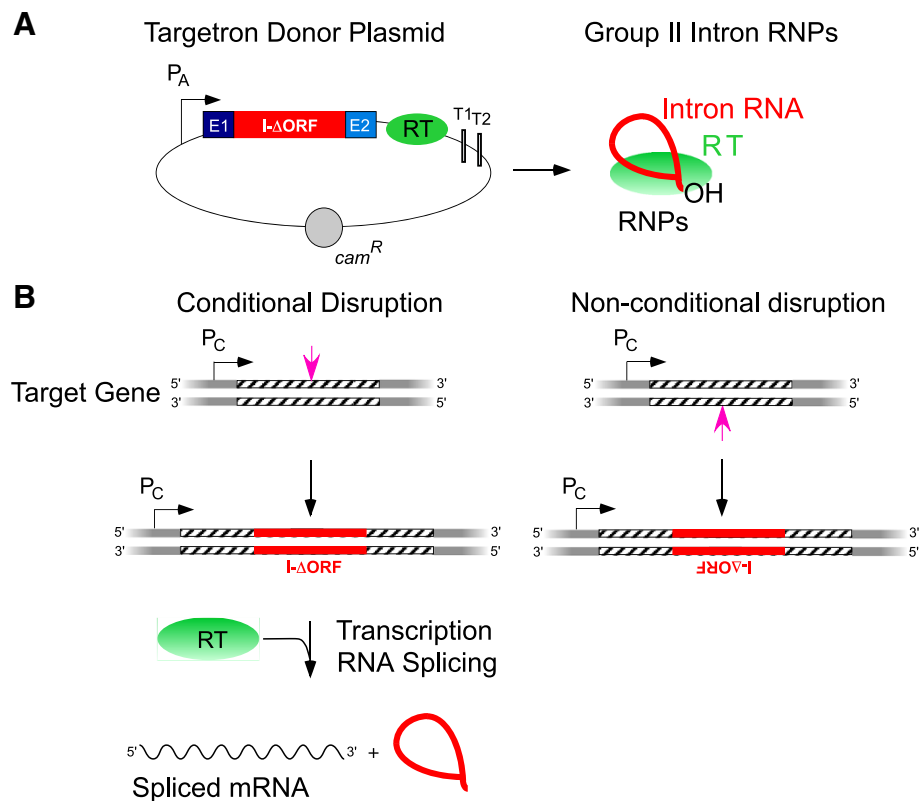
Although group II introns can and have been retargeted by the method mentioned above, in which the closest match to the native recognition site in a sequence to be targeted is identified, and the base-pairing sequences of the intron are modified to accommodate discrepancies, the rules by which introns recognize their target sites are actually more complex. For instance, the RT recognizes different residues at the target site with different stringencies, and none of these recognition events are absolutely required for retrohoming to occur [5,58,59]. If only the wild-type recognition sequence is used, then new targeting sites may be hard to come by, but knowing which bases can be varied and how is not a simple matter. The EBS/ $\delta$  sequences may also differ in the stringency of required base-pairing interactions at different positions. Algorithms have therefore been developed for retargeting the Ll.LtrB [7] and Ecl5 [42] introns. These algorithms were developed by examining libraries of inserted mobile group II introns with randomized base-pairing motifs for the most frequently conserved residues and base-pairing interactions, and using these frequencies to generate weighting schemes for the various interactions. Potential target sites are then assessed using the weighted criteria and assigned a score. Although the algorithms have limitations and do

not always correctly predict insertion frequency, typically a targetron efficient enough to be screened for site-specific insertion via colony PCR without selection can be found for any given stretch of 1,000 base pairs of DNA. Off-target integrations by the Ll.LtrB and Ecl5 targetrons are rare and can generally be avoided by the prudent step of scanning the genome for closely matching target sites. However, the specificity of targetrons could vary for different target sites, making it important to confirm desired single integrations by Southern hybridization.

The actual retargeting process is carried out by using PCRs that modify the EBS/ $\delta$  sequences within the intron to base pair to the DNA target site and simultaneously modify the IBS sequences upstream of the intron to base pair to the retargeted EBS sequences to allow the intron to splice out of a precursor RNA [6,7]. The PCR product corresponding to a segment of the intron and upstream exon is then cloned into a targetron expression vector (see below). Alternatively, the entire region covering the IBS1 and 2 and the EBS1, 2 and  $\delta$  sequences can be commercially synthesized in a single DNA molecule (for example, as a gBlock sold by IDT) that can be cloned directly into the vector [60]. The outlying  $\delta'$  or EBS3/IBS3 positions are typically adjusted by cloning the PCR product into one of four parallel targetron vectors already containing the correct bases for these interactions.

For biotechnological applications, targetrons are typically expressed from a donor plasmid that is transformed or conjugated into the desired host (Figure 5A). In *E. coli*, targetron donor plasmids have used a T7 promoter driven by T7 RNA polymerase integrated into the chromosome or expressed from a separate plasmid [6]. However, endogenous host or plasmid promoters can also be used in *E. coli* and are commonly employed for targetron expression in other bacteria [61-63]. A broad-host-range targetron expression plasmid, pBL1, uses an m-toluic acid-inducible promoter, which is not dependent upon specific host factors for induction [64]. The typical configuration for the targetron cassette is one in which the ORF encoding the RT is removed from DIV of the intron and expressed in tandem. This increases the efficiency of retrohoming and allows for disruptions of the targeted gene to be either conditional or non-conditional, depending on whether the intron is targeted to insert into the sense or antisense strand of the gene and whether or not the RT remains present to aid in splicing of the intron from the mRNA (Figure 5B) [61,62].

It is also possible to select for targetron integration by using a retrotransposition-activated marker (RAM) [59,63] (Figure 6). This involves including in the targetron a selectable marker, such as a *kan*<sup>R</sup> antibiotic-resistance gene, that is inserted in the antisense orientation and is itself disrupted by a self-splicing group I intron in the sense orientation, a configuration first developed to detect rare



**Figure 5 Targetron donor plasmid and use of targetrons for conditional and non-conditional gene disruptions.** (A) Targetron donor plasmid. The plasmid expresses a modified group II intron with the reverse transcriptase (RT) ORF deleted (I-ΔORF) and flanked by short exons under control of an active promoter ( $P_A$ ), which can be either inducible or constitutive. The RT ORF is expressed in tandem from a location just downstream of E2. Protein-assisted splicing of the primary transcript produces a ribonucleoprotein (RNP) complex, which contains the group II intron RT bound to the excised intron lariat RNA and which promotes site-specific integration of the intron into DNA target sites via retrohoming (see Figure 3). (B) Use of targetrons for conditional and non-conditional gene disruptions. Conditional disruptions are obtained when the intron is targeted to insert into the top or sense strand of the target gene. Transcription of the target gene from its own promoter in the host chromosome ( $P_C$ ) results in a primary transcript from which the intron can be removed by providing the RT, which promotes protein-assisted RNA splicing. Non-conditional disruptions are obtained by targeting of the intron to the bottom or antisense strand, which results in the insertion of the intron in an antisense orientation relative to that of the target gene. Transcription of the target gene then yields a primary transcript containing the complement of the intron, which is inactive and cannot be removed by RNA splicing.

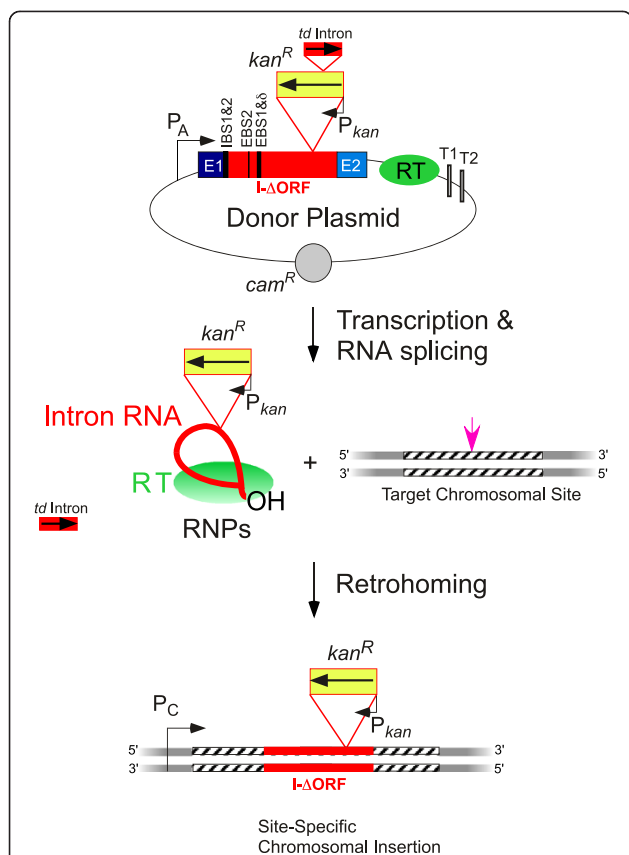
retrotranspositions to ectopic sites [65]. The marker can only be expressed after splicing out of the group I intron and reverse transcription of the RNA intermediate into DNA, as occurs during the process of retrohoming. An Ll.LtrB targetron containing a trimethoprim-resistance-*RAM* ( $Tp^R$ -RAM) and randomized EBS/ $\delta$  sequences was used to construct an *E. coli* gene disruption library [59]. After targetron expression,  $Tp^R$  colonies contained targetrons inserted into different genes with complementary IBS/ $\delta'$  sequences, and these validated targetrons could be recovered by simple PCR and used to obtain the same disruption in other *E. coli* strains [66], providing an alternative to the use of a targeting algorithm.

#### Targetron use in bacteria

Targetrons have been used in a wide range of bacteria, including medically and commercially important species

that had been recalcitrant to gene targeting by other methods (Table 1). Compared to other bacterial gene targeting methods, advantages of targetrons are their wide host range (the Ll.LtrB targetron works in virtually all bacteria), ability to function in either  $RecA^+$  or  $RecA^-$  hosts, very high integration efficiencies (typically 1 to 100% without selection), and ease of retargeting via a computer algorithm and simple PCR reactions that are amenable to high-throughput approaches.

It is relatively simple to tailor the commercially available Ll.LtrB or EcI5 targetron cassettes for use in different bacterial hosts. This typically requires re-cloning the targetron cassette from the provided donor plasmid into an established host-specific or broad-host-range expression plasmid with a strong promoter. Continuous targetron expression, which can lead to off-target integrations, can be avoided by using an inducible promoter or a donor



**Figure 6 Use of a retrotranscription-activated marker (RAM) to select for targetron integrations.** A targetron with a RAM cassette is expressed from a donor plasmid. The RAM cassette consists of a selectable marker gene, such as an antibiotic-resistance gene (*kan<sup>R</sup>*), inserted within the intron DIV in the antisense orientation, but interrupted by an efficiently self-splicing group I intron (the phage T4 *td* intron) in the sense orientation, thereby preventing expression of the *kan<sup>R</sup>* marker gene from the donor plasmid. Transcription of the targetron carrying the RAM from the donor plasmid yields a primary transcript from which the group I intron is spliced to reconstitute the *kan<sup>R</sup>* marker and the group II intron is spliced to yield RNPs that retrohome into a DNA target site. After retrohoming, the reconstituted *kan<sup>R</sup>* marker is expressed from its own promoter, enabling selection for targetron integrations.

plasmid that is readily curable in the absence of selection. A RAM capable of functioning in the desired bacteria can also be introduced into the intron, but targeting frequencies are typically high enough to screen for targetron insertions by colony PCR, making such a marker dispensable. The ClosTron, which has made possible gene targeting in a wide range of notoriously difficult *Clostridium* spp., is a highly successful example of adaptation of the Ll.LtrB targetron from a commercial kit [63,80], and similar adaptations of the Ll.LtrB targetron have been made for a variety of other bacteria (for example, [62,64,71,81]). Because the initial reverse splicing and target-DNA-primed reverse transcription reactions are catalyzed by group II intron RNPs, and because the late steps of second-strand

**Table 1 Bacteria in which targetrons have been used successfully**

Genus	Primary references
<i>Agrobacterium</i>	[64]
<i>Azospirillum</i>	[67]
<i>Bacillus</i>	[68]
<i>Clostridium</i>	[63,69]
<i>Ehrlichia</i>	[70]
<i>Escherichia</i>	[6,42]
<i>Francisella</i>	[71]
<i>Lactococcus</i>	[61]
<i>Listeria</i>	[72]
<i>Paenibacillus</i>	[73]
<i>Pasteurella</i>	[74]
<i>Proteus</i>	[75]
<i>Pseudomonas</i>	[64]
<i>Ralstonia</i>	[76]
<i>Salmonella</i>	[6]
<i>Shewanella</i>	[60]
<i>Shigella</i>	[6]
<i>Sinorhizobium</i>	[57]
<i>Sodalis</i>	[77]
<i>Staphylococcus</i>	[62]
<i>Vibrio</i>	[78]
<i>Yersinia</i>	[79]

synthesis and cDNA integration are performed by common host factors [30,31,37,82,83], there are in principle no limitations to the number of bacterial species in which targetrons might function. As mobile group II introns are present in the genomes of some archaea [84], it seems likely that targetrons will prove useful in archaea, as well.

#### Applications of targetrons in bacteria

Targetrons have most frequently been used to generate knock-outs in bacteria. A great deal of work has been done using this method, with examples including identifying virulence factors [70,72,74,85-88] and potential drug targets [89,90], and examining the combinatorial effect of different genomic loci on protein expression [91].

Targetrons have been particularly widely used in strains of the genus *Clostridium*. Suicide plasmids were previously the only method of utility in these strains [63], but since Clostridia typically have very low transformations frequencies (for instance, more than a milligram of plasmid is required to transform *Clostridium acetobutylicum* [92]), suicide plasmids are difficult to use in these organisms. Targetrons have thus greatly increased our understanding of and ability to engineer Clostridia, many of which are of medical and industrial



importance. For instance, Clostridia include a number of biofuel-producing strains, and targetrons have come into frequent use to aid in understanding the metabolism of these strains and to engineer them for higher yields [92-110]. Targetron-mediated knockouts have been used in a large number of studies on sporulation, germination, and other aspects of the biology of *Clostridium difficile*, a leading cause of diarrhea in hospitals [88,111-143]. Targetron technology has also benefitted the study of toxin production, sporulation, and other biological processes in *Clostridium botulinum* [144-153], *Clostridium perfringens* [69,85,154-164], and *Clostridium sordellii* [87,165]. Work on developing targetrons for the thermophilic bacterium *Clostridium thermocellum* is discussed in more detail below.

Many bacteria of interest are difficult to transform due to restriction-modification systems. In *Staphylococcus aureus* [81], *Clostridium acetobutylicum* [166], and *Clostridium cellulolyticum* [167], targetrons were used to knock out restriction enzymes, thereby opening clinical and environmental isolates to systematic mutational analysis. Besides *S. aureus* and the *Clostridium* species mentioned previously, targetrons have been developed for use in other pathogenic bacteria, such as *Francisella tularensis* [71], *Bacillus anthracis* [68,168], *Listeria monocytogenes* [72], *Pasteurella multocida* [74], *Vibrio cholerae* [78], and *Ehrlichia chaffeensis* [70], opening up the possibility of using targetrons to develop vaccine strains of these organisms.

Targetrons have also been used to deliver cargo genes, including genes for fluorescent proteins [91], phage resistance [61], and antigens for release into a host's digestive system as a live vaccine [169]. Unstructured sequences of less than 100 nts in length can usually be carried without impacting intron mobility. Longer sequences may impair functionality, and sequences above 1,000 nts can drastically decrease efficiency. DIV, particularly the DIVb loop, has been shown to be the best location to insert cargo genes for minimal impact on intron mobility [170]. Targetrons have also been used to induce targeted genomic deletions via homologous recombination, albeit at much lower efficiencies than are possible in tandem use with recombinases [171].

Finally, the relative simplicity of targetron retargeting, combined with the falling costs of gene synthesis [172] and the increasing ability to automate the laboratory techniques involved [173,174], opens the door to high-throughput construction of targetrons for simultaneous integration into a multiplicity of loci. Applications could include rapidly generating whole-genome knock-out libraries for novel organisms and testing in parallel different combinations of mutants discovered in random screens in order to, for example, improve the yield of a target metabolite or develop a suitable vaccine strain for

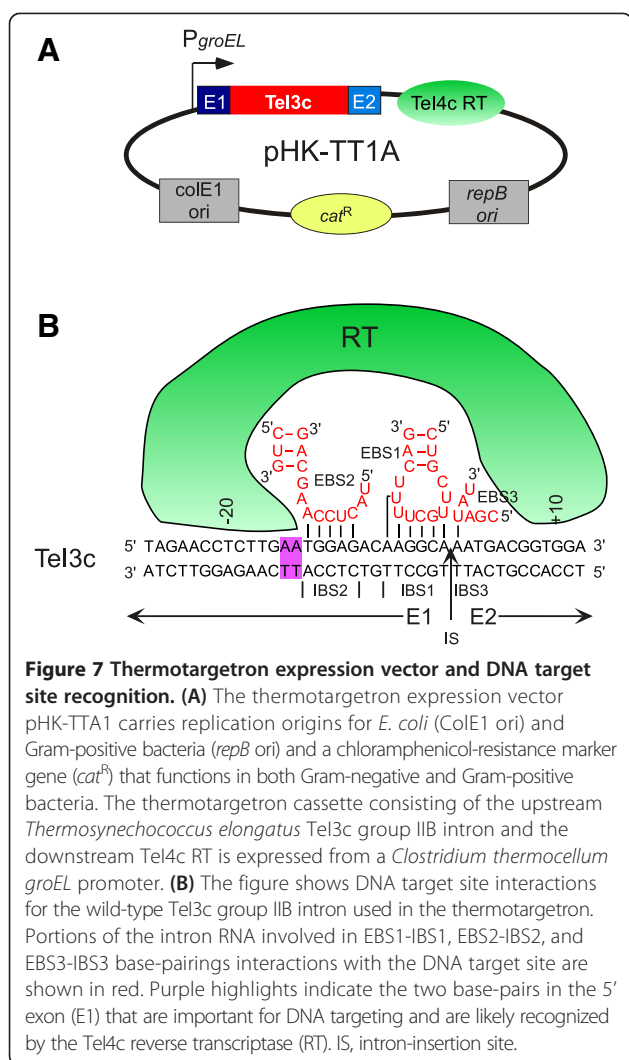
a pathogenic organism. Two other recent extensions of targetron technology in bacteria are discussed below.

#### **A thermotargetron for gene targeting in thermophiles**

Bacterial thermophiles are widely used for the production of chemicals and thermostable proteins but in many cases have inefficient transformation systems and have proven difficult to genetically engineer by conventional methods relying on homologous recombination [175-177]. Recently, a thermotargetron for gene targeting in thermophiles was developed based on a group IIB intron (denoted TeI3c) and RT (TeI4c RT) from the thermophilic cyanobacterium *Thermosynechococcus elongatus* [178] (Figure 7A; see also the diagram of the TeI4c RT in Figure 2B). Unlike other group II introns that have been made into targetrons, TeI3c is a naturally ORF-less group II intron, and TeI4c is an RT that is encoded by another group II intron but mobilizes TeI3c efficiently.

This TeI3c/4c thermotargetron was used for efficient gene targeting in *Clostridium thermocellum*, an organism used in the consolidated bioprocessing of lignocellulose biomass [178,179]. Like many species of Clostridia, *C. thermocellum* has low, variable transformation frequencies. An important feature of the thermotargetron is its high integration efficiency, 67 to 100% without selection for seven successful gene disruptions, making it possible to identify disruptants by colony PCR of only a small number of transformants. Gene disruptions that block pathways leading to by-products of cellulose degradation increased cellulolytic ethanol production in *C. thermocellum* [178].

Another notable feature of the thermotargetron is that it recognizes DNA target sites almost entirely by base pairing of the intron RNA (11-bp), while the RT recognizes only two bases (Figure 7B). The contribution of the RT to DNA melting appears to be largely dispensable at higher temperatures. This feature is advantageous because it increases the number of potential target sites and should facilitate the targeting of short ORFs and small non-coding RNAs, not only in thermophiles but also potentially in mesophiles that can tolerate short times at elevated temperatures (45 to 48°C). A downside of the more limited protein recognition, however, is that it decreases DNA target specificity, thus requiring greater attention to targetron design to avoid integration into closely matching off-target sites. The decreased target specificity may also contribute to the lower success rate for gene disruptions (7 of 25 targetrons in initial tests), which could be due in part to deleterious off-target integrations. This situation should be ameliorated by the development of algorithms to minimize off target integrations, as done for other targetrons. The TeI3c/4c thermotargetron functions in both Gram-negative and Gram-positive bacteria and should be adaptable to a wide variety of thermophiles.



### Use of targetrons for large-scale genome engineering

Targetrons have recently been adapted for carrying *lox* sites to facilitate large-scale genome engineering [60]. While recombinase sites have been previously included in targetrons, they had rarely been used for any purpose other than removing selectable markers after integration [59,80]. *Lox* sites and other recombinase recognition motifs with palindromic sequences can form stable hairpin structures upon transcription into RNA. In the absence of a selectable marker, such hairpin structures can significantly impair the functionality of both the Ll.LtrB and EcI5 targetrons. This effect was mitigated by adding non-base-pairing regions to the base of the hairpin structures, which presumably made the hairpins more flexible, such that they no longer interfered with the catalytic structures of the intron. These results point out the importance of considering structure when designing targetrons to carry cargo.

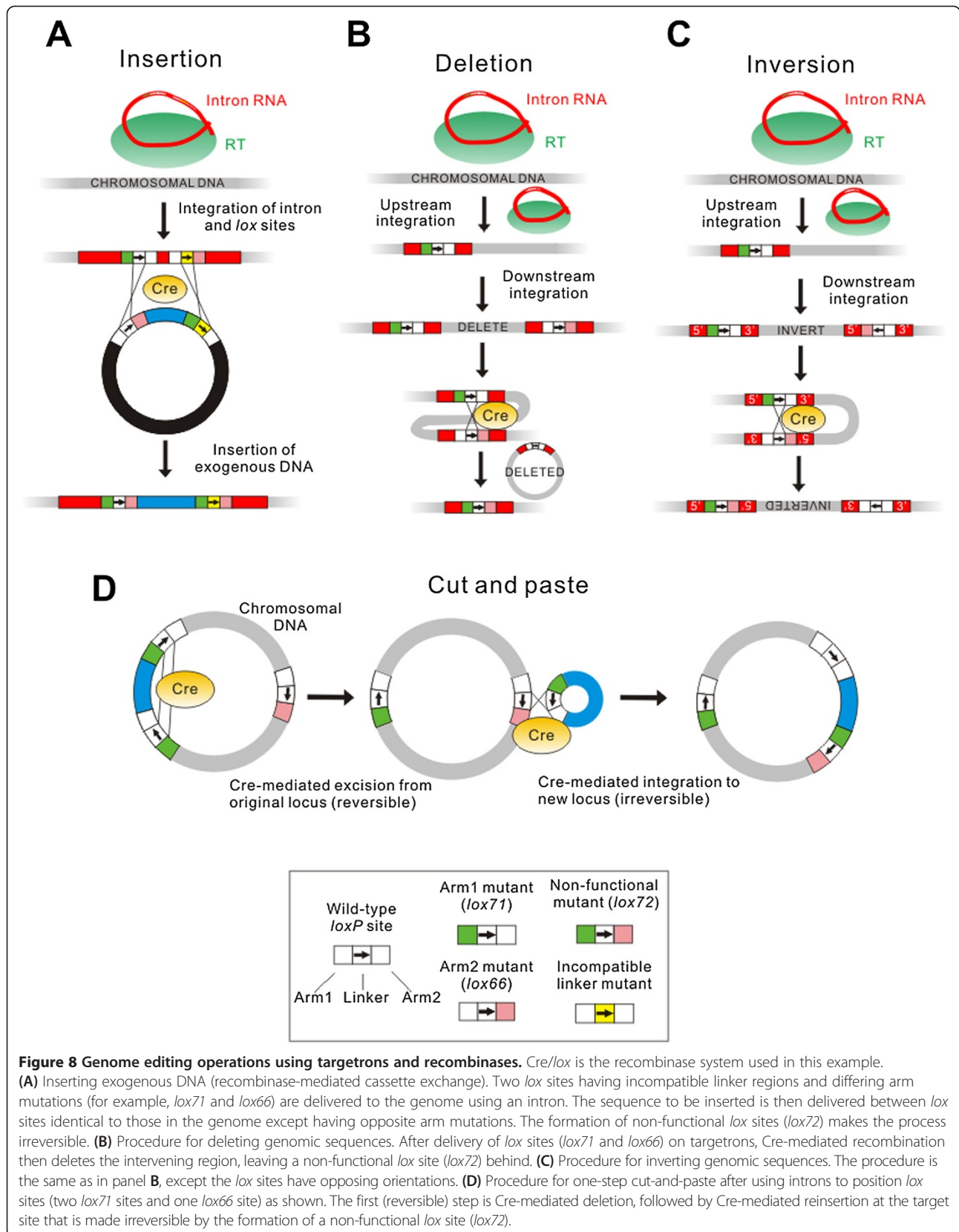
Since both targetrons and the *Cre/lox* system [180] function well in a wide variety of organisms, the combination

of both allows for a powerful and generalized genome engineering system, where previously engineering solutions typically needed to be developed for each organism. Once the *lox* or other recombinase sites are positioned using the targetrons, a variety of operations are possible. Figure 8 shows schematics for using the system to engineer large-scale insertions, deletions, inversions, and one-step 'cut-and-paste' operations in which large DNA segments are translocated from one genomic site to another. Manipulations of any size are possible, within the constraints of gene content and genome structure. Among other manipulations, the system was used to deliver a 12-kb polyketide synthase operon to the genome of *E. coli*; move 100 kb of the *E. coli* genome to another locus 1.5 Mb away; delete a pathogenicity island from *Staphylococcus aureus*; and invert approximately one third of the *Bacillus subtilis* genome. Intragenomic recombinations mediated by *Cre-lox* occurred at essentially 100% efficiency, and intermolecular recombinations occurred at 40 to 100% efficiency, without the need to place any selectable markers in the genome. The method can be expected to function in any organism in which targetrons can be made to work.

These examples are likely but the first in a series of innovations that will allow targetrons to be used for large-scale genomic engineering. There are currently few alternatives that allow the facile, site-specific introduction of genetic material into microorganisms. While some organisms, such as *Streptococcus pneumoniae* [181] and *Acinetobacter* [182], have relatively robust systems for homologous recombination, most others do not. Similarly, while methods such as recombineering [183,184] and MAGE [174] have been developed that allow PCR products and oligonucleotides to be readily introduced into *E. coli* in a site-specific manner, these methods do not scale to most other microorganisms. Targetrons are essentially the only tool that can be used to site-specifically 'punctuate' the genomes of a wide array of bacteria, as has previously been observed for recalcitrant thermophilic strains and Clostridia, discussed elsewhere in this review. While *lox* sites have been introduced to promote site-specific recombination, the option also exists to introduce a wide variety of other short genetic elements that can impact phenotype, including promoters, terminators, leader sequences, affinity tags, and even origins of replication. The use of targetron libraries [59,66] to seek out sites that lead to improved functionality, combined with the use of efficient targetron insertion to rapidly introduce multiple targetrons into a single strain, either serially or in parallel, makes targetrons the tool of choice for the engineering of industrially-relevant microorganisms.

### Prospects for targetron use in eukaryotes

Although efficient eukaryotic gene targeting technologies have been developed, including Zn-finger nucleases,



TALENS, and CRISPR-based systems, targetrons offer the advantages of greater ease of retargeting than Zn-finger nucleases or TALENS and potentially higher DNA target specificity than any of the other methods. However, the barriers to targetron use in eukaryotes include the requirement for delivering RNPs containing a large, structured group II intron RNA to the nucleus, as well as the relatively high  $Mg^{2+}$  concentrations required for group II intron RNA function. Group II introns evolved to function in bacteria, whose free  $Mg^{2+}$  concentrations are generally 1 to 4 mM [185], whereas in eukaryotes,  $Mg^{2+}$  concentrations are <1 mM and possibly lower in nuclei, where  $Mg^{2+}$  is sequestered by binding to large amounts of DNA [186,187]. These lower  $Mg^{2+}$  concentrations constitute a barrier to group II intron invasion of nuclear genomes and limit their efficiency for gene targeting in eukaryotes. Additional host defense and innate immunity mechanisms could also come into play.

Initial studies showed that Ll.LtrB targetron RNPs introduced into mammalian cells by transfection could integrate into separately transfected plasmid target sites albeit at low efficiency [5] and envisioned methods that might be used for targeted repair of human genes [188]. In a later systematic study testing the feasibility of using targetrons in eukaryotes, Ll.LtrB targetron RNPs were microinjected directly into *Xenopus laevis* oocyte nuclei and tested for retrohoming and gene targeting via double-strand-break-stimulated homologous recombination in plasmid assays [34]. These studies showed that retrohoming and targeting via group II intron-stimulated homologous recombination occurred efficiently (up to 38% and 4.8% of plasmid target sites, respectively), but required the injection of additional  $Mg^{2+}$ , sufficient to obtain an intracellular concentration of 5 to 10 mM. A similar requirement for the injection of additional  $Mg^{2+}$  for retrohoming was found for targetron RNPs injected into *Drosophila* and zebrafish embryos [34]. Injection of targetron RNPs plus  $Mg^{2+}$  have given targeted integration into the chromosomal yellow gene in flies at frequencies up to 0.12% in pooled embryos and 0.021% in pooled adult flies [189], and in *X. laevis*, a different approach, using group II intron RNPs for site-specific DNA modification in sperm nuclei followed by *in vitro* fertilization, gave targeted integration at frequencies sufficiently high to detect knockouts in a single copy *mitF* gene by PCR screening of tail clippings (M Mastroianni, J Yao, and AM Lambowitz, unpublished data). However, the frequencies were variable and further improvements in efficiency and reliability are needed for these to become routine methods.

There is some prospect that more active group II introns with enhanced retrohoming in eukaryotes can be selected by directed evolution approaches. Recent work showed that Ll.LtrB introns that retrohome more efficiently

at lower  $Mg^{2+}$  could be selected in an *E. coli* mutant deficient in  $Mg^{2+}$ -transport [190], laying the groundwork for direct selections of group II introns that function more efficiently in eukaryotic cells. The recent group II intron RNA X-ray crystal structures [19-21] may also enable rational design approaches to enhancing group II intron function. If these efforts prove successful, the same rationales that are driving the use of targetrons for genomic engineering in bacteria will extend to genomic engineering in eukaryotes.

#### **Thermostable group II intron reverse transcriptases**

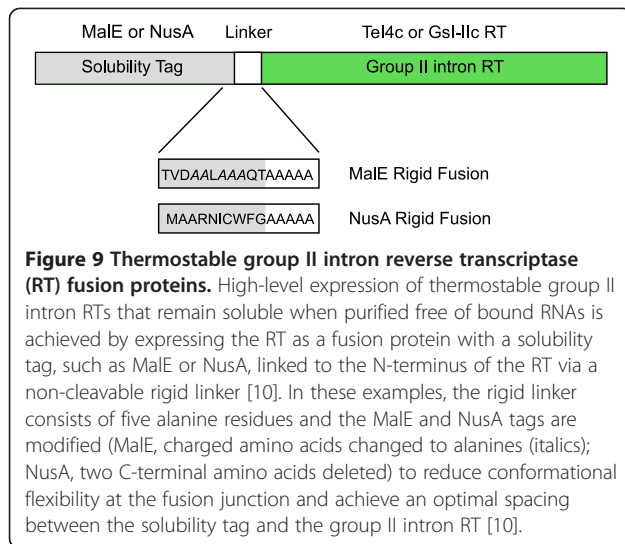
Reverse transcriptases are widely used in biotechnology for applications involving cDNA synthesis, such as qRT-PCR and RNA-seq. Most if not all of these applications would benefit from using RTs that synthesize cDNAs with high processivity and fidelity. However, the retroviral RTs that are commonly used for these methods have inherently low fidelity and processivity, reflecting that these enzymes evolved to help retroviruses evade host defenses by introducing sequence variations and rapidly propagating successful ones by RNA recombination [191]. Vast sums have been expended to engineer variants of retroviral RTs that overcome these inherent deficiencies.

By contrast, group II intron RTs evolved to have high processivity and fidelity, reflecting their function in retrohoming, which requires synthesis of an accurate, full-length cDNA copy of a highly structured group II intron RNA [8,9]. Other advantageous characteristics of group II intron RTs are their lack of RNase H activity, which enables reuse of RNA templates, and their difficulty in initiating on DNA templates, which preserves RNA strand information by minimizing recopying of cDNAs [23,31].

In a recent technical advance that makes group II intron RTs available for widespread use as tools for research and biotechnology, general methods were developed that enable their high-level expression in bacteria and their purification in active form free of tightly bound RNA [10]. These methods involve the expression of group II intron RTs as fusion proteins with a solubility tag, such as MalE or NusA, attached to the N-terminus of the RT via a non-cleavable rigid linker (Figure 9). The attached solubility tag enables the protein to remain soluble when freed of the intron RNA, and the rigid linker minimizes interference of the tag with RT function.

By applying the above methods to group II intron RTs from the bacterial thermophiles *Thermosynechococcus elongatus* and *Geobacillus stearothermophilus* (Figure 2B), it was possible to obtain thermostable group II intron RT fusion proteins that synthesize cDNAs at temperatures up to 81°C with much higher processivity and two- to four-fold higher fidelity than retroviral RTs [10]. The high

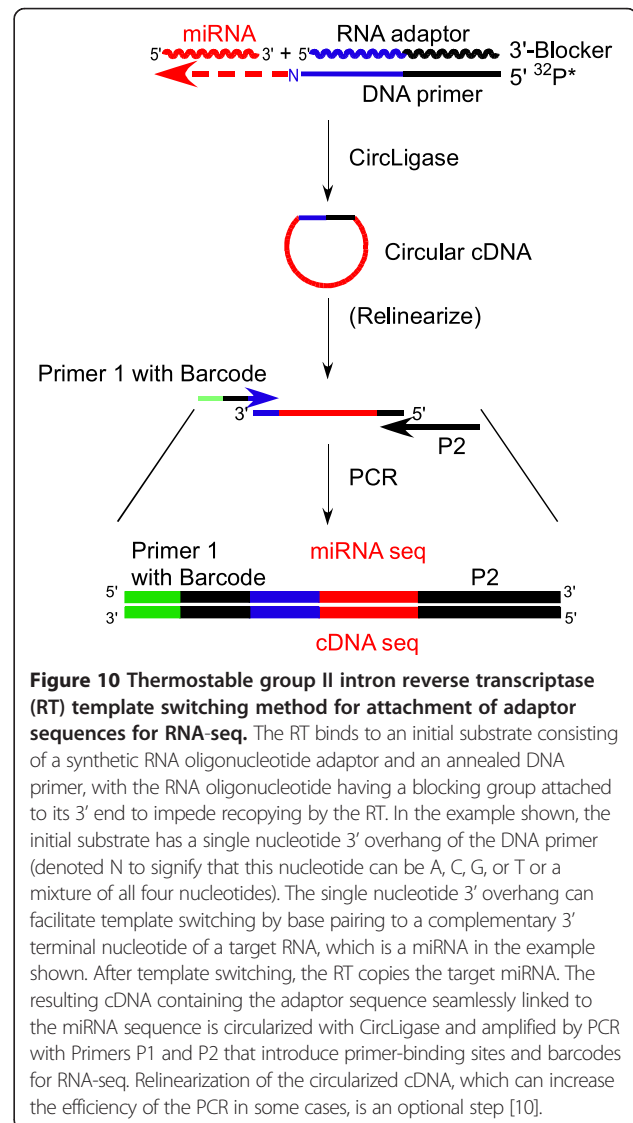




processivity of group II intron RTs is advantageous for synthesizing long cDNAs that preserve information about alternatively spliced RNAs and for RNA footprinting and structure mapping using RNA modification reagents, where premature terminations by retroviral RTs at unmodified sites result in high background and loss of information. The higher fidelity of group II intron RT should benefit applications, such as tumor profiling, that require the analysis of sequence variants.

We are still at the early stages of developing methods and applications utilizing the novel properties of these enzymes. Group II intron RTs behave differently from retroviral RTs, both in terms of optimal conditions for different applications and their tighter binding to nucleic acids, which necessitates different types of clean-up procedures for cDNA products. Consequently, group II intron RTs cannot simply be substituted into protocols developed for retroviral RTs and must be optimized for each application. In a published application, a thermostable group II intron RT was used to generate RNA-seq libraries of human mRNAs, using an oligo(dT)<sub>42</sub> primer [10]. The resulting libraries showed relatively uniform 5' to 3' coverage of all size classes of human mRNAs, including those >7 kb, whereas parallel libraries prepared using the thermostable retroviral RT, SuperScript III, showed a strong bias for reads near the 3' ends of mRNAs, reflecting premature terminations. The ability to obtain RNA-seq libraries with uniform 5' to 3' coverage using an oligo(dT) primer avoids steps such as ribodepletion and RNA fragmentation, which are needed to minimize rRNA contamination and obtain uniform coverage in libraries prepared using retroviral RTs. The minimal manipulation needed to prepare whole cell RNA-seq libraries using group II intron RTs may be useful for procedures that start with small amounts of RNA, such as transcriptome analysis from single cells.

In addition to their higher processivity and fidelity than retroviral RTs, group II intron RTs have a novel end-to-end template switching activity in which the RT synthesizes a cDNA copy of one template and then switches directly to the 3' end of a second template [10]. As illustrated in Figure 10, this template-switching activity can be used to seamlessly link RNA-seq adaptor sequences containing primer-binding sites and barcodes to cDNAs during reverse transcription, thereby avoiding an additional inefficient and bias-inducing step of using RNA ligase for linker ligation. In the example shown, the group II intron RT initiates from a short synthetic RNA oligonucleotide adaptor with an annealed DNA primer. It then switches to the 3' end of a target miRNA, yielding a continuous cDNA containing the RNA-seq adaptor sequence seamlessly linked to the miRNA sequence.





Like other DNA and RNA polymerases, group II intron RTs are prone to add extra non-templated nucleotides to DNA upon reaching the end of an RNA template (ref. [10] and references therein), which could lead to non-seamless junctions and biases during template switching. This problem is avoided by using an initial template/primer substrate consisting of a synthetic RNA oligonucleotide annealed to a DNA primer that leaves a single nucleotide 3' overhang. This 3' overhang nucleotide base pairs with the 3' terminal nucleotide of the second RNA template, resulting in a seamless switch to the second template. A specific 3' overhang nucleotide can be used to direct the RT to a specific target RNA, while a mixture of 3' overhang nucleotides is used to minimize bias for mixtures of templates having different 3' RNA ends.

After template switching, the resulting cDNA linked to adapter sequences is circularized with CircLigase and PCR amplified to generate an RNA-seq library (Figure 10). By incorporating an additional step for removal of a 3' phosphate of the target RNAs, the method can also be applied to protein- and ribosome-bound RNA fragments in procedures such as HITS-CLIP, CRAC, RIP-Seq, and ribosome profiling. Because thermostable group II intron RTs can template-switch to RNAs with modified 3' ends and reverse transcribe through highly structured RNAs, the method can potentially lead to the identification of novel miRNAs and other non-coding RNAs that cannot be cloned by current methods using retroviral RTs.

## Conclusions

The biotechnological applications of mobile group II introns and their RTs are an example of how basic research into biochemical mechanisms and evolution can lead to unexpected practical applications. Targetrons, which arose from studies of the mechanism of group II intron mobility, now provide a broad-host-range solution to knock-outs and, when combined with other technologies, such as site-specific recombinases, can be employed to make a wide variety of changes in almost any bacteria, including previously recalcitrant medically and industrially important species. Together with the prospect of targetron-mediated mutagenesis in archaea and the possibility of adapting targetrons for use in eukaryotes, targetrons are well-positioned to play an expanding role in the analysis and engineering of genomes for biotechnological and medical applications. The unique properties of group II RTs, enzymes that were discovered solely as a consequence of basic research, may ameliorate many of the problems of current *in vitro* methodologies for RNA analysis, qRT-PCR, and RNA-seq, with potentially wide applications in biomedical research, diagnostics, and biotechnology.

## Abbreviations

CRAC: Cross-linking and analysis of cDNA; DI-DVI: Group II intron RNA domains I-VI; D: DNA-binding domain of group II intron reverse transcriptases; E1 and E2: 5' and 3' exons; EBS: Exon-binding site; En: DNA endonuclease domain of group II intron reverse transcriptases; HITS-CLIP: High-throughput sequencing by cross-linking immunoprecipitation; IBS: Intron-binding site; I-ΔORF: Group II intron with ORF encoding the RT deleted; IEP: Intron-encoded protein; PAM: Protospacer adjacent motif; RAM: Retrotransposition-activated marker; RIP-Seq: RNA immunoprecipitation and sequencing; RNP: Ribonucleoprotein; RT: Reverse transcriptase; Tp: Trimethoprim.

## Competing interests

Targetron technology and thermostable group II intron reverse transcriptases and methods for their use are subject to issued U.S. and international patents and patent applications that are licensed by the Ohio State University and the University of Texas to InGex, LLC, which sublicenses these technologies to others for commercial applications. GM, AML, the Ohio State University, and the University of Texas are minority equity holders in InGex, LLC, and AML serves as an advisor to InGex LLC. GM and AML receive royalties for commercial use of the technologies.

## Authors' contributions

PJE, GM, ADE, and AML all contributed to the writing of this manuscript. All authors read and approved the final manuscript.

## Acknowledgements

We thank Dr. Gregory Davis (Sigma-Aldrich) for comments on the manuscript. Research on targetrons and group II intron reverse transcriptases in AML's laboratory is supported by NIH grants GM037949 and GM037951 and Welch Foundation grant F-1607. Research on targetrons in ADE's laboratory has been supported by a National Science Foundation Graduate Research Fellowship for PJE under Grant No. DGE-1110007, a National Security Science and Engineering Faculty Fellowship (FA9550-10-1-0169), and the Welch Foundation (F-1654).

Received: 19 September 2013 Accepted: 19 November 2013

Published: 13 January 2014

## References

1. Pyle AM, Lambowitz AM: **Group II introns: ribozymes that splice RNA and invade DNA.** In *The RNA World, Third Edition*. Edited by Gesteland RF, Cech T, Atkins JF. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2006:469–505.
2. Toro N, Jiménez-Zurdo JI, García-Rodríguez FM: **Bacterial group II introns: not just splicing.** *FEMS Microbiol Rev* 2007, **31**:342–358.
3. Lambowitz AM, Zimmerly S: **Group II introns: mobile ribozymes that invade DNA.** *Cold Spring Harb Perspect Biol* 2011, **3**:a003616.
4. Lambowitz AM, Zimmerly S: **Mobile group II introns.** *Annu Rev Genet* 2004, **38**:1–35.
5. Guo H, Karberg M, Long M, Jones JP 3rd, Sullenger B, Lambowitz AM: **Group II introns designed to insert into therapeutically relevant DNA target sites in human cells.** *Science* 2000, **289**:452–457.
6. Karberg M, Guo H, Zhong J, Coon R, Perutka J, Lambowitz AM: **Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria.** *Nat Biotechnol* 2001, **19**:1162–1167.
7. Perutka J, Wang W, Goerlitz D, Lambowitz AM: **Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes.** *J Mol Biol* 2004, **336**:421–439.
8. Cousineau B, Smith D, Lawrence-Cavanagh S, Mueller JE, Yang J, Mills D, Manias D, Dunny G, Lambowitz AM, Belfort M: **Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination.** *Cell* 1998, **94**:451–462.
9. Conlan LH, Stanger MJ, Ichihyanagi K, Belfort M: **Localization, mobility and fidelity of retrotransposed group II introns in rRNA genes.** *Nucleic Acids Res* 2005, **33**:5262–5270.
10. Mohr S, Ghanem E, Smith W, Sheeter D, Qin Y, King O, Polioudakis D, Iyer VR, Hunnicke-Smith S, Swamy S, Kuersten S, Lambowitz AM: **Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing.** *RNA* 2013, **19**:958–970.

11. Vellore J, Moretz SE, Lampson BC: A group II intron-type open reading frame from the thermophile *Bacillus (Geobacillus) stearotherophilus* encodes a heat-stable reverse transcriptase. *Appl Environ Microbiol* 2004, **70**:7140–7147.
12. Ng B, Nayak S, Gibbs MD, Lee J, Bergquist PL: Reverse transcriptases: intron-encoded proteins found in thermophilic bacteria. *Gene* 2007, **393**:137–144.
13. Collins K, Nilsen TW: Enzyme engineering through evolution: thermostable recombinant group II intron reverse transcriptases provide new tools for RNA research and biotechnology. *RNA* 2013, **19**:1017–1018.
14. Martin W, Koonin EV: Introns and the origin of nucleus-cytosol compartmentalization. *Nature* 2006, **440**:41–45.
15. Keating KS, Toor N, Perlman PS, Pyle AM: A structural analysis of the group II intron active site and implications for the spliceosome. *RNA* 2010, **16**:1–9.
16. Candales MA, Duong A, Hood KS, Li T, Neufeld RA, Sun R, McNeil BA, Wu L, Jarding AM, Zimmerly S: Database for bacterial group II introns. *Nucleic Acids Res* 2012, **40**:D187–D190.
17. Michel F, Ferat JL: Structure and activities of group II introns. *Annu Rev Biochem* 1995, **64**:435–461.
18. Qin PZ, Pyle AM: The architectural organization and mechanistic function of group II intron structural elements. *Curr Opin Struct Biol* 1998, **8**:301–308.
19. Toor N, Keating KS, Taylor SD, Pyle AM: Crystal structure of a self-spliced group II intron. *Science* 2008, **320**:77–82.
20. Toor N, Keating KS, Fedorova O, Rajashankar K, Wang J, Pyle AM: Tertiary architecture of the *Oceanobacillus iheyensis* group II intron. *RNA* 2010, **16**:57–69.
21. Marcia M, Somarowthu S, Pyle AM: Now on display: a gallery of group II intron structures at different stages of catalysis. *Mob DNA* 2013, **4**:14.
22. Muñoz-Adelantado E, San Filippo J, Martínez-Abarca F, García-Rodríguez FM, Lambowitz AM, Toro N: Mobility of the *Sinorhizobium meliloti* group II intron Rmlnt1 occurs by reverse splicing into DNA, but requires an unknown reverse transcriptase priming mechanism. *J Mol Biol* 2003, **327**:931–943.
23. Blocker FJ, Mohr G, Conlan LH, Qi L, Belfort M, Lambowitz AM: Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. *RNA* 2005, **11**:14–28.
24. Zimmerly S, Guo H, Perlman PS, Lambowitz AM: Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* 1995, **82**:545–554.
25. Zimmerly S, Guo H, Eskes R, Yang J, Perlman PS, Lambowitz AM: A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell* 1995, **83**:529–538.
26. Yang J, Zimmerly S, Perlman PS, Lambowitz AM: Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature* 1996, **381**:332–335.
27. Zhong J, Lambowitz AM: Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription. *EMBO J* 2003, **22**:4555–4565.
28. Martínez-Abarca F, Barrientos-Durán A, Fernández-López M, Toro N: The Rmlnt1 group II intron has two different retrohoming pathways for mobility using predominantly the nascent lagging strand at DNA replication forks for priming. *Nucleic Acids Res* 2004, **32**:2880–2888.
29. Toro N, Martínez-Abarca F: Comprehensive phylogenetic analysis of bacterial group II intron-encoded ORFs lacking the DNA endonuclease domain reveals new varieties. *PLoS One* 2013, **8**:e55102.
30. Yao J, Truong DM, Lambowitz AM: Genetic and biochemical assays reveal a key role for replication restart proteins in group II intron retrohoming. *PLoS Genet* 2013, **9**:e1003469.
31. Smith D, Zhong J, Matsuura M, Lambowitz AM, Belfort M: Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming. *Genes Dev* 2005, **19**:2477–2487.
32. Podar M, Chu VT, Pyle AM, Perlman PS: Group II intron splicing *in vivo* by first-step hydrolysis. *Nature* 1998, **391**:915–918.
33. Vogel J, Börner T: Lariat formation and a hydrolytic pathway in plant chloroplast group II intron splicing. *EMBO J* 2002, **21**:3794–3803.
34. Mastroianni M, Watanabe K, White TB, Zhuang F, Vernon J, Matsuura M, Wallingford J, Lambowitz AM: Group II intron-based gene targeting reactions in eukaryotes. *PLoS One* 2008, **3**:e3121.
35. Gaj T, Gersbach CA, Barbas CF 3rd: ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* 2013, **31**:397–405.
36. Zhuang F, Mastroianni M, White TB, Lambowitz AM: Linear group II intron RNAs can retrohome in eukaryotes and may use nonhomologous end-joining for cDNA ligation. *Proc Natl Acad Sci USA* 2009, **106**:18189–18194.
37. White TB, Lambowitz AM: The retrohoming of linear group II intron RNAs in *Drosophila melanogaster* occurs by both DNA ligase 4-dependent and -independent mechanisms. *PLoS Genet* 2012, **8**:e1002534.
38. Guo H, Zimmerly S, Perlman PS, Lambowitz AM: Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. *EMBO J* 1997, **16**:6835–6848.
39. Singh NN, Lambowitz AM: Interaction of a group II intron ribonucleoprotein endonuclease with its DNA target site investigated by DNA footprinting and modification interference. *J Mol Biol* 2001, **309**:361–386.
40. Noah JW, Park S, Whitt JT, Perutka J, Frey W, Lambowitz AM: Atomic force microscopy reveals DNA bending during group II intron ribonucleoprotein particle integration into double-stranded DNA. *Biochemistry* 2006, **45**:12424–12435.
41. Costa M, Michel F, Westhof E: A three-dimensional perspective on exon binding by a group II self-splicing intron. *EMBO J* 2000, **19**:5007–5018.
42. Zhuang F, Karberg M, Perutka J, Lambowitz AM: Ecl5, a group IIB intron with high retrohoming frequency: DNA target site recognition and use in gene targeting. *RNA* 2009, **15**:432–449.
43. Jiménez-Zurdo JI, García-Rodríguez FM, Barrientos-Durán A, Toro N: DNA target site requirements for homing *in vivo* of a bacterial group II intron encoding a protein lacking the DNA endonuclease domain. *J Mol Biol* 2003, **326**:413–423.
44. Toor N, Robart AR, Christianson J, Zimmerly S: Self-splicing of a group IIC intron: 5' exon recognition and alternative 5' splicing events implicate the stem-loop motif of a transcriptional terminator. *Nucleic Acids Res* 2006, **34**:6461–6471.
45. Robart AR, Seo W, Zimmerly S: Insertion of group II intron retroelements after intrinsic transcriptional terminators. *Proc Natl Acad Sci USA* 2007, **104**:6620–6625.
46. Léon G, Roy PH: Group IIC intron mobility into attC sites involves a bulged DNA stem-loop motif. *RNA* 2009, **15**:1543–1553.
47. Coros CJ, Landthaler M, Piazza CL, Beaugard A, Esposito D, Perutka J, Lambowitz AM, Belfort M: Retrotransposition strategies of the *Lactococcus lactis* LI.LtrB group II intron are dictated by host identity and cellular environment. *Mol Microbiol* 2005, **56**:509–524.
48. Xiang Q, Qin PZ, Michels WJ, Freeland K, Pyle AM: Sequence specificity of a group II intron ribozyme: multiple mechanisms for promoting unusually high discrimination against mismatched targets. *Biochemistry* 1998, **37**:3839–3849.
49. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E: A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012, **337**:816–821.
50. Cho SW, Kim S, Kim JM, Kim JS: Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* 2013, **31**:230–232.
51. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F: Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013, **339**:819–823.
52. Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK: Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* 2013, **31**:227–229.
53. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA: RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 2013, **31**:233–239.
54. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM: RNA-guided human genome engineering via Cas9. *Science* 2013, **339**:823–826.
55. Hou Z, Zhang Y, Propson NE, Howden SE, Chu LF, Sontheimer EJ, Thomson JA: Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc Natl Acad Sci USA* 2013, **110**:15644–15649.
56. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD: High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 2013, **31**:822–826.
57. García-Rodríguez FM, Barrientos-Durán A, Díaz-Prado V, Fernández-López M, Toro N: Use of Rmlnt1, a group IIB intron lacking the intron-encoded

- protein endonuclease domain, in gene targeting. *Appl Environ Microbiol* 2011, **77**:854–861.
58. Mohr G, Smith D, Belfort M, Lambowitz AM: Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. *Genes Dev* 2000, **14**:559–573.
59. Zhong J, Karberg M, Lambowitz AM: Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucleic Acids Res* 2003, **31**:1656–1664.
60. Enyeart PJ, Chirieleison SM, Dao MN, Perutka J, Quandt EM, Yao J, Whitt JT, Keatinge-Clay AT, Lambowitz AM, Ellington AD: Generalized bacterial genome editing using mobile group II introns and Cre-lox. *Mol Syst Biol* 2013, **9**:685.
61. Frazier CL, San Filippo J, Lambowitz AM, Mills DA: Genetic manipulation of *Lactococcus lactis* by using targeted group II introns: generation of stable insertions without selection. *Appl Environ Microbiol* 2003, **69**:1121–1128.
62. Yao J, Zhong J, Fang Y, Geisinger E, Novick RP, Lambowitz AM: Use of targetrons to disrupt essential and nonessential genes in *Staphylococcus aureus* reveals temperature sensitivity of LI.LtrB group II intron splicing. *RNA* 2006, **12**:1271–1281.
63. Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP: The Clostron: a universal gene knock-out system for the genus *Clostridium*. *J Microbiol Methods* 2007, **70**:452–464.
64. Yao J, Lambowitz AM: Gene targeting in Gram-negative bacteria by use of a mobile group II intron (“targetron”) expressed from a broad-host-range vector. *Appl Environ Microbiol* 2007, **73**:2735–2743.
65. Ichiyanagi K, Beaugregard A, Lawrence S, Smith D, Cousineau B, Belfort M: Retrotransposition of the LI.LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. *Mol Microbiol* 2002, **46**:1259–1272.
66. Yao J, Zhong J, Lambowitz AM: Gene targeting using randomly inserted group II introns (targetrons) recovered from an *Escherichia coli* gene disruption library. *Nucleic Acids Res* 2005, **33**:3351–3362.
67. Malhotra M, Srivastava S: An *ipdC* gene knock-out of *Azospirillum brasilense* strain SM and its implications on indole-3-acetic acid biosynthesis and plant growth promotion. *Antonie Van Leeuwenhoek* 2008, **93**:425–433.
68. Akhtar P, Khan SA: Two independent replicons can support replication of the anthrax toxin-encoding plasmid pXO1 of *Bacillus anthracis*. *Plasmid* 2012, **67**:111–117.
69. Chen Y, McClane BA, Fisher DJ, Rood JI, Gupta P: Construction of an alpha toxin gene knockout mutant of *Clostridium perfringens* type A by use of a mobile group II intron. *Appl Environ Microbiol* 2005, **71**:7542–7547.
70. Cheng C, Nair AD, Indukuri WV, Gong S, Felsheim RF, Jaworski D, Munderloh UG, Ganta RR: Targeted and random mutagenesis of *Ehrlichia chaffeensis* for the identification of genes required for *in vivo* infection. *PLoS Pathog* 2013, **9**:e1003171.
71. Rodriguez SA, Yu JJ, Davis G, Arulanandam BP, Klose KE: Targeted inactivation of *Francisella tularensis* genes by group II introns. *Appl Environ Microbiol* 2008, **74**:2619–2626.
72. Alonzo F 3rd, Port GC, Cao M, Freitag NE: The posttranslocation chaperone PrsA2 contributes to multiple facets of *Listeria monocytogenes* pathogenesis. *Infect Immun* 2009, **77**:2612–2623.
73. Zarschler K, Janesch B, Zayni S, Schäffer C, Messner P: Construction of a gene knockout system for application in *Paenibacillus alvei* CCM 2051 T, exemplified by the S-layer glycan biosynthesis initiation enzyme WsfP. *Appl Environ Microbiol* 2009, **75**:3077–3085.
74. Steen JA, Steen JA, Harrison P, Seemann T, Wilkie I, Harper M, Adler B, Boyce JD: Fis is essential for capsule production in *Pasteurella multocida* and regulates expression of other important virulence factors. *PLoS Pathog* 2010, **6**:e1000750.
75. Pearson MM, Mobley HL: The type III secretion system of *Proteus mirabilis* HI4320 does not contribute to virulence in the mouse model of ascending urinary tract infection. *J Med Microbiol* 2007, **56**:1277–1283.
76. Park JM, Jang YS, Kim TY, Lee SY: Development of a gene knockout system for *Ralstonia eutropha* H16 based on the broad-host-range vector expressing a mobile group II intron. *FEMS Microbiol Lett* 2010, **309**:193–200.
77. Smith CL, Weiss BL, Aksoy S, Runyen-Janecky LJ: Characterization of the achromobactin iron acquisition operon in *Sodalis glossinidius*. *Appl Environ Microbiol* 2013, **79**:2872–2881.
78. Kumar S, Smith KP, Floyd JL, Varela MF: Cloning and molecular analysis of a mannitol operon of phosphoenolpyruvate-dependent phosphotransferase (PTS) type from *Vibrio cholerae* O395. *Arch Microbiol* 2011, **193**:201–208.
79. Palonen E, Lindstrom M, Karttunen R, Somervuo P, Korkeala H: Expression of signal transduction system encoding genes of *Yersinia pseudotuberculosis* IP32953 at 28°C and 3°C. *PLoS One* 2011, **6**:e25063.
80. Heap JT, Kuehne SA, Ehsaan M, Cartman ST, Cooksley CM, Scott JC, Minton NP: The Clostron: mutagenesis in *Clostridium* refined and streamlined. *J Microbiol Methods* 2010, **80**:49–55.
81. Corvaglia AR, Francois P, Hernandez D, Perron K, Linder P, Schrenzel J: A type III-like restriction endonuclease functions as a major barrier to horizontal gene transfer in clinical *Staphylococcus aureus* strains. *Proc Natl Acad Sci USA* 2010, **107**:11954–11958.
82. Eskes R, Yang J, Lambowitz AM, Perlman PS: Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. *Cell* 1997, **88**:865–874.
83. Eskes R, Liu L, Ma H, Chao MY, Dickson L, Lambowitz AM, Perlman PS: Multiple homing pathways used by yeast mitochondrial group II introns. *Mol Cell Biol* 2000, **20**:8432–8446.
84. Rest JS, Mindell DP: Retroids in archaea: phylogeny and lateral origins. *Mol Biol Evol* 2003, **20**:1134–1142.
85. Sayeed S, Uzal FA, Fisher DJ, Saputo J, Vidal JE, Chen Y, Gupta P, Rood JI, McClane BA: Beta toxin is essential for the intestinal virulence of *Clostridium perfringens* type C disease isolate CN3685 in a rabbit ileal loop model. *Mol Microbiol* 2008, **67**:15–30.
86. Buchan BW, McCaffrey RL, Lindemann SR, Allen LA, Jones BD: Identification of *migR*, a regulatory element of the *Francisella tularensis* live vaccine strain *igIABCD* virulence operon required for normal replication and trafficking in macrophages. *Infect Immun* 2009, **77**:2517–2529.
87. Carter GP, Awad MM, Hao Y, Thelen T, Bergin IL, Howarth PM, Seemann T, Rood JI, Aronoff DM, Lyras D: Tcsl is an essential virulence factor in *Clostridium sordellii* ATCC 9714. *Infect Immun* 2011, **79**:1025–1032.
88. Francis MB, Allen CA, Shrestha R, Sorg JA: Bile acid recognition by the *Clostridium difficile* germinant receptor, CspC, is important for establishing infection. *PLoS Pathog* 2013, **9**:e1003356.
89. Zoraghi R, See RH, Gong H, Lian T, Swayze R, Finlay BB, Brunham RC, McMaster WR, Reiner NE: Functional analysis, overexpression, and kinetic characterization of pyruvate kinase from methicillin-resistant *Staphylococcus aureus*. *Biochemistry* 2010, **49**:7733–7747.
90. Zoraghi R, Worrall L, See RH, Strangman W, Popplewell WL, Gong H, Samaai T, Swayze RD, Kaur S, Vuckovic M, Finlay BB, Brunham RC, McMaster WR, Davies-Coleman MT, Strynadka NC, Andersen RJ, Reiner NE: Methicillin-resistant *Staphylococcus aureus* (MRSA) pyruvate kinase as a target for bis-indole alkaloids with antibacterial activities. *J Biol Chem* 2011, **286**:44716–44725.
91. Rawsthorne H, Turner KN, Mills DA: Multicopy integration of heterologous genes, using the lactococcal group II intron targeted to bacterial insertion sequences. *Appl Environ Microbiol* 2006, **72**:6088–6093.
92. Shao L, Hu S, Yang Y, Gu Y, Chen J, Yang Y, Jiang W, Yang S: Targeted gene disruption by use of a group II intron (targetron) vector in *Clostridium acetobutylicum*. *Cell Res* 2007, **17**:963–965.
93. Jiang Y, Xu C, Dong F, Yang Y, Jiang W, Yang S: Disruption of the acetoacetate decarboxylase gene in solvent-producing *Clostridium acetobutylicum* increases the butanol ratio. *Metab Eng* 2009, **11**:284–291.
94. Tolonen AC, Chilaka AC, Church GM: Targeted gene inactivation in *Clostridium phytofermentans* shows that cellulose degradation requires the family 9 hydrolase Cphy3367. *Mol Microbiol* 2009, **74**:1300–1313.
95. Cai G, Jin B, Saint C, Monis P: Genetic manipulation of butyrate formation pathways in *Clostridium butyricum*. *J Biotechnol* 2011, **155**:269–274.
96. Lehmann D, Lütke-Eversloh T: Switching *Clostridium acetobutylicum* to an ethanol producer by disruption of the butyrate/butanol fermentative pathway. *Metab Eng* 2011, **13**:464–473.
97. Cooksley CM, Zhang Y, Wang H, Redl S, Winzer K, Minton NP: Targeted mutagenesis of the *Clostridium acetobutylicum* acetone-butanol-ethanol fermentation pathway. *Metab Eng* 2012, **14**:630–641.
98. Jang YS, Lee JY, Lee J, Park JH, Im JA, Eom MH, Lee SH, Song H, Cho JH, Seung Do Y, Lee SY: Enhanced butanol production obtained by reinforcing the direct butanol-forming route in *Clostridium acetobutylicum*. *mBio* 2012, **3**:12.
99. Jia K, Zhang Y, Li Y: Identification and characterization of two functionally unknown genes involved in butanol production of *Clostridium acetobutylicum*. *PLoS One* 2012, **7**:e38815.



100. Kuit W, Minton NP, López-Contreras AM, Eggink G: **Disruption of the acetate kinase (*ack*) gene of *Clostridium acetobutylicum* results in delayed acetate production.** *Appl Microbiol Biotechnol* 2012, **94**:729–741.
101. Lehmann D, Hönicke D, Ehrenreich A, Schmidt M, Weuster-Botz D, Bahl H, Lütke-Eversloh T: **Modifying the product pattern of *Clostridium acetobutylicum*: physiological effects of disrupting the acetate and acetone formation pathways.** *Appl Microbiol Biotechnol* 2012, **94**:743–754.
102. Li Y, Tschapinski TJ, Engle NL, Hamilton CY, Rodriguez M Jr, Liao JC, Schadt CW, Guss AM, Yang Y, Graham DE: **Combined inactivation of the *Clostridium cellulolyticum* lactate and malate dehydrogenase genes substantially increases ethanol yield from cellulose and switchgrass fermentations.** *Biotechnol Biofuels* 2012, **5**:2.
103. Steiner E, Scott J, Minton NP, Winzer K: **An *agr* quorum sensing system that regulates granule formation and sporulation in *Clostridium acetobutylicum*.** *Appl Environ Microbiol* 2012, **78**:1113–1122.
104. Wietzke M, Bahl H: **The redox-sensing protein Rex, a transcriptional regulator of solventogenesis in *Clostridium acetobutylicum*.** *Appl Microbiol Biotechnol* 2012, **96**:749–761.
105. Cai G, Jin B, Monis P, Saint C: **A genetic and metabolic approach to redirection of biochemical pathways of *Clostridium butyricum* for enhancing hydrogen production.** *Biotechnol Bioeng* 2013, **110**:338–342.
106. Celik H, Blouzard JC, Voigt B, Becher D, Trotter V, Fierobe HP, Tardif C, Pagès S, de Philip P: **A two-component system (*XydsR*) controls the expression of genes encoding CBM6-containing proteins in response to straw in *Clostridium cellulolyticum*.** *PLoS One* 2013, **8**:e56063.
107. Fendri I, Abdou L, Trotter V, Dedieu L, Maamar H, Minton NP, Tardif C: **Regulation of *cel* genes of *C. cellulolyticum*: identification of GlyR2, a transcriptional regulator regulating *cel5D* gene expression.** *PLoS One* 2013, **8**:e44708.
108. Ferdinand PH, Borne R, Trotter V, Pagès S, Tardif C, Fierobe HP, Perret S: **Are cellulosome scaffolding protein CipC and CBM3-containing protein HycP, involved in adherence of *Clostridium cellulolyticum* to cellulose?** *PLoS One* 2013, **8**:e69360.
109. Jang YS, Woo HM, Im JA, Kim IH, Lee SY: **Metabolic engineering of *Clostridium acetobutylicum* for enhanced production of butyric acid.** *Appl Microbiol Biotechnol* 2013, **97**:9355–9363.
110. Wang Y, Li X, Milne CB, Janssen H, Lin W, Phan G, Hu H, Jin Y-S, Price ND, Blaschek HP: **Development of a gene knockout system using mobile group II introns (targetron) and genetic disruption of acid production pathways in *Clostridium beijerinckii*.** *Appl Environ Microbiol* 2013, **79**:5853–5863.
111. Emerson JE, Reynolds CB, Fagan RP, Shaw HA, Goulding D, Fairweather NF: **A novel genetic switch controls phase variable expression of CwpV, a *Clostridium difficile* cell wall protein.** *Mol Microbiol* 2009, **74**:541–556.
112. Kirby JM, Ahern H, Roberts AK, Kumar V, Freeman Z, Acharya KR, Shone CC: **Cwp84, a surface-associated cysteine protease, plays a role in the maturation of the surface layer of *Clostridium difficile*.** *J Biol Chem* 2009, **284**:34666–34673.
113. Underwood S, Guan S, Vijayasubhash V, Baines SD, Graham L, Lewis RJ, Wilcox MH, Stephenson K: **Characterization of the sporulation initiation pathway of *Clostridium difficile* and its role in toxin production.** *J Bacteriol* 2009, **191**:7296–7305.
114. Burns DA, Heap JT, Minton NP: **SleC is essential for germination of *Clostridium difficile* spores in nutrient-rich medium supplemented with the bile salt taurocholate.** *J Bacteriol* 2010, **192**:657–664.
115. Antunes A, Martin-Verstraete I, Dupuy B: **CcpA-mediated repression of *Clostridium difficile* toxin gene expression.** *Mol Microbiol* 2011, **79**:882–899.
116. Barketi-Klai A, Hoys S, Lambert-Bordes S, Collignon A, Kansau I: **Role of fibronectin-binding protein A in *Clostridium difficile* intestinal colonization.** *J Med Microbiol* 2011, **60**:1155–1161.
117. Dawson LF, Donahue EH, Cartman ST, Barton RH, Bundy J, McNERNEY R, Minton NP, Wren BW: **The analysis of *para*-cresol production and tolerance in *Clostridium difficile* 027 and 012 strains.** *BMC Microbiol* 2011, **11**:86.
118. Dingle TC, Mulvey GL, Armstrong GD: **Mutagenic analysis of the *Clostridium difficile* flagellar proteins, FliC and FliD, and their contribution to virulence in hamsters.** *Infect Immun* 2011, **79**:4061–4067.
119. Ho TD, Ellermeier CD: **PrsW is required for colonization, resistance to antimicrobial peptides, and expression of extracytoplasmic function sigma factors in *Clostridium difficile*.** *Infect Immun* 2011, **79**:3229–3238.
120. Kuehne SA, Cartman ST, Minton NP: **Both, toxin A and toxin B, are important in *Clostridium difficile* infection.** *Gut Microbes* 2011, **2**:252–255.
121. McBride SM, Sonenshein AL: **The *dlt* operon confers resistance to cationic antimicrobial peptides in *Clostridium difficile*.** *Microbiology* 2011, **157**:1457–1465.
122. de la Riva L, Willing SE, Tate EW, Fairweather NF: **Roles of cysteine proteases Cwp84 and Cwp13 in biogenesis of the cell wall of *Clostridium difficile*.** *J Bacteriol* 2011, **193**:3276–3285.
123. Saujet L, Monot M, Dupuy B, Soutourina O, Martin-Verstraete I: **The key sigma factor of transition phase, SigH, controls sporulation, metabolism, and virulence factor expression in *Clostridium difficile*.** *J Bacteriol* 2011, **193**:3186–3196.
124. Bakker D, Smits WK, Kuijper EJ, Corver J: **TcdC does not significantly repress toxin expression in *Clostridium difficile* 630ΔErm.** *PLoS One* 2012, **7**:e43247.
125. Dawson LF, Valiente E, Faulds-Pain A, Donahue EH, Wren BW: **Characterisation of *Clostridium difficile* biofilm formation, a role for Spo0A.** *PLoS One* 2012, **7**:e50527.
126. Deakin LJ, Clare S, Fagan RP, Dawson LF, Pickard DJ, West MR, Wren BW, Fairweather NF, Dougan G, Lawley TD: **The *Clostridium difficile* *spo0A* gene is a persistence and transmission factor.** *Infect Immun* 2012, **80**:2704–2711.
127. Govind R, Dupuy B: **Secretion of *Clostridium difficile* toxins A and B requires the holin-like protein TcdE.** *PLoS Pathog* 2012, **8**:e1002727.
128. Olling A, Seehase S, Minton NP, Tatge H, Schröter S, Kohlscheen S, Pich A, Just I, Gerhard R: **Release of TcdA and TcdB from *Clostridium difficile* cdi 630 is not affected by functional inactivation of the *tcdE* gene.** *Microb Pathog* 2012, **52**:92–100.
129. Adams CM, Eckenroth BE, Putnam EE, Doublé S, Shen A: **Structural and functional analysis of the CspB protease required for *Clostridium* spore germination.** *PLoS Pathog* 2013, **9**:e1003165.
130. Bouillaut L, Self WT, Sonenshein AL: **Proline-dependent regulation of *Clostridium difficile* Stickland metabolism.** *J Bacteriol* 2013, **195**:844–854.
131. Dapa T, Leuzzi R, Ng YK, Baban ST, Adamo R, Kuehne SA, Scarselli M, Minton NP, Serruto D, Unnikrishnan M: **Multiple factors modulate biofilm formation by the anaerobic pathogen *Clostridium difficile*.** *J Bacteriol* 2013, **195**:545–555.
132. Martin MJ, Clare S, Goulding D, Faulds-Pain A, Barquist L, Browne HP, Pettit L, Dougan G, Lawley TD, Wren BW: **The *agr* locus regulates virulence and colonization genes in *Clostridium difficile* 027.** *J Bacteriol* 2013, **195**:3672–3681.
133. Putnam EE, Nock AM, Lawley TD, Shen A: **SpoIVA and Sipl are *Clostridium difficile* spore morphogenetic proteins.** *J Bacteriol* 2013, **195**:1214–1225.
134. Permpoonpattana P, Phetcharaburanin J, Mikelsone A, Dembek M, Tan S, Brisson MC, La Ragione R, Brisson AR, Fairweather N, Hong HA, Cutting SM: **Functional characterization of *Clostridium difficile* spore coat proteins.** *J Bacteriol* 2013, **195**:1492–1503.
135. Ammam F, Meziane-cherif D, Mengin-Lecreux D, Blanot D, Patin D, Boneca IG, Courvalin P, Lambert T, Candela T: **The functional *vanG<sub>cd</sub>* cluster of *Clostridium difficile* does not confer vancomycin resistance.** *Mol Microbiol* 2013, **89**:612–625.
136. Baban ST, Kuehne SA, Barketi-Klai A, Cartman ST, Kelly ML, Hardie KR, Kansau I, Collignon A, Minton NP: **The role of flagella in *Clostridium difficile* pathogenesis: comparison between a non-epidemic and an epidemic strain.** *PLoS One* 2013, **8**:e73026.
137. Ng KM, Ferreyra JA, Higginbottom SK, Lynch JB, Kashyap PC, Gopinath S, Naidu N, Choudhury B, Weimer BC, Monack DM, Sonnenburg JL: **Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens.** *Nature* 2013, **502**:96–99.
138. Tulli L, Marchi S, Petracca R, Shaw HA, Fairweather NF, Scarselli M, Soriani M, Leuzzi R: **CbpA: a novel surface exposed adhesin of *Clostridium difficile* targeting human collagen.** *Cell Microbiol* 2013, **15**:1674–1687.
139. Saujet L, Pereira FC, Serrano M, Soutourina O, Monot M, Shelyakin PV, Gelfand MS, Dupuy B, Henriques AO, Martin-Verstraete I: **Genome-wide analysis of cell type-specific gene transcription during spore formation in *Clostridium difficile*.** *PLoS Genet* 2013, **9**:e1003756.
140. McKee RW, Mangalea MR, Purcell EB, Borchardt EK, Tamayo R: **The second messenger cyclic di-GMP regulates *Clostridium difficile* toxin production by controlling expression of *sigD*.** *J Bacteriol* 2013, **195**:5174–5185.
141. Pereira FC, Saujet L, Tome AR, Serrano M, Monot M, Couture-Tosi E, Martin-Verstraete I, Dupuy B, Henriques AO: **The spore differentiation pathway in the enteric pathogen *Clostridium difficile*.** *PLoS Genet* 2013, **9**:e1003782.

142. Janoir C, Denève C, Bouttier S, Barbut F, Hoys S, Caleechum L, Chapetón-Montes D, Pereira FC, Henriques AO, Collignon A, Monot M, Dupuy B: **Adaptive strategies and pathogenesis of *Clostridium difficile* from in vivo transcriptomics.** *Infect Immun* 2013, **81**:3757–3769.
143. Fimlaid KA, Bond JP, Schutz KC, Putnam EE, Leung JM, Lawley TD, Shen A: **Global analysis of the sporulation pathway of *Clostridium difficile*.** *PLoS Genet* 2013, **9**:e1003660.
144. Bradshaw M, Marshall KM, Heap JT, Tepp WH, Minton NP, Johnson EA: **Construction of a nontoxicogenic *Clostridium botulinum* strain for food challenge studies.** *Appl Environ Microbiol* 2010, **76**:387–393.
145. Cooksley CM, Davis IJ, Winzer K, Chan WC, Peck MW, Minton NP: **Regulation of neurotoxin production and sporulation by a putative *agrBD* signaling system in proteolytic *Clostridium botulinum*.** *Appl Environ Microbiol* 2010, **76**:4448–4460.
146. Marshall KM, Bradshaw M, Johnson EA: **Conjugative botulinum neurotoxin-encoding plasmids in *Clostridium botulinum*.** *PLoS One* 2010, **5**:e11087.
147. Selby K, Lindström M, Somervuo P, Heap JT, Minton NP, Korkeala H: **Important role of class I heat shock genes *hrcA* and *dnaK* in the heat shock response and the response to pH and NaCl stress of group I *Clostridium botulinum* strain ATCC 3502.** *Appl Environ Microbiol* 2011, **77**:2823–2830.
148. Söderholm H, Lindström M, Somervuo P, Heap J, Minton N, Linden J, Korkeala H: ***cspB* encodes a major cold shock protein in *Clostridium botulinum* ATCC 3502.** *Int J Food Microbiol* 2011, **146**:23–30.
149. Kirk DG, Dahlsten E, Zhang Z, Korkeala H, Lindström M: **Involvement of *Clostridium botulinum* ATCC 3502 sigma factor K in early-stage sporulation.** *Appl Environ Microbiol* 2012, **78**:4590–4596.
150. Lindström M, Dahlsten E, Söderholm H, Selby K, Somervuo P, Heap JT, Minton NP, Korkeala H: **Involvement of two-component system CBO0366/CBO0365 in the cold shock response and growth of group I (proteolytic) *Clostridium botulinum* ATCC 3502 at low temperatures.** *Appl Environ Microbiol* 2012, **78**:5466–5470.
151. Dahlsten E, Kirk D, Lindström M, Korkeala H: **Alternative sigma factor SigK has a role in stress tolerance of group I *Clostridium botulinum* strain ATCC 3502.** *Appl Environ Microbiol* 2013, **79**:3867–3869.
152. Zhang Z, Korkeala H, Dahlsten E, Sahala E, Heap JT, Minton NP, Lindström M: **Two-component signal transduction system CBO0787/CBO0786 represses transcription from botulinum neurotoxin promoters in *Clostridium botulinum* ATCC 3502.** *PLoS Pathog* 2013, **9**:e1003252.
153. Derman Y, Isokallio M, Lindström M, Korkeala H: **The two-component system CBO2306/CBO2307 is important for cold adaptation of *Clostridium botulinum* ATCC 3502.** *Int J Food Microbiol* 2013, **167**:87–91.
154. Paredes-Sabja D, Torres JA, Setlow P, Sarker MR: ***Clostridium perfringens* spore germination: characterization of germinants and their receptors.** *J Bacteriol* 2008, **190**:1190–1201.
155. Vidal JE, Ohtani K, Shimizu T, McClane BA: **Contact with enterocyte-like Caco-2 cells induces rapid upregulation of toxin production by *Clostridium perfringens* type C isolates.** *Cell Microbiol* 2009, **11**:1306–1328.
156. Cheung JK, Keyburn AL, Carter GP, Lanckriet AL, Van Immerseel F, Moore RJ, Rood JI: **The VirSR two-component signal transduction system regulates NetB toxin production in *Clostridium perfringens*.** *Infect Immun* 2010, **78**:3064–3072.
157. Li J, McClane BA: **Evaluating the involvement of alternative sigma factors SigF and SigG in *Clostridium perfringens* sporulation and enterotoxin synthesis.** *Infect Immun* 2010, **78**:4286–4293.
158. Chen J, Rood JI, McClane BA: **Epsilon-toxin production by *Clostridium perfringens* type D strain CN3718 is dependent upon the *agr* operon but not the VirS/VirR two-component regulatory system.** *mBio* 2011, **2**:e00275. 11.
159. Fujita M, Tsuchida A, Hirata A, Kobayashi N, Goto K, Osumi K, Hirose Y, Nakayama J, Yamanoi T, Ashida H, Mizuno M: **Glycoside hydrolase family 89  $\alpha$ -N-acetylglucosaminidase from *Clostridium perfringens* specifically acts on GlcNAc<sub>4</sub>Gal $\beta$ 1R at the non-reducing terminus of O-glycans in gastric mucin.** *J Biol Chem* 2011, **286**:6479–6489.
160. Li J, Chen J, Vidal JE, McClane BA: **The Agr-like quorum-sensing system regulates sporulation and production of enterotoxin and beta2 toxin by *Clostridium perfringens* type A non-food-borne human gastrointestinal disease strain F5603.** *Infect Immun* 2011, **79**:2451–2459.
161. Li J, Sayeed S, Robertson S, Chen J, McClane BA: **Sialidases affect the host cell adherence and epsilon toxin-induced cytotoxicity of *Clostridium perfringens* type D strain CN3718.** *PLoS Pathog* 2011, **7**:e1002429.
162. Chen J, McClane BA: **Role of the Agr-like quorum-sensing system in regulating toxin production by *Clostridium perfringens* type B strains CN1793 and CN1795.** *Infect Immun* 2012, **80**:3008–3017.
163. Banawas S, Paredes-Sabja D, Korza G, Li Y, Hao B, Setlow P, Sarker MR: **The *Clostridium perfringens* germinant receptor protein GerKC is located in the spore inner membrane and is crucial for spore germination.** *J Bacteriol* 2013, **195**:5084–5091.
164. Li J, Ma M, Sarker MR, McClane BA: **CodY is a global regulator of virulence-associated properties for *Clostridium perfringens* type D strain CN3718.** *mBio* 2013, **4**:e00770–13.
165. Sirigi Reddy AR, Girinathan BP, Zapotocny R, Govind R: **Identification and characterization of *Clostridium sordellii* toxin gene regulator.** *J Bacteriol* 2013, **195**:4246–4254.
166. Dong H, Zhang Y, Dai Z, Li Y: **Engineering *Clostridium* strain to accept unmethylated DNA.** *PLoS One* 2010, **5**:e9038.
167. Cui GZ, Hong W, Zhang J, Li WL, Feng Y, Liu YJ, Cui Q: **Targeted gene engineering in *Clostridium cellulolyticum* H10 without methylation.** *J Microbiol Methods* 2012, **89**:201–208.
168. Saldanha RJ, Pemberton A, Shiflett P, Perutka J, Whitt JT, Ellington A, Lambowitz AM, Kramer R, Taylor D, Lamkin TJ: **Rapid targeted gene disruption in *Bacillus anthracis*.** *BMC Biotechnol* 2013, **13**:72.
169. Chen Y, Caruso L, McClane B, Fisher D, Gupta P: **Disruption of a toxin gene by introduction of a foreign gene into the chromosome of *Clostridium perfringens* using targetron-induced mutagenesis.** *Plasmid* 2007, **58**:182–189.
170. Plante I, Cousineau B: **Restriction for gene insertion within the *Lactococcus lactis* LI.LtrB group II intron.** *RNA* 2006, **12**:1980–1992.
171. Jia K, Zhu Y, Zhang Y, Li Y: **Group II intron-anchored gene deletion in *Clostridium*.** *PLoS One* 2011, **6**:e16693.
172. Carr PA, Church GM: **Genome engineering.** *Nat Biotechnol* 2009, **27**:1151–1162.
173. King RD, Rowland J, Oliver SG, Young M, Aubrey W, Byrne E, Liakata M, Markham M, Pir P, Soldatova LN, Sparks A, Whelan KE, Clare A: **The automation of science.** *Science* 2009, **324**:85–89.
174. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM: **Programming cells by multiplex genome engineering and accelerated evolution.** *Nature* 2009, **460**:894–898.
175. Haki GD, Rakshit SK: **Developments in industrially important thermostable enzymes: a review.** *Bioresour Technol* 2003, **89**:17–34.
176. Turner P, Mamo G, Karlsson EN: **Potential and utilization of thermophiles and thermostable enzymes in biorefining.** *Microb Cell Fact* 2007, **6**:9.
177. Taylor MP, van Zyl L, Tuffin IM, Leak DJ, Cowan DA: **Genetic tool development underpins recent advances in thermophilic whole-cell biocatalysts.** *Microb Biotechnol* 2011, **4**:438–448.
178. Mohr G, Hong W, Zhang J, Cui GZ, Yang Y, Cui Q, Liu YJ, Lambowitz AM: **A targetron system for gene targeting in thermophiles and its application in *Clostridium thermocellum*.** *PLoS One* 2013, **8**:e69032.
179. Lynd LR, Grethlein HE, Wolkin RH: **Fermentation of cellulosic substrates in batch and continuous culture by *Clostridium thermocellum*.** *Appl Environ Microbiol* 1989, **55**:3131–3139.
180. Nagy A: **Cre recombinase: the universal reagent for genome tailoring.** *Genesis* 2000, **26**:99–109.
181. Claverys JP, Prudhomme M, Mortier-Barrière I, Martin B: **Adaptation to the environment: *Streptococcus pneumoniae*, a paradigm for recombination-mediated genetic plasticity?** *Mol Microbiol* 2000, **35**:251–259.
182. Metzgar D, Bacher JM, Pezo V, Reader J, Döring V, Schimmel P, Marlière P, de Crécy-Lagard V: ***Acinetobacter* sp. ADP1: an ideal model organism for genetic analysis and genome engineering.** *Nucleic Acids Res* 2004, **32**:5780–5790.
183. Ellis HM, Yu D, DiTizio T, Court DL: **High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides.** *Proc Natl Acad Sci USA* 2001, **98**:6742–6746.
184. Costantino N, Court DL: **Enhanced levels of  $\lambda$  Red-mediated recombinants in mismatch repair mutants.** *Proc Natl Acad Sci USA* 2003, **100**:15748–15753.
185. Lusk JE, Williams RJ, Kennedy EP: **Magnesium and the growth of *Escherichia coli*.** *J Biol Chem* 1968, **243**:2618–2624.
186. Horowitz SB, Tluczek LJ: **Gonadotropin stimulates oocyte translation by increasing magnesium activity through intracellular potassium-magnesium exchange.** *Proc Natl Acad Sci USA* 1989, **86**:9652–9656.
187. Günther T: **Concentration, compartmentation and metabolic function of intracellular free Mg<sup>2+</sup>.** *Magnes Res* 2006, **19**:225–236.



188. Jones JP 3rd, Kierlin MN, Coon RG, Perutka J, Lambowitz AM, Sullenger BA: **Retargeting mobile group II introns to repair mutant genes.** *Mol Ther* 2005, **11**:687–694.
189. White TB: *Group II intron retrohoming and gene targeting reactions in Drosophila melanogaster*, PhD thesis. University of Texas: Cell and Molecular Biology; 2011.
190. Truong DM, Sidote DJ, Russell R, Lambowitz AM: **Enhanced group II intron retrohoming in magnesium-deficient *Escherichia coli* via selection of mutations in the ribozyme core.** *Proc Natl Acad Sci USA* 2013, **110**:E3800–E3809.
191. Hu WS, Hughes SH: **HIV-1 reverse transcription.** *Cold Spring Harb Perspect Med* 2012, **2**:a006882.

doi:10.1186/1759-8753-5-2

**Cite this article as:** Enyeart *et al.*: Biotechnological applications of mobile group II introns and their reverse transcriptases: gene targeting, RNA-seq, and non-coding RNA analysis. *Mobile DNA* 2014 5:2.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

