



Generalized bacterial genome editing using mobile group II introns and Cre-*lox*

Peter J Enyeart¹, Steven M Chirieleison^{2,5}, Mai N Dao^{3,4}, Jiri Perutka^{1,3}, Erik M Quandt¹, Jun Yao^{1,3,4}, Jacob T Whitt^{1,3,4}, Adrian T Keatinge-Clay^{1,3}, Alan M Lambowitz^{1,3,4} and Andrew D Ellington^{1,3,*}

¹ Institute for Cell and Molecular Biology, University of Texas at Austin, Austin, TX, USA, ² Department of Biomedical Engineering, University of Texas at Austin, Austin, TX, USA, ³ Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, TX, USA and ⁴ Section of Molecular Genetics and Microbiology, School of Biological Sciences, University of Texas at Austin, Austin, TX, USA

⁵Present address: School of Medicine, Case Western Reserve University, Cleveland, OH 44106, USA.

* Corresponding author. Institute for Cell and Molecular Biology, University of Texas at Austin, Austin, TX 78712, USA. Tel.: + 1 512 232 3424; Fax: + 1 512 471 7014; E-mail: andy.ellington@mail.utexas.edu

Received 31.5.13; accepted 23.7.13

Efficient bacterial genetic engineering approaches with broad-host applicability are rare. We combine two systems, mobile group II introns ('targetrons') and Cre-*lox*, which function efficiently in many different organisms, into a versatile platform we call GETR (Genome Editing via Targetrons and Recombinases). The introns deliver *lox* sites to specific genomic loci, enabling genomic manipulations. Efficiency is enhanced by adding flexibility to the RNA hairpins formed by the *lox* sites. We use the system for insertions, deletions, inversions, and one-step cut-and-paste operations. We demonstrate insertion of a 12-kb polyketide synthase operon into the *lacZ* gene of *Escherichia coli*, multiple simultaneous and sequential deletions of up to 120 kb in *E. coli* and *Staphylococcus aureus*, inversions of up to 1.2 Mb in *E. coli* and *Bacillus subtilis*, and one-step cut-and-pastes for translocating 120 kb of genomic sequence to a site 1.5 Mb away. We also demonstrate the simultaneous delivery of *lox* sites into multiple loci in the *Shewanella oneidensis* genome. No selectable markers need to be placed in the genome, and the efficiency of Cre-mediated manipulations typically approaches 100%.

Molecular Systems Biology 9: 685; published online 3 September 2013; doi:10.1038/msb.2013.41

Subject Categories: synthetic biology; genome stability & dynamics

Keywords: bacterial genome engineering; Cre-*lox*; mobile group II introns; *Staphylococcus aureus*; *Shewanella oneidensis*

Introduction

Though synthetic biology has thus far been focused primarily on building circuits of small numbers of genes to perform tasks of interest (Kaern *et al.*, 2003; Lu *et al.*, 2009), in recent years, more interest is being taken in the genome as a whole as the unit of engineering (Gibson *et al.*, 2010; Dymond *et al.*, 2011; Isaacs *et al.*, 2011). As interest in engineering bacterial genomes increases, the need for efficient tools for manipulating these genomes will also increase. Though a variety of methods exist for engineering bacterial genomes (Miller, 1991; Hughes and Maloy, 2007), each has specific limitations in terms of site specificity, efficiency, versatility, and/or range of applicable bacterial species. Recombineering and related methods making use of phage recombinases have come into widespread use for small-scale modifications in *Escherichia coli* (Datsenko and Wanner, 2000; Yu *et al.*, 2000; Costantino and Court, 2003; Wang *et al.*, 2009), but use of this approach in other species has so far been limited and often requires developing new recombineering functions for each system (Datta *et al.*, 2008; van Kessel and Hatfull, 2008; Swingle *et al.*, 2010). On the other

hand, site-specific recombinases such as the Cre-*lox* system are quite efficient and function in many organisms; indeed, the Cre-*lox* system has been claimed to function efficiently 'in any cellular environment and on any kind of DNA' (Nagy, 2000). In bacteria, the system has thus far been primarily used for selective marker removal, but it has, for example, been used to create large deletions in *E. coli* (Fukuya *et al.*, 2004) and large inversions in *Lactococcus lactis* (Campo *et al.*, 2004). However, positioning the recombination-recognition targets requires complementary genome-engineering approaches (typically with selectable markers), thus creating a chicken-and-egg problem.

Retargetable mobile group II introns are another tool that has been developed relatively recently. These so-called 'targetrons' can be designed to insert into a given DNA site at efficiencies high enough that selectable markers need not be used. Mobile group II introns occur naturally in bacteria, eukaryotic organelles, and some archaea, and are thought to be precursors to the eukaryotic spliceosome (Lambowitz and Zimmerly, 2004). In these introns, the intron-encoded protein

(IEP) aids in self-splicing and in the process of ‘retrohoming,’ in which the intron site-specifically reverse splices into DNA. Diagrams of the intron structure and the mechanism of intron retrohoming are shown in Figure 1. Retrohoming sites are recognized primarily by base-pairing interactions between the intron RNA and target DNA, and it is therefore possible to change the specificity of intron insertion by modifying the target-site recognition sequences in the intron RNA. Algorithms have been developed for efficiently retargeting both the Ll.LtrB intron from *L. lactis* (Perutka *et al*, 2004) and the Ecl5 intron from *E. coli* (Zhuang *et al*, 2009) and are available online (www.targettrons.com). The two intron types have little apparent sequence homology.

Targettrons function in a wide variety of bacteria. Rodriguez *et al* (2009) compiled a list of 11 different bacterial genera in which targettrons have been shown to function, and at least 9 more genera have been added to the list of known targets since that time (Alonzo *et al*, 2009; Zarschler *et al*, 2009; Park *et al*, 2010; Steen *et al*, 2010; Kumar *et al*, 2011; Palonen *et al*, 2011; Akhtar and Khan, 2012; Cheng *et al*, 2013; Smith *et al*, 2013). For some genera, such as *Clostridia*, targettrons have proven to be the first genetic tool of significant utility beyond suicide plasmids, which have low efficiency and are unstable (Heap *et al*, 2007).

While targettrons are conventionally used for gene knock-outs, their efficiency, specificity, and broad applicability make them attractive for tandem use with other general-utility genome-engineering tools, such as site-specific recombinases. Here, we modified targettrons to efficiently carry *lox* sites to defined genomic loci and thereby developed a generalizable approach to genome editing that can be adapted with minimal modification to a wide variety of bacterial strains. We use this system, called GETR (Genome Editing via Targettrons and Recombinases), to generate large-scale chromosomal insertions, deletions, inversions, and one-step cut-and-pastes, and we demonstrate its use in the Gram-negative *E. coli* and *Shewanella oneidensis* bacteria, as well as in the Gram-positive *Staphylococcus aureus* and *Bacillus subtilis* bacteria.

Results

Engineering *lox*-site inserts for improved intron efficiency

One of the most well-understood site-specific recombinases is the Cre/*lox* system, first discovered in 1981 in the

P1 bacteriophage (Sternberg and Hamilton, 1981). The Cre protein catalyzes recombination between *lox* sites. *Lox* sites are 34 nucleotides long and consist of 13-nucleotide palindromic repeats flanking an 8-nucleotide linker (Hoess and Abremski, 1984). The linkers are asymmetrical and thus have a specific orientation that controls the directionality of recombination by the Cre protein (see Supplementary

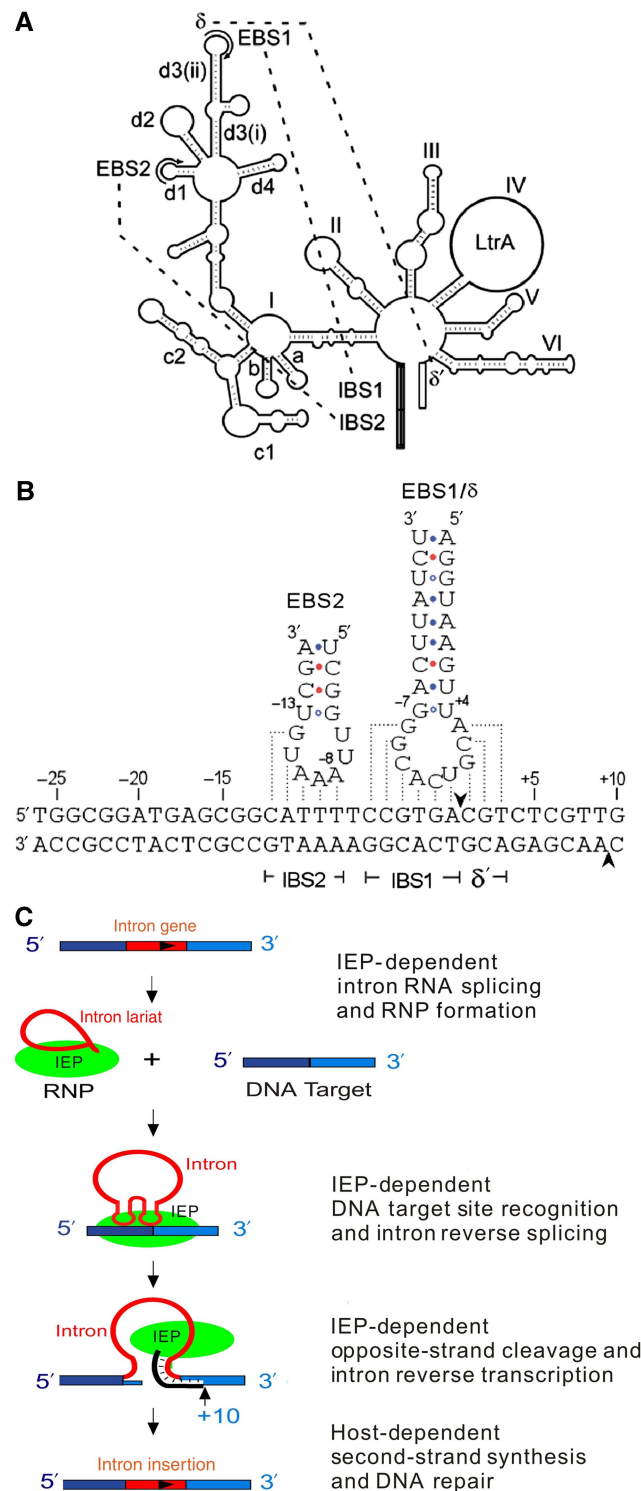


Figure 1 Targettron structure and mechanism. (A) Schematic of the structure of the Ll.LtrB intron, adapted from Perutka *et al* (2004). Domains are labeled, and dotted lines represent contacts made during splicing. For biotechnology applications such as that presented here, the *LtrA* gene (intron-encoded protein) is removed and expressed separately. The *MluI* restriction site used for cloning into the intron occupies the location where the intron-encoded protein is found in the wild type. (B) Base-pair contacts involved in DNA target-site recognition, using the LtrB.lacZ.635s intron (see Supplementary Table S1) as an example (analogous contacts are made during intron splicing from RNA). Dotted lines show base-pairing interactions, and the arrows show the sites where the DNA is cut. Numbering is relative to the insertion site of the intron, which is between the -1 and +1 bases. (C) General mechanism of intron splicing and targeting. RNP, ribonucleoprotein; IEP, intron-encoded protein. While the structure of the Ecl5 intron differs from that of Ll.LtrB (Zhuang *et al*, 2009), the target-site base-pairing contacts and overall mechanism are similar.

Figures S1A and B). The linker also determines the specificity of recombination, such that any given *lox* site can only recombine with other *lox* sites having compatible linkers. Many linker mutants with orthogonal specificities are known (Siegel *et al*, 2001; Langer *et al*, 2002). Mutations in the flanking repeats (the ‘arms’) affect the binding affinity of Cre and can be used to control the direction of recombination. For example, variant *lox66* and *lox71* sites are functional but upon recombination with each other form a *lox72* site that is no longer recognized by Cre (Albert *et al*, 1995) (see Supplementary Figure S1C).

We began by inserting *lox* sites into intron domain IV (the site of the IEP open reading frame (ORF) in wild-type introns) of the *lacZ*-targeting introns LtrB.LacZ.635s and Ecl5.LacZ.912s (Supplementary Table S1; the numbers 635 and 912 indicate the position in the *lacZ* gene at which the introns insert, and ‘s’ (as opposed to ‘a’) indicates that the introns insert into the sense strand of the gene). However, some of the

initial *lox*-site constructs significantly reduced the integration efficiency of the introns. We hypothesized that the tight hairpins that were predicted to be formed by the symmetric *lox* sites (Zuker, 2003) might disrupt the tertiary structure of the intron, and thereby inhibit splicing or insertion. The *lox*-site inserts were therefore redesigned to include the sequence ‘GAGAG’ at both ends of the insert to increase the flexibility of the structures, as judged by the presence and size of non-base-pairing regions in the predicted structures, particularly at the base of the stem. This largely restored insertion efficiency (see Figure 2). We hypothesize that this trend occurs because inserts having inflexible structures are more likely to interfere with proper folding of the catalytic structures of the intron than inserts having flexible structures, which can be moved out of the way of other formations.

Statistical analyses (see Supplementary information) confirmed that the inserts fall into two classes: one of wild-type efficiency and one of impaired efficiency. The 2ML5 insert was

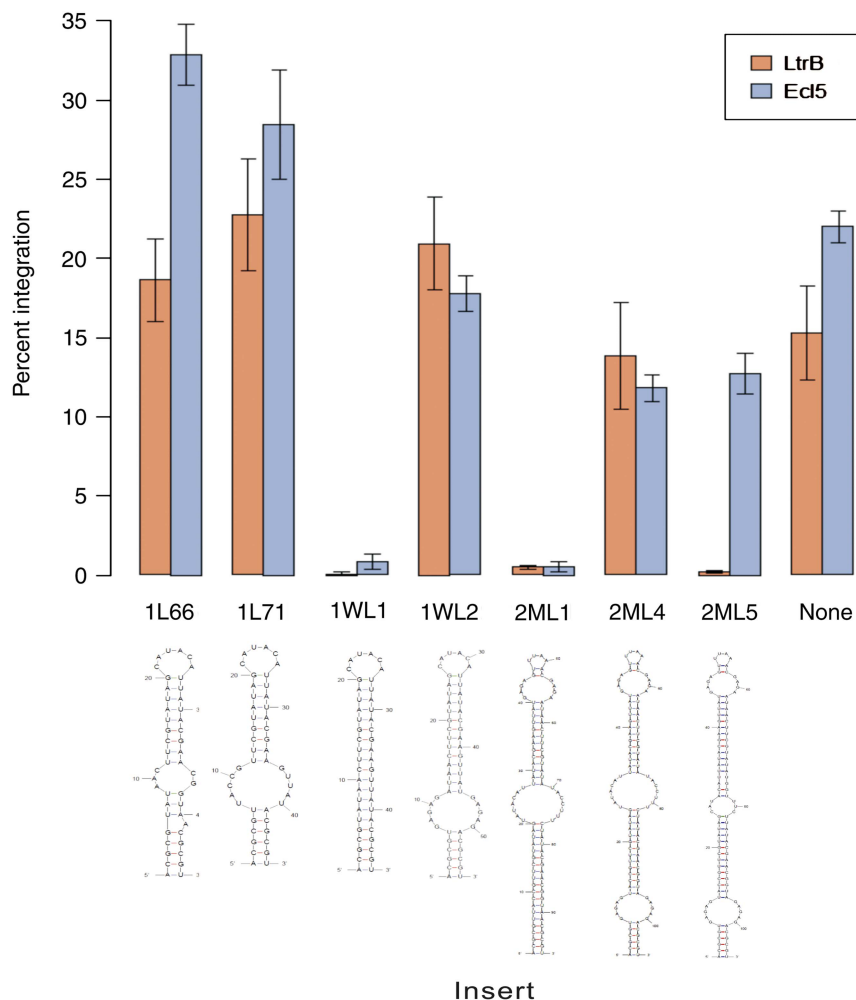


Figure 2 Effect of *lox* insert on intron efficiency. Different *lox* sequences were inserted into the *Mlu*I site of the LtrB.LacZ.635s (LtrB) and Ecl5.LacZ.912s (Ecl5) introns (see Supplementary Table S1), and efficiency was screened by counting the number of white colonies obtained. Error bars are standard error of three replicates, each representing a separate transformation of the intron plasmid into the recipient strain. The identities of the *lox* inserts are as follows, where all sequences are flanked by *Mlu*I sites: 1L66–*lox66*; 1L71–*lox71*; 1WL1–*loxP* (wild-type *lox* site); 1WL2–1WL1 plus a flexible base; 2ML1–*lox511* with the *lox71* arm mutation (*lox511/71*) and *loxFAS* with the *lox66* arm mutation (*loxFAS/66*), separated by a *Pme*I site and a short linker; 2ML4–2ML1 plus a flexible base; and 2ML5–identical to 2ML4 except with *lox71* and *loxm2/66* instead of *lox511/71* and *loxFAS/66*. At the bottom are the RNA structures of the inserts as determined by Mfold (Zuker, 2003).

the only insert that performed markedly differently in the Ll.LtrB intron versus the Eci5 intron. The poor performance of the 2ML5 insert in the Ll.LtrB intron may be due to its relative inflexibility at the central non-base-pairing region as compared with the 2ML4 insert, which is otherwise similar in structure.

Overview of genomic manipulations of *E. coli* chromosome

Introns were targeted to a variety of insertion sites in *E. coli* (Figure 3); these sites were chosen to flank genomic regions that had previously been shown to be non-essential and amenable to deletion (Kolisnychenko *et al*, 2002; Fukiya *et al*, 2004). A list of introns used in the present work is given in Supplementary Table S1. Figures 4A–D show schematics for using this system to implement insertions, deletions, inversions, and cut-and-paste operations, respectively.

Insertions (recombination-mediated cassette exchange)

After targetron integration, genomic insertions were performed by recombination-mediated cassette exchange (RMCE), using the Eci5.LacZ.1806s intron to deliver an incompatible pair of *loxP* and *loxm2* (Langer *et al*, 2002) sites to the genome (Figure 4A). The 1806s intron for inserting into *lacZ* was used for most subsequent modifications instead of the 912s intron due to its higher efficiency, approaching 97% (Zhuang *et al*, 2009). The use of incompatible linker mutations prevents inversion or deletion of the sequence between the *lox*

sites, and the use of arm mutants makes the recombination reaction unidirectional and allows multiple insertions to be made without cross-reactivity. To examine the effect of various experimental parameters (in particular, incubation time, copy number of the delivery plasmid, and strain background) on the efficiency of RMCE, we first delivered a T7 promoter to the genome along with the *lox* sites in the Eci5.LacZ.1806s intron, and separately provided both a promoterless GFPuv gene, flanked by *lox* sites on a pUC19 vector or a pACD vector (derived from pACYC184), and a Cre-expressing plasmid (pQL269; Liu *et al*, 1998). The pUC19 high-copy plasmid is present at about 500 copies per cell (Chambers *et al*, 1988), whereas pACYC is present at only about 20 copies per cell (Chang and Cohen, 1977). In co-transformed cells, GFP expression (*via* the endogenous T7 RNA polymerase) should only occur upon insertion into the genomic target site (see Figure 5A). Two *E. coli* strains, HMS174(DE3) (a K-12 strain related to MG1655) and BL21(DE3) (a B strain), that contained intron-delivered *lox* sites were used and were plated at 1, 2, and 3 days after transformation. Efficiency was gauged by manually counting colonies. The results are shown in Figure 5B, and a statistical analysis of the results is presented in Supplementary information.

In interpreting these results, we first note the significant effect of increasing time on efficiency of insertion. This is likely because interaction between the delivery plasmid and the chromosome occurs at random during any given period, and thus the chance of an interaction occurring increases with time (though in general little is gained by waiting 3 days as opposed to 2). The better performance of the lower-copy vector versus the high-copy vector is surprising at first but may be a result of the lower opportunity for Cre-mediated swapping of cassettes between plasmids in the lower-copy case. The effect of strain, which was not statistically significant except in interaction with other factors, seems to be in modulating the influence of time and copy number. In particular, the effect of time was more pronounced in HMS174(DE3), and the effect of vector copy number was more pronounced in BL21(DE3).

We then further examined the effect of genomic location on RMCE insertion efficiency by repeating the experiment using the lower-copy pACD vector in HMS174(DE3) at two new loci, the *galK* gene and the *malT* gene. The results are shown in Figure 5C. While efficiency of integration into the *malT* locus was worse than at the other loci (see Supplementary information for a full statistical analysis), in all cases the efficiency was high enough by the second day that screening for insertions via colony PCR could be easily performed.

To demonstrate not just the efficiency but the broad utility of this system, we then proceeded to insert the 12-kb DEBS1-TE polyketide synthase operon (Wiesmann *et al*, 1995; Kodumal *et al*, 2004) into the *lacZ* gene of *E. coli* K207-3 (Murli *et al*, 2003). The delivery-plasmid *lox* sites used in earlier experiments were inserted on either side of the operon in the pET26b-DEBS1-TE plasmid using conventional cloning methods. The pET vectors are built on the pBR322 backbone (Rosenberg *et al*, 1987), which is similar in copy number to the pACYC backbone used for the pACD plasmids (Green and Sambrook, 2012). Insertion of the entire operon into the *lacZ* gene was facile (as judged by PCR across an insertion junction) and also showed a trend of increased insertion efficiency as a

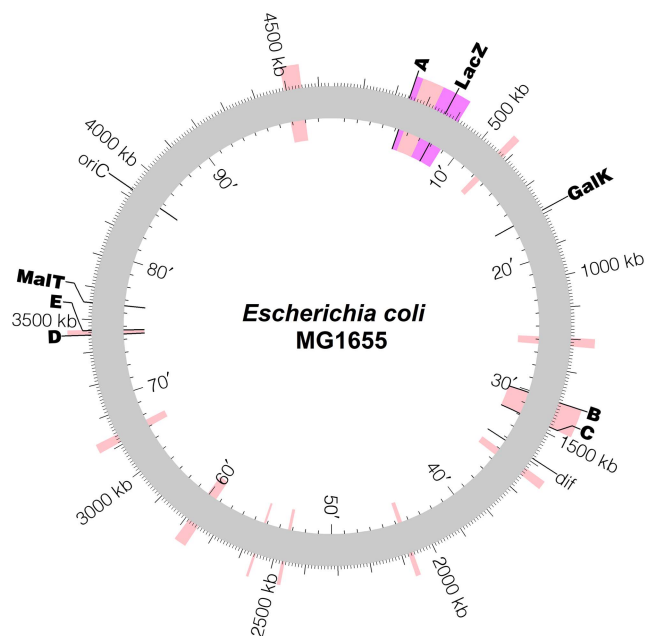


Figure 3 Genomic integration sites of the introns. Insertion sites of introns used in the present work are labeled in bold type. Pink highlights are regions previously deleted by Kolisnychenko *et al* (2002), and the purple highlight is a region previously deleted by Fukiya *et al* (2004). The intron used for *lacZ* is Eci5.LacZ.1806s (see Supplementary Table S1 online) unless otherwise noted. Image made using Circos (Krzywinski *et al*, 2009).

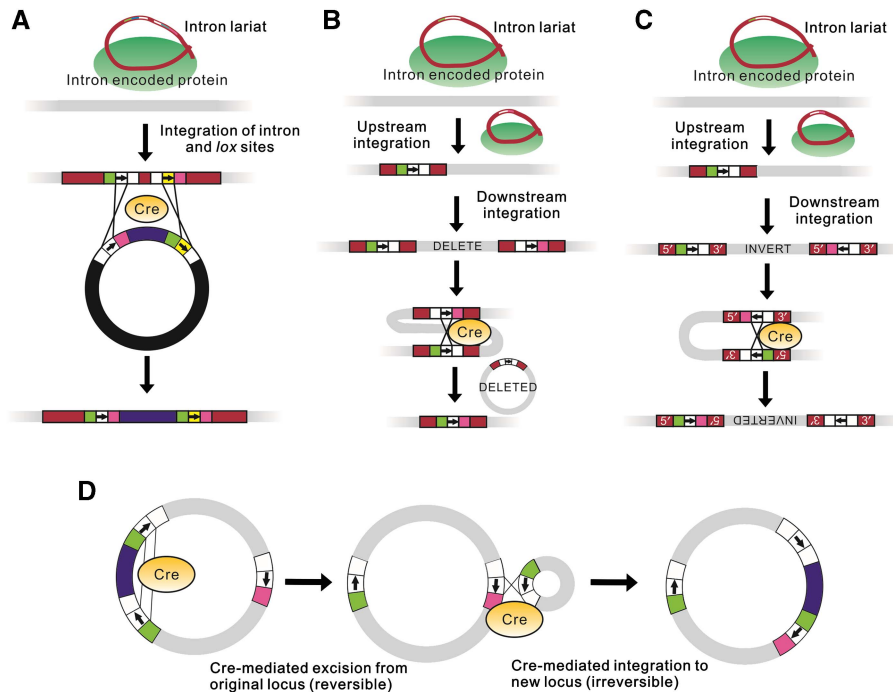


Figure 4 Genome edits performed. In the figure, *lox* sites are represented by three boxes (arm, linker, and arm), where white represents wild-type *loxP* sequence, green represents the *lox71* mutant arm, pink represents the *lox66* mutant arm, yellow represents an incompatible *lox* linker, and the arrows represent the linker orientation. **(A)** Inserting exogenous DNA (recombinase-mediated cassette exchange). Two *lox* sites having incompatible linker regions and differing arm mutations are delivered to the genome using an intron. The sequence to be inserted is then delivered between *lox* sites identical to those in the genome except having opposite arm mutations. The formation of non-functional *lox72* sites makes the process irreversible. **(B)** Procedure for deleting genomic sequences. A *lox71* site is carried by an intron upstream of the region to be deleted, and a *lox66* site is carried downstream. Cre-mediated recombination then deletes the intervening region, leaving a non-functional *lox72* site behind. **(C)** Procedure for inverting genomic sequences. The procedure is the same as in **(B)**, except the *lox* sites have opposing orientations. In the example shown, inverted repeats result from the recombination and would be subsequently deleted by the cell. **(D)** Procedure for one-step cut-and-paste (after placing *lox* sites using introns (not shown)). The first (reversible) step is Cre-mediated deletion, followed by Cre-mediated reinsertion at the target site that is made irreversible by the formation of a *lox72* site.

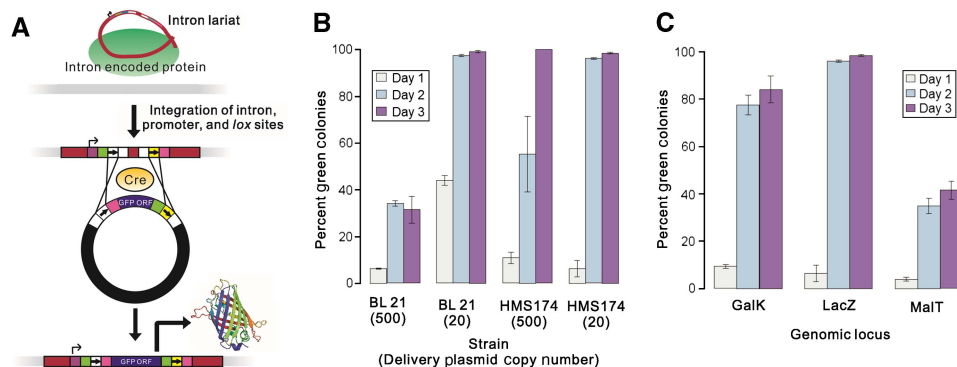


Figure 5 GFP reporter assay for Cre/*lox*-mediated gene insertion. **(A)** Overview of the method. A T7 promoter is first delivered to the genome with an intron. A promoterless GFP ORF (with ribosome binding site) is then inserted via Cre/*lox*, such that GFP expression is only seen upon insertion. Color-coding as in Figure 4. **(B)** Results as a percentage of green colonies, by strain, delivery-plasmid copy number, and incubation time. Error bars are the standard error of three replicates. On day 3, the HMS174(DE3) (High) colonies were visually homogenous and were thus also assayed by PCR. **(C)** Results as a percentage of green colonies, by genomic location, in HMS174 using the lower-copy vector. The data for *lacZ* are identical to those for HMS174(DE3) (20) in **(B)**. Error bars are the standard error of three replicates.

function of incubation time; after 3 days of incubation, 25/25 screened colonies tested positive for the insertion.

One of the features of this manipulation is that insertion also inactivated the *LacY* gene. Thus, IPTG sensitivity is reduced due to a smaller amount being transported into the cell (Mehdi *et al*, 1971), and protein production can be more precisely modulated, resulting in a lower fitness load on the cell. When two of these colonies were screened more fully using

overlapping PCR amplifications that were subsequently sequenced, both were found in fact to contain the entire polyketide synthase operon without error (see Supplementary Figure S2). In principle, insertions of any size could be made at similar efficiency, limited only by the constraints of genome structure (Esnault *et al*, 2007).

We have devised a set of vectors to facilitate the use of RMCE insertion of cassettes into the genome: pX10, pX11, pX20,

and pX21. All the vectors contain the *sacB* gene for counter-selection on sucrose, as well as two T7 terminators between the *sacB* gene and the *lox* sites to prevent read-through from the *sacB* promoter to the delivery target between the *lox* sites. Vectors pX10 and pX11 contain the incompatible *lox* pair of *lox*FAS/71 and *lox*511/66 for use with the 2ML4 pair of *lox* sites (see Figure 2) at the integration site, and vectors pX20 and pX21 contain the incompatible *lox* pair of *lox*m2/71 and *lox*66 for use with the 2ML5 pair of *lox* sites (see Figure 2) at the integration site. For ease of cloning, vectors pX10 and pX20 contain a *PmeI* restriction site between the *lox* sites, whereas vectors pX11 and pX21 have the multiple cloning site from the pUC vectors between the *lox* sites.

Deletions

We used the same methods to demonstrate larger scale manipulations of the cellular genome of *E. coli* MG1655(DE3). First, we attempted large-scale deletions (Figure 4B). The *A-lacZ*, *D-E*, and *B-C* regions (see Figure 3) were deleted both sequentially (in the order given) and simultaneously. A set of three mutually incompatible *lox* sites (*loxP*, *lox2272*, and *loxN*; Livet *et al*, 2007) was used for the simultaneous triple deletion. The use of arm mutants once again allowed multiple deletions to be made without cross-reactivity.

Screening for the deletions was performed via colony PCR as depicted in Figure 6A. Three PCR amplifications can be used to characterize each deletion: two amplicons that bridge the genomic sites at which *lox*-carrying targetrons are inserted,

and one amplicon that bridges the expected deletion between those sites. The first two PCRs testing for insertion are expected to give relatively small bands (several hundred base pairs) when performed on the wild-type strain and larger bands (about 1 kb larger) when performed on a strain harboring targetrons at the expected sites. Upon successful deletion of the intervening region, all of these bands should no longer be generated. Instead, the PCR that bridges the two sites should yield a new band of a predicted size. Doing all three PCR amplifications on all three strains (wild-type, wild-type harboring insertions, and recombined (induced with Cre)) provided clear diagnostic signatures of the recombination events. When artifact bands near the sizes of expected bands were observed, these were further analyzed by sequencing to ensure that they did not represent an off-target rearrangement. The predicted sizes of all of the expected amplicons of the present work (as shown in the gels of Figures 6–8 and Supplementary Figures S3–S5) are listed in Supplementary Table S2.

The gels used for verifying the occurrence of these deletions are shown in Figure 6. In particular, Figure 6B shows the three PCR amplifications performed on the three successive, engineered strains that ultimately resulted in a deletion of 121 kb between the *A* and *lacZ* loci. The deletion-bridging amplicon (Au/Ld-I in Figure 6) from the strain that was finally exposed to the Cre protein (*E. coli* MG1655 E1) was sequenced and found to conform to expectations.

As deletions are added, verification becomes more complex but is performed in exactly the same manner. Figure 6C shows the PCR amplifications for verifying the sequential double-

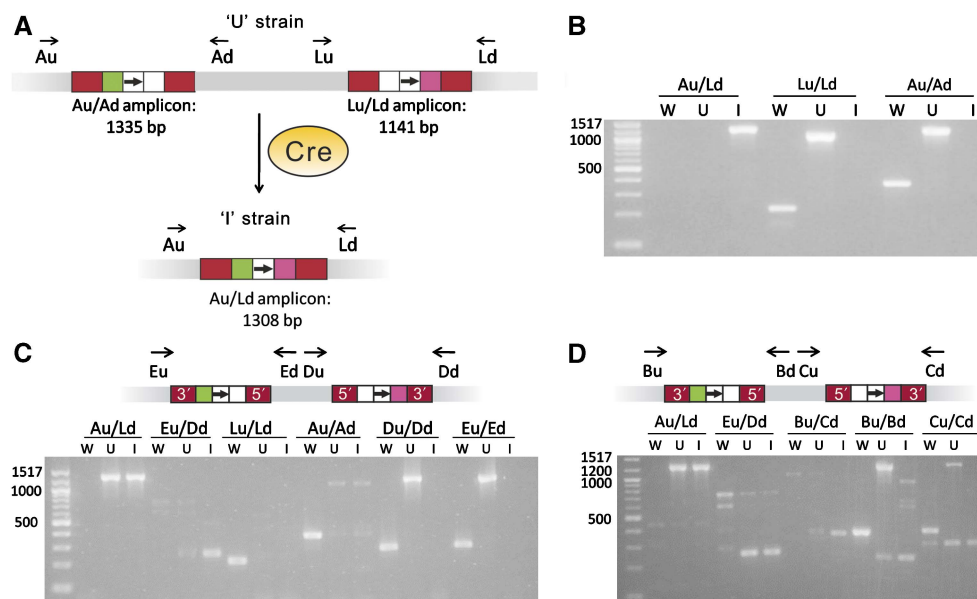


Figure 6 Verification of genomic deletions. In the figure, 'W' refers to the wild-type *E. coli* strain MG1655(DE3); 'U' refers to the relevant uninduced strain, in which introns and *lox* sites have been placed but Cre has not been added; and 'I' refers to the induced strain, which results from Cre-mediated recombination of the 'U' strain. For primers, the first letter indicates the genomic location the primer amplifies (where 'L' refers to the *lacZ* locus), and the subsequent 'u' or 'd' designates the primer as 'up' or 'down'. PCR products are designated by the two primer names separated by a slash. '5'' or '3'' refers to the sense strand of the intron. (A) Methodology, using the deletion of the *A-lacZ* region as an example. (B) Verification of the strain (*E. coli* MG1655 E1) containing a deletion of the *A-lacZ* region, as shown in (A). (C) Verification of the sequential double-deletion strain (*E. coli* MG1655 E6), with schematic corresponding to the 'U' strain. 'U' here is *E. coli* MG1655 E1 with introns inserted to delete the *D-E* region. The Eu/Dd PCR amplifies the *D-E* deletion site (the *D-E* deletion leaves an inverted repeat behind). (D) Verification of the sequential triple-deletion strain (*E. coli* MG1655 E10), with schematic corresponding to the 'U' strain. 'U' here is *E. coli* MG1655 E6 with intron insertions for the deletion of the *B-C* region. Bu/Cd amplifies the *B-C* deletion site (the *B-C* deletion leaves behind an inverted repeat).

deletion strain (*E. coli* MG1655 E6) that harbors deletions of both the *A-lacZ* and *D-E* regions. The Eu/Dd-I band for verifying the deletion of the *D-E* region is small because the recombination results in an inverted repeat of intron sequences that is subsequently removed by homologous recombination. Sequencing results confirmed this interpretation. There were some unexpected PCR-amplified bands. These bands, in particular the Eu/Dd-I, Au/Ad-U, and Au/Ad-I bands, were sequenced and found to match genomic sequences from unrelated regions, and not to off-target rearrangements.

Similarly, Figure 6D shows the PCR amplifications for verifying the sequential triple-deletion strain (*E. coli* MG1655 E10) that contains a deletion of the *B-C* region in addition to the *A-lacZ* and *D-E* regions. The unexpected band in the Bu/Cd-U lane was sequenced and found to be from an unrelated genomic region. The Bu/Cd-I amplicon that confirms the deletion also represents the formation and removal of an inverted repeat, which was confirmed by sequencing. The simultaneous triple-deletion strain (E9) and a strain containing a single deletion of the *D-E* region (E11) were verified in the same manner.

As was the case with insertions, the efficiency of deletions approached 100%, with the expected deletion being found in every colony tested. Off-target recombination was rare; in strains designed for deletion, only 1/60 was found to have an inversion when screened after Cre induction; recombination between *lox72* sites was not detected in any of the modifications reported herein. The removal of inverted repeats upon formation of the *D-E* and *B-C* deletions is interesting in that the *lox* sites are removed entirely and the size of the scar is reduced from hundreds to tens of base pairs.

Inversions

GETR proved to be robust for other types of recombination that are not easily achieved by other methods. The same methods used to detect deletions can be used to detect inversions, except that four different PCRs are used to verify an inversion: two amplicons bridging the insertions sites and two amplicons bridging the new ends of the inversion (see Figure 7A). Several inversions (see Figure 4C) were executed: namely, between the *A-lacZ*, *B-lacZ*, *E-lacZ*, and *D-E* loci. All colonies screened by PCR soon after addition of Cre tested positive for the expected inversion. Some inversions were only detected immediately after adding Cre and were not detected at later time points. This is in line with previous studies of inversions in the *E. coli* genome, some of which are not well tolerated (Esnault *et al*, 2007). Inversions into the *lacZ* locus were transient when an inverted repeat was formed and subsequently deleted but were stable when non-homologous introns were used, suggesting that the intron sequences may function as a buffer against otherwise deleterious rearrangements at this site.

Recombination back to the original state via homologous recombination of the introns could be detected in some cases but was not seen when non-homologous introns were used. In other words, inversions between *lox* sites in homologous introns may be reversible, but inversions between *lox* sites in non-homologous introns are irreversible. We also tested for the

presence of uninverted chromosomes soon after induction of an irreversible inversion between the *E* and *lacZ* loci. All 10 colonies assayed tested positive for both inverted and uninverted chromosomes, though uninverted chromosomes were not found after restreaking.

Gels for verifying the stable inversions are found in Figure 7. Figure 7B shows the PCRs used to verify an inversion between the *A* and *lacZ* loci, as depicted in Figure 7A. Figure 7C shows an analogous set of PCRs for verifying an inversion between the *D* and *E* loci. As these introns are present in opposite orientations upon integration, inversions can occur via homologous recombination, and this inversion is in fact detected at low levels before the addition of Cre. Similarly, the uninverted (reverted) state can still be detected after the addition of Cre. These bands were confirmed by sequencing. In those strains where inversion had occurred in the absence of Cre, unrecombined *lox* sites were found, whereas in those strains where inversions back to the wild-type state had apparently occurred after the induction of Cre, recombined *lox* sites were found. These results are consistent with homologous recombination between the introns rather than catalyzed recombination between the *lox* sites. No artifacts were seen in these instances, consistent with the presence of a template that could be amplified by the primers.

Figure 7D demonstrates the difference between using homologous introns (where both introns are of the EcI5 type) versus non-homologous introns (one EcI5 and one Ll.LtrB) for delivering *lox* sites to create substantially identical inversions. When non-homologous introns are used, as in the E5 strain, the inversion is only detected (via the Lu₀/Ed-I and Eu/Ld₀-I bands, which are of the expected size and were confirmed by sequencing) after adding Cre, and reversions back to the original state via homologous recombination were not detected. However, when homologous introns were used, as in the E2 strain, PCR products that should only have been seen upon inversion were also seen in the absence of Cre; and furthermore, the uninverted state could still be detected after the addition of Cre, consistent with homologous recombination between introns. The bands of the sizes expected for recombination events, whether due to Cre-*lox* or homologous recombination, were confirmed by sequencing.

One-step cut-and-paste

Finally, we used combinations of three *lox* sites to effect unique one-step cut-and-paste reactions (Figure 4D). We use the term 'one-step' because the designated region moves directly from one part of the genome to another upon adding Cre, without the requirement for a stable intermediate, such as a plasmid, to act as a shuttle. In particular, we transferred the *D-E* region to the *B* locus, and the *A-lacZ* region to the *E* locus. Six different PCR amplifications were used to validate a stable cut-and-paste reaction, as shown in Figure 8: three amplicons bridging the three intron integration sites (the left set of three triplets in Figures 8A and B), one amplicon bridging the site of the 'cut' (the fourth set of triplets in Figures 8A and B), and two amplicons bridging the boundaries of the 'paste' region (the fifth and sixth set of triplets in Figures 8A and B). The *D-E* to

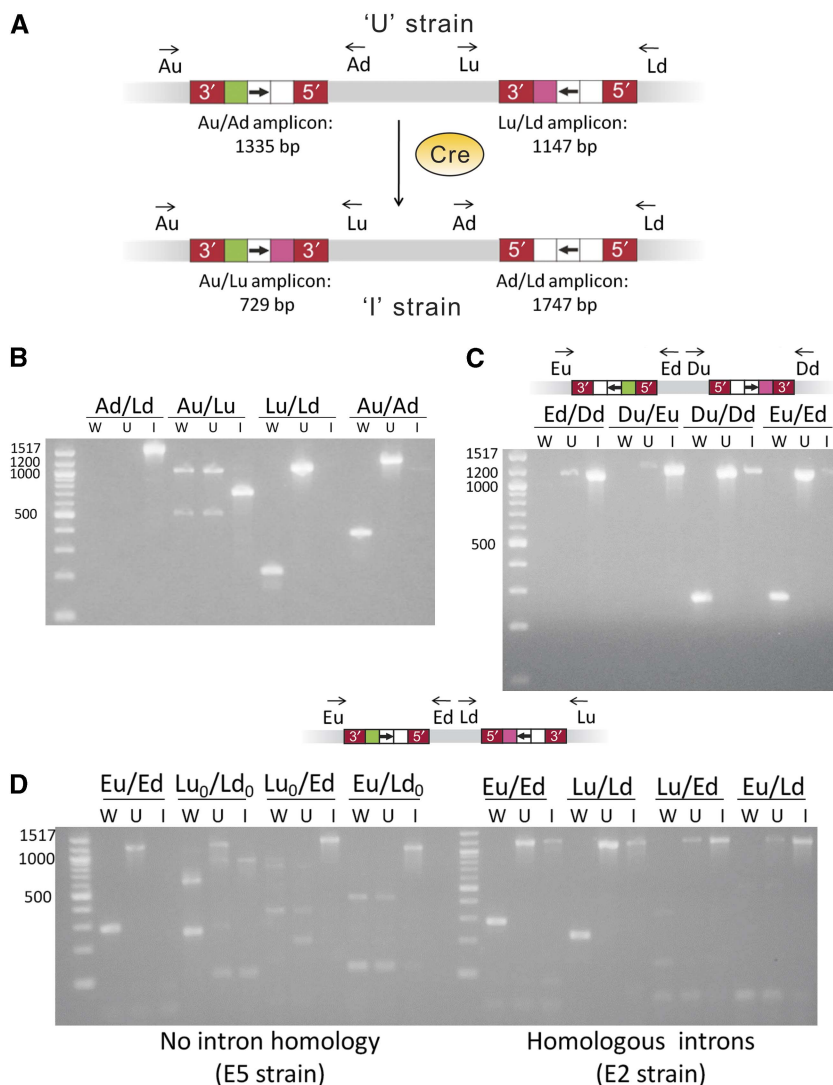


Figure 7 Verification of genomic inversions. Letter designations are as in Figure 6. (A) Methodology, using the inversion of the *A-lacZ* region as an example. '5'' or '3'' refers to the sense strand of the intron. (B) Verification of the strain containing an inversion of the *A-lacZ* region (*E. coli* MG1655 E3), as shown in (A). (C) Verification of the strain containing an inversion of the *D-E* region (*E. coli* MG1655 E4), with schematic corresponding to the 'U' strain. (D) Comparison of a strain containing an inversion of the *E-lacZ* region using homologous introns (*E. coli* MG1655 E2) and a strain containing the same inversion using non-homologous introns (*E. coli* MG1655 E5), with schematic corresponding to the 'U' strains. The subscript in Lu₀ and Ld₀ signifies that these primers amplify the insertion site of the LtrB.LacZ.635s intron rather than the Ecl5.LacZ.1806s intron used elsewhere.

B transfer was detected (via the 'cut' and 'paste' bridging amplicons) in every colony soon after exposure to Cre but was not stable upon restreaking.

The *A-lacZ* region was stably translocated to the *E* locus in both possible orientations, with 3/5 and 4/5 colonies positive for the translocations after overnight growth in liquid culture after transformation. Gels for verifying these recombinations can be found in Figure 8. In the case shown in Figure 8A, intron homology allows inversions to occur back and forth between the *lacZ* and *E* loci, but the complete cut-and-paste was only seen upon addition of the Cre protein. Figure 8B shows a similar case where the orientation of the *lox* site at the target (*E*) locus is reversed with respect to the case shown in Figure 8A. The expected rearrangement was obtained, but since the insertion at *E* is in the opposite orientation, inversions and reversions resulting from intron homology

were avoided. The key bands for confirming the cut-and-pastes (Au/Ld-I, Lu/Ed-I, Eu/Ad-I, Eu/Lu-I, and Ed/Ad-I) as well as the bands expected to result from homologous recombination (Lu/Ld-U and Lu/Ed-U in Figure 8A) were confirmed by sequencing.

Growth of *E. coli* strains with chromosomal rearrangements

A summary of the genomic rearrangements generated in *E. coli* is given in Table I, and a list of *E. coli* strains containing these rearrangements is given in Table II, along with doubling times measured for these strains as a proxy for fitness. A statistical analysis of the doubling times indicated that the strains fall broadly into two groups, one group having approximately

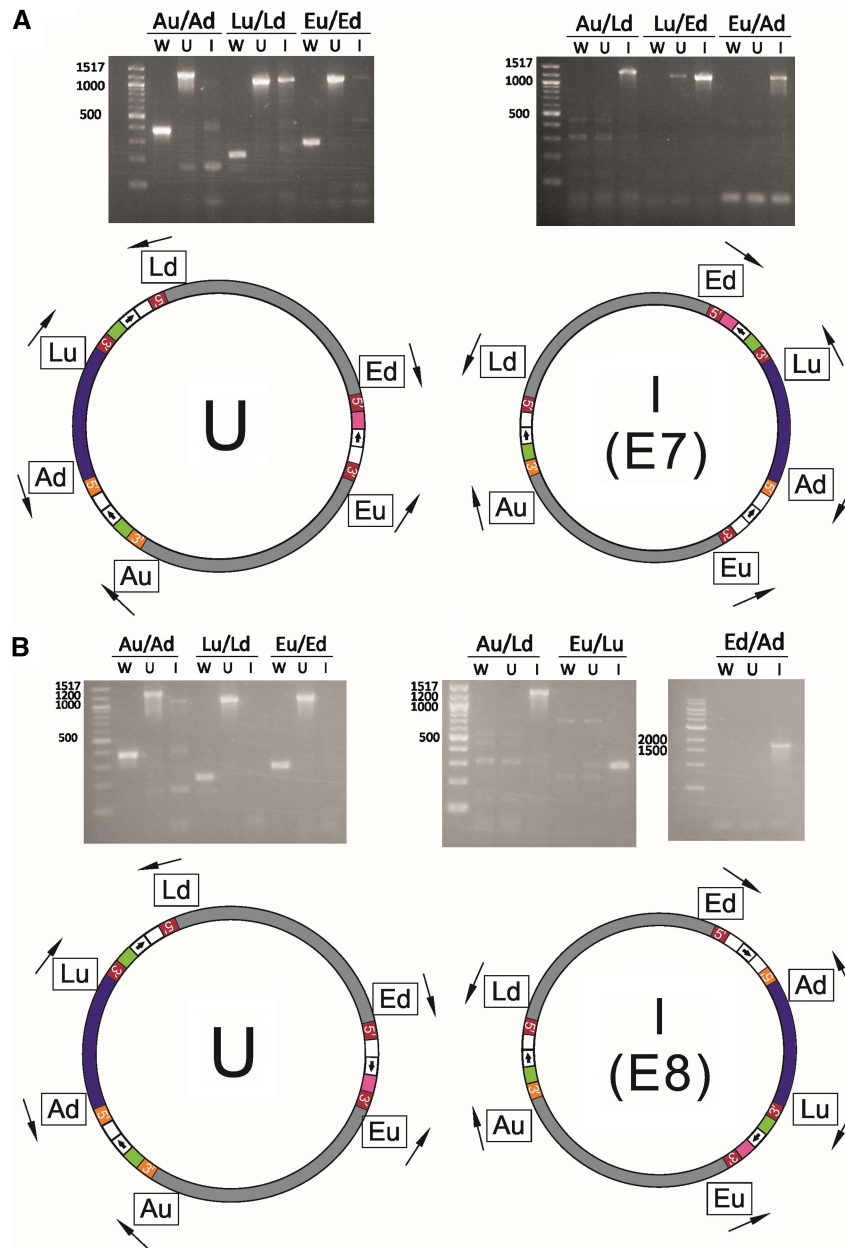


Figure 8 Verification of one-step cut-and-pastes. Orange is L1.LtrB intron sequence, which is non-homologous with respect to Ecl5 intron sequence shown in red. Letter and number designations are as in Figure 6. **(A)** Verification of the strain (*E. coli* MG1655 E7) containing a cut-and-paste (translocation) of the *A-lacZ* region to the *E* locus in the reverse orientation. **(B)** Verification of the strain (*E. coli* MG1655 E8) containing a cut-and-paste (translocation) of the *A-lacZ* region to the *E* site in the forward orientation.

wild-type growth, and the other group having impaired growth (see Supplementary information). All strains that either lacked the *A-lacZ* region or had an inversion between the *E* and *lacZ* regions showed significantly impaired growth. Interestingly, the strains containing a cut-and-paste of the *A-lacZ* region to the *E* site displayed wild-type growth rates.

Genome engineering in diverse bacteria

While GETR is obviously broadly useful for creating virtually any type of rearrangement, the real utility of the method

appears when moving beyond *E. coli* as a model system. We therefore applied the method to make genomic modifications in three additional phylogenetically diverse species: *Staphylococcus aureus*, *B. subtilis*, and *S. oneidensis*.

The Gram-positive bacterium *Staphylococcus aureus* is an intensely studied human pathogen, and the rise of drug-resistant strains in recent years has given new urgency to the development of prophylactic and therapeutic approaches to treatment (Otto, 2012). We therefore attempted to delete the 15-kb *Staphylococcus aureus* pathogenicity island 1 (SaPI-1) from *Staphylococcus aureus* RN10628 (Ubeda *et al*, 2009), to create a strain that might serve as a live vaccine. Highly

Table I Summary of intragenomic rearrangements in *E. coli*

Loci recombined ^a	Largest distance between recombined loci	Type of modification	Inverted repeat generated	Stability ^b	Recombination in the absence of Cre ^c
<i>A-lacZ</i>	121 kb	Deletion	No	Stable	Yes
<i>A-lacZ</i>	121 kb	Deletion	No	Stable	No
<i>B-C</i>	81 kb	Deletion	Yes	Stable	No
<i>D-E</i>	14 kb	Deletion	Yes	Stable	No
<i>A-lacZ</i>	121 kb	Inversion	Yes	Unstable	No
<i>A-lacZ</i>	121 kb	Inversion	No	Stable	No
<i>B-lacZ</i>	1 Mb	Inversion	Yes	Unstable	No
<i>E-lacZ</i>	1.5 Mb	Inversion	No	Quasi-stable	Yes
<i>E-lacZ</i>	1.5 Mb	Inversion	No	Stable	No
<i>D-E</i>	14 kb	Inversion	No	Quasi-stable	Yes
<i>A-lacZ</i> to <i>E</i>	1.5 Mb	Cut-and-paste	Yes	Stable	Some
<i>D-E</i> to <i>B</i>	2.1 Mb	Cut-and-paste	Yes	Unstable	Some

^aThe LtrB.LacZ.635s intron was used in the first *A-lacZ* deletion (with recombination in the absence of Cre) and in the stable *A-lacZ* and *E-lacZ* inversions; the Ecl5.LacZ.1806s intron was used in all other cases (see Supplementary Table S1 online).

^b'Stable' means the recombination remained present unchanged through multiple rounds of regrowth. 'Unstable' means the recombination was detected initially but was not detected after multiple rounds of regrowth. 'Quasi-stable' means the recombination was still detected after multiple rounds of regrowth, but back-recombination due to homologous recombination was also detected.

^cRecombination was seen only in the absence of Cre when (1) the introns were homologous and (2) the introns were oriented so as to allow the homologous recombination to occur. 'Some' recombination in the absence of Cre for the cut-and-pastes refers to the fact that inversions caused by homologous recombination were detected, but the complete cut-and-paste did not occur without Cre.

Table II Doubling times of *E. coli* strains with intragenomic rearrangements

Strain	Doubling time	Standard error ^a	Description
MG1655	24.94	0.1	
MG1655(DE3)	24.44	0.2	Base strain for E1-E11
E1	28.33	0.5	<i>A-lacZ</i> deletion
E2	33.25	0.6	<i>E-lacZ</i> inversion (reversible)
E3	22.03	0.1	<i>A-lacZ</i> inversion (irreversible)
E4	24.65	0.4	<i>D-E</i> inversion (reversible)
E5	30.5	1.2	<i>E-lacZ</i> inversion (irreversible)
E6	33.94	0.8	<i>lacZ-A</i> , <i>D-E</i> deletion
E7	25.24	0.5	<i>lacZ-A</i> region to <i>E</i> , reverse orientation
E8	23.65	0.4	<i>lacZ-A</i> region to <i>E</i> , forward orientation
E9	29.27	1.0	<i>lacZ-A</i> , <i>D-E</i> , <i>B-C</i> simultaneous deletion
E10	31.39	0.2	<i>lacZ-A</i> , <i>D-E</i> , <i>B-C</i> sequential deletion
E11	24.45	0.5	<i>D-E</i> del

^aError is from three replicates.

efficient L1.LtrB-type introns were generated that could integrate in the first gene (*int*) of SaPI-1 and also downstream of the pathogenicity island. After transformation of the Cre-expressing plasmid pRAB1 (Leibig *et al*, 2008), 40/40 colonies tested contained cells harboring the expected deletion, and 19/40 colonies tested still harbored SaPI-1. The deletion was detected and verified via PCR and sequencing (see Supplementary Figure S3). No chromosomes containing SaPI-1 were detected in restreaked colonies. The deletion was stably maintained.

Bacillus subtilis is a model system for the study of Gram-positive bacteria, including studies on sporulation (Earl *et al*, 2008; Higgins and Dworkin, 2012). We designed and built two L1.LtrB-type introns that inserted into the sense strands of the *sacB* and *yhcS* (*srtA*) genes of *B. subtilis* at efficiencies 98 and

91%, respectively (Yao, 2008; Whitt, 2011). We used these introns to deliver a *lox71* site to the *sacB* locus and a *lox66* site to the *yhcS* locus in *B. subtilis* 168, positioning the intervening region for inversion. Upon addition of the Cre-expressing plasmid pCrePA (Pomerantsev *et al*, 2006), 4/11 screened colonies tested positive for the inversion via colony PCR (Supplementary Figure S4). Sequencing of the PCR products gave a sequence consistent with the expected inversion. The inversion covers about 1.5 Mb of the 4.2 Mb genome and was not seen upon restreaking, indicating that it was not well tolerated.

Shewanella oneidensis is a Gram-negative bacterium that is a model system for extracellular electron transfer, with potential applications in bioremediation and energy (Fredrickson *et al*, 2008). We designed an L1.LtrB targetron that inserted into the ribosomal *rrs* genes (*rfsA* through *rfsI*) in *S. oneidensis* (see Supplementary Figure S5A). This targetron included a *loxP* site in the intron to facilitate subsequent genomic rearrangements, was cloned onto the broad-host range plasmid RP4, and was introduced into *S. oneidensis* via conjugation. We screened single colonies for insertions at each site by PCR using one primer complementary to the intron and another primer complementary to a unique chromosomal sequence near the insertion site. We initially found insertions in all copies of the *rfs* gene except *rfsC* (Supplementary Figure S5B); subsequent PCRs using other primers to more specifically detect insertions in *rfsC* did yield bands. We found the same pattern of PCR bands after growing one of the colonies overnight in liquid culture, freezing at -80°C , restreaking, and repeating the PCR amplifications on one of the resultant colonies (Supplementary Figure S5C).

Discussion

As synthetic biology continues to advance, there will be an increasing emphasis on the genome as the unit of engineering,

which allows much larger swaths of DNA to be manipulated than is possible with plasmid-based methods and enhances our ability to study the structure and function of genomes. This is already evidenced by the synthesis and transplantation of whole genomes by the Venter Institute (Lartigue *et al*, 2007; Gibson *et al*, 2008; Lartigue *et al*, 2009; Gibson *et al*, 2010) and the development of technologies such as MAGE that can site-specifically perturb multiple sites in a genome (Wang *et al*, 2009; Isaacs *et al*, 2011). However, both technologies are still time and resource intensive and are currently limited to a relatively small number of organisms (Enyeart and Ellington, 2011).

We have therefore combined the well-known Cre-lox recombinase system and the adaptable targetron technologies to create a method we dub GETR. GETR presents several advantages in comparison with recombineering and related methods that make use of the lambda Red functions. One of these advantages is the fact that GETR is very efficient, and, while the use of selectable markers is required for temporary plasmid maintenance, markers need not be used for selecting or maintaining genomic modifications. Recombineering using single-stranded DNA is simple to execute and useful for making small changes such as point mutations, but percent efficiencies are typically in the single digit range in mutator strains (specifically, *mutS* mutants) and are much lower in wild-type strains. Even then, the efficiency of inserting a sequence as large as a *lox* site is in the neighborhood of 1%, and the efficiency of deleting 10 000 bases of genomic sequence is ~0.1% (Wang *et al*, 2009). The use of the full complement of lambda Red proteins allows larger pieces of double-stranded DNA to be inserted (Datsenko and Wanner, 2000), but selectable markers are typically required and the size of possible insertions is limited to several thousands of bases. Manipulations such as inversions and cut-and-paste operations are also impossible using these methods alone.

Another advantage is that targetrons function at high efficiency in many bacterial strains and thus provide an appealing alternative to recombineering functions in many contexts. While the lambda Red system has been used outside of *E. coli*, it typically does not function as well in other organisms and in such cases generally requires 500 nucleotides of target-site homology on either side of the integration cassette to obtain reliable results (Beloin *et al*, 2003; Derbise *et al*, 2003; Rossi *et al*, 2003; Lesic and Rahme, 2008; Jia *et al*, 2010). This requires significantly more labor than the 30–40 nucleotides of homologous sequence required in *E. coli*. In *Pantoea ananatis*, the system only worked well after selection of mutants resistant to the toxic effects of the lambda Red proteins (Katashkina *et al*, 2009). In some organisms, such as *Pseudomonas syringae* (Swingle *et al*, 2010) and *Mycobacterium tuberculosis* (van Kessel and Hatfull, 2007, 2008), alternative recombineering functions have been discovered, but these do not exceed 0.1% efficiency without selection and also typically require at least 500 nucleotides of homology on either side for reliable results with selectable markers. Recombineering using single-stranded oligonucleotides for making point mutations has been reported in *Lactobacillus* species, but the electroporation of 100 µg of DNA (1,000 times the optimal amount in *E. coli*) was required for efficient mutagenesis

(van Pijkeren and Britton, 2012). Wang *et al* (2012) were also able to demonstrate gene disruption in *B. subtilis* using single-stranded DNA, but the method required the use of selectable markers and the generation of single-stranded DNA long enough to encode those markers. Datta *et al* (2008) have identified a number of other possible recombineering proteins from a variety of species, but to our knowledge none of these have yet been demonstrated as recombineering tools in their natural hosts.

Another common method of genome engineering is the use of suicide plasmids. For instance, temperature-sensitive integrable plasmids have been developed for all the systems described here other than *S. oneidensis* (Hamilton *et al*, 1989; Luchansky *et al*, 1989; Biswas *et al*, 1993; Link *et al*, 1997), and systems based on plasmids requiring expression of the *pir* (Kolter *et al*, 1978; Miller and Mekalanos, 1988) or *repA* (Leenhouts *et al*, 1996) genes to replicate have also been frequently employed. These systems are most useful for gene replacements. For the types of modifications discussed in the present work, suicide plasmids present many of the same limitations as recombineering, such as requirements for selectable markers and large regions of homology, and are limited by poor efficiency and relatively high background. The profusion of research into alternative recombineering systems in recent years, described above, is symptomatic of broader dissatisfaction with suicide plasmids as genetic tools, and the present system represents a favorable alternative to suicide plasmids for large-scale genomic modifications.

A more recent addition to the set of tools available for genome engineering is the CRISPR/Cas9 system, which adapts the site-specific RNA-mediated restriction system of bacteria toward making targeted double-strand breaks in genomic DNA (Cho *et al*, 2013; Cong *et al*, 2013; Hwang *et al*, 2013; Mali *et al*, 2013). Methods of genome engineering relying solely on the creation of double-strand breaks have not traditionally gained much traction in bacterial systems. Besides the requirement for selectable markers, the efficiency of double-strand break repair tends to be poor in bacteria, since most prokaryotes are only capable of repairing breaks via homologous recombination, and those that can carry out non-homologous end joining have only a rudimentary system for doing so (Aravind and Koonin, 2001; Hefferin and Tomkinson, 2005). CRISPR-Cas9-mediated cutting of genomic DNA has been shown to be lethal to bacteria (Bikard *et al*, 2012), but Jiang *et al* (2013) have recently reported that this method can be used to select for the integration of mutated DNA homologous into the cut site.

However, the CRISPR/Cas9 system alone is only of functional efficiency in bacteria that have very active recombination systems, such as *Streptococcus pneumoniae*, and in those systems the CRISPR-Cas9 expression construct must also be integrated into the genome along with a selectable marker and then subsequently removed. In *E. coli*, the CRISPR-Cas9 system has been shown to increase the efficiency of recombineering by cleaving the genome at unmodified sites (and thereby selecting for modified strains), but this method also has the inherent limitations of recombineering; that is, the requirement of a mutator strain for high efficiency, limitation to relatively small changes, and generally poor efficiency in systems other than *E. coli*. That said, it is

possible that a more general application of CRISPR-Cas9 could be to increase the efficiency of targetron-mediated mutagenesis. Finally, recent work by Fu *et al* (2013) demonstrates extensive off-target mutagenesis by CRISPR-Cas9, often at efficiencies comparable to the degree of on-target mutagenesis.

We have demonstrated the utility of targetron-delivered *lox* sites by deleting up to 120 kb of the *E. coli* genome and 15 kb of the *S. aureus* genome, inverting up to 1.5 megabases (one-third) of both the *E. coli* and *B. subtilis* genomes, and stably translocating 121 kb of the *E. coli* genome to another locus 1.5 megabases away. Efficiencies of the Cre-mediated recombinations are typically near 100%. This method compares favorably with another recently reported method for using targetrons to make genomic deletions (Jia *et al*, 2011) that relied on homologous recombination between introns and reported an efficiency of 2/648 for the deletion of a two-gene operon, requiring seven rounds of growth and transfer to new media.

The use of the Cre/*lox* system allows large pieces of foreign DNA to be integrated into genomes at high efficiency. An initial recombination occurs between a *lox* site on the plasmid and a *lox* site in the genome, serving to integrate the entire plasmid into the genome, and a second recombination event then occurs between the other two *lox* sites and removes the plasmid sequence. We found no evidence for a difference in efficiency between inserting 1 and 13 kb into the *E. coli* genome via RMCE. Given the high efficiency observed during the construction of large deletions and inversions, the limiting factor in RMCE would thus seem to be the initial encounter between the plasmid and the genome, and not the size of the insertion. The speed and efficiency of the second recombination event is presumed to be rapid and essentially 100% efficient, similar to the other intragenomic recombinations we report.

While large-scale inversions were presented here primarily as a demonstration of the lack of size limits for generating rearrangements using our method, artificial inversions have traditionally been used for studying genome structure and its constraints (Hill and Gray, 1988; Rebollo *et al*, 1988; Segall *et al*, 1988; Guijo *et al*, 2001; Campo *et al*, 2004; Garcia-Russell *et al*, 2004; Valens *et al*, 2004; Esnault *et al*, 2007), and the approaches presented herein allow such studies to be more easily performed in many more systems.

The one-step cut-and-paste method we present is of particular interest given that it allows one piece of a genome to be inserted within another site, without the accumulation of intervening intermediates, an operation that is not possible with any other technique. The cut-and-paste method could also be applied to more nuanced studies of genome structure constraints. For instance, the effect of moving different structural domains or of swapping two domains, such as the Ori and Ter domains, could be examined. Additionally, expression levels tend to be dependent on genomic location, with, for instance, genes nearer the origin tending to be more highly expressed (Cooper and Helmstetter, 1968; Rocha, 2008); and thus, cut-and-pastes could be used as a simple means for modulating the overall expression levels of super-operons (Lathe *et al*, 2000; Rogozin *et al*, 2002) or other large genetic units. The ability to move DNA between species without regard for inherent similarities or phylogenetic

relationships opens up the possibility of using genomic editing for rapidly adapting bacterial genomes.

Targetron genomic engineering technology can be readily practiced by almost any laboratory. The algorithm for retargeting the targetrons is available online. The targeting sites in the intron can be changed via restriction cloning of a short fragment of DNA that can be created via two PCRs or synthesized in its entirety (see Materials and methods), followed by the typical time required for ligation, transformation, and sequence validation. Retargeting and the addition of *lox* sites can be performed for multiple introns in parallel. Following electroporation into the target strain, intron induction requires only 1 day, and plated induction colonies grow after 1 day. The method is similar in complexity to *lox*-site placements with lambda Red, but is an improvement on recombineering in that no selectable markers are required and it can be used in strains where lambda Red performs poorly. Similarly, Cre-mediated recombination requires 1 day for electroporation of the Cre-expressing plasmid (and, for RMCE insertions, the delivery plasmid), and 1–2 days for the cells to grow and for recombination to occur. Though we used plasmids (one plasmid carrying the targetron, one plasmid carrying the Cre gene, and, as necessary, a plasmid or other vector carrying DNA to be integrated, delivered by electroporation or conjugation) to deliver targetrons in the present study, phage, direct electroporation or other methods could potentially be used, as well.

The scars left by the GETR method and the possibility of unplanned homologous recombination between introns are potential drawbacks, but we have shown that these can be avoided by careful planning. If intron and *lox*-site orientations are designed so that inverted repeats form upon Cre-mediated recombination, then the repeats will be deleted by the cell, removing most of the intron scar. However, the fact that certain inversions into the *lacZ* locus were viable when a scar was present but not when the scar was removed indicates that such scars may serve as a buffer against deleterious genomic rearrangements. Unwanted homologous recombination between introns can be prevented by the use of non-homologous introns (EcI5 and Ll.LtrB), or by targetron-mediated disruption of the *recA* gene.

Removing the genome-modifying plasmids was also simple. Except in the case of *S. oneidensis*, which required the continued presence of the IEP to allow the intron to splice out from the rRNA genes, a significant fraction (at least 1/3) of colonies were found to have lost the intron-expressing plasmid after the induction process. The Cre-expressing plasmids employed all contained temperature-sensitive origins of replication, and the delivery plasmids for RMCE had the *sacB* gene for counter-selection on sucrose, allowing these plasmids to be easily removed, as well.

In summary, GETR is a new method for genome engineering that can be adapted for use in a variety of bacteria with minimal modifications and without a significant loss of functionality. Large, specific, and varied changes can be made with high efficiency. This approach presents certain advantages over recombineering, particularly when working in strains not closely related to *E. coli*, or when the use of selectable markers is impractical or undesirable. In the case of *Staphylococcus aureus* in particular, recent work has made it

possible to transform clinical strains (Corvaglia *et al*, 2010), opening the way to genome editing of otherwise drug-resistant bacteria to create vaccine strains. As concerns about increasing drug resistance of pathogenic bacteria continue to mount, such strains may prove to be a viable alternative to antibiotics. We also expect the system to be of general utility to synthetic biologists looking to engineer entire genomes, particularly those looking to work in systems other than *E. coli*.

Materials and methods

Details of plasmid construction can be found in Supplementary Methods, and lists of the introns, plasmids, strains, and oligomers used in the present work are in Supplementary Tables S2–S5.

Intron retargeting

Introns were designed as described elsewhere (Perutka *et al*, 2004; Zhuang *et al*, 2009). The algorithm is available at <http://www.targetrons.com>. Ll.LtrB-type introns were retargeted according to the Sigma-Aldrich User Guide for the TargeTron Gene Knockout System (http://www.sigmaaldrich.com/etc/medialib/docs/Sigma/General_Information/targetron-user-guide.Par.0001.File.tmp/targetron-user-guide.pdf), except that the primers were prepared differently to improve the yield of the PCR amplification. Specifically, 1 μ l each of 20- μ M solutions of the EBS2 and EBS2AS primers were diluted into 26 μ l of water. In all, 2 μ l of this mixture and 1.4 μ l each of 20- μ M solutions of the IBS and EBS1 primers were used in the PCR amplification. The rest of the protocol was not substantially different from the Sigma-Aldrich protocol. Alternatively, the entire retargeted *HindIII/BsrGI* fragment was ordered as a gBlock from IDT and cloned directly into the introns to be retargeted.

For the Ecl5 introns, two different PCR amplifications were first executed using the IBS1/2S and EBS2AS primers in one reaction, and the EBS1S and EBSR primers in the other. In these reactions, 2 μ l each of 10- μ M solutions of the two primers and at least 5 ng template (an Ecl5 intron having the proper base at the +1 position) were used in 50 μ l. The products were subjected to PCR clean-up, and then at least 5 ng of each was combined for use as the template of a second PCR amplification similar to the first except doubled to a total volume of 100 μ l, with 8 μ l of 10- μ M EBSR and 2 μ l of 10- μ M IBS1/2S as the primers. The product was subjected to PCR clean-up, digested with *AvaII* and *XbaI*, and ligated into the Ecl5 vector (having the proper base at the +1 position) cut with *AvaII* and *XbaI*.

Intron induction

In *E. coli* strains, cells transformed with the intron-expressing plasmid were grown overnight at 37°C in Luria-Bertani (LB) broth plus 34 μ g/ml chloramphenicol, diluted to an OD₆₀₀ of 0.05 in 5 ml of LB plus 34 μ g/ml chloramphenicol, and then grown for 1 h at 37°C. In all, 250 μ l of that culture was then inoculated into 5 ml of LB containing 200 μ M IPTG (no antibiotic) and grown for 20 min (for Ll.LtrB-type introns) or 3 h (for Ecl5-type introns) at 37°C. (Ecl5 introns can also be induced for the shorter time period, but efficiency is somewhat better using the longer period.) The cultures were then put on ice, and 50 μ l of a 100 \times dilution (for Ll.LtrB-type introns) or a 1000 \times dilution (for Ecl5-type introns) was then streaked on LB plates (non-selective) pre-warmed to 37°C. The plates were then incubated overnight, and intron integration was screened using colony PCR. A subset of positive colonies was then screened for loss of antibiotic resistance to indicate the absence of the intron-expressing plasmid.

Intron induction in *S. aureus* RN10628 and *B. subtilis* 168 was performed as described elsewhere (Yao *et al*, 2006). Tryptic soy broth (TSB) was used as the growth medium for *S. aureus*, and LB broth was used for *B. subtilis* (with 5 μ g/ml erythromycin) and *S. oneidensis* (with 50 μ g/ml kanamycin). The T5.rDNA.798s.1WL2R intron was not formally induced.

Induction of Cre-mediated recombination

For intramolecular recombinations in *E. coli*, the plasmid pQL269 (Liu *et al*, 1998) was electroporated into cells that were then plated on LB plus 100 μ g/ml spectinomycin and grown at 30°C until colonies appeared. Occurrence of recombination was screened using colony PCR, and a subset of positive colonies were restreaked on LB (non-selective) and grown overnight at 42°C to cure the plasmid. Freezer stocks were made from these cells, and the analyses shown in Figure 6 through 8 and Supplementary Figure S3 were performed on cells streaked from these stocks. The procedure was essentially the same in *S. aureus* RN10628, except that the cells were electroporated with pRAB1 (Leibig *et al*, 2008) and grown initially on tryptic soy agar (TSA) plus 10 μ g/ml chloramphenicol. *B. subtilis* 168 was electroporated with pCrePA and grown on LB plus 5 μ g/ml erythromycin.

Cre-mediated genomic insertion (RMCE)

To assay insertion efficiency, delivery plasmids pACDX3S-GFP and pUC19X3S-GFP were used. These plasmids contain the GFP ORF flanked by T7 terminators and the *lox71* and *loxm2/66* sites. Each of the GFP delivery plasmids was transformed into *E. coli* BL21(DE3) Gold and *E. coli* HMS174(DE3) having a T7 promoter, as well as *lox66* and *loxm2/71* sites complementary to those in the delivery plasmids, integrated at the *lacZ*, *galK*, or *malT* locus. The Cre-expressing plasmid pQL269 was transformed into the strains, which were then grown at 30°C in liquid culture. At days 1, 2, and 3 after transformation of pQL269, aliquots from each culture were spread on LB plates. The plates were grown overnight at 37°C, and then imaged using a UV backlight and a B&W 061 dark-green filter. Identical strains lacking pQL269 were used as negative controls at each time point. Green colonies were counted manually to determine the insertion efficiency. The entire 3-day procedure was performed three times separately.

The insertion of DEBS1-TE was performed similarly, using pET26b-DEBS1TE-i as the delivery plasmid. Insertion was assayed by colony PCR using primers flanking the 5' end of the insertion 3 days after transformation of pQL269. Selected positive clones were then further assayed by overlapping PCRs covering the entire operon after removal of the delivery plasmid.

Doubling time measurements

Overnight cultures of the strains to be measured were diluted in LB to an OD₆₀₀ of 0.001, and triplicates of 500 μ l of that culture were placed in a 96-well plate (Nunc). All other wells (including all wells on edges) were filled with 500 μ l of sterile media (LB). Growth was measured using a plate reader (Bio-Tek PowerWave 340), pre-heated to and maintained at 37°C with a shaking intensity of 4 for 540 s at a time, with measurements taken every 560 s.

The results were plotted as log₂(OD₆₀₀) versus time (min). To select the linear region of the curve, each point was assigned a correlation coefficient R^2 corresponding to the value of R^2 for the line consisting of that point and the three points before and after. Since variance was lower when the same time window was used for all three replicates, the resulting R^2 values were averaged for all three replicates at each time point. The longest stretch in which all these averaged R^2 values were ≥ 0.99 was taken as the linear range. The slope of the least-squares linear fit of each replicate in that time range was then taken as the doubling time.

Data availability

GenBank accession numbers for the plasmids pX10, pX11, pX20, and pX21, are KF155402, KF155403, KF155404, and KF155405, respectively.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank S Elledge (Harvard University) for the gift of plasmid pQL269, R Bertram (University of Tübingen) for the gift of plasmid pRAB1, G Christie (Virginia Commonwealth University) for the gift of strain RN10628, S Leppa (National Institutes of Health) for the gift of the plasmid pCrePA, and Andrew Harper (University of Texas at Austin) for constructing plasmid pET26b-DEBS1TE. This material is based upon work supported in part by the National Science Foundation Graduate Research Fellowship under Grant No DGE-1110007, the National Security Science and Engineering Faculty Fellowship (FA9550-10-1-0169), and the Welch Foundation (F-1654). Work in AML's laboratory was supported by NIH grant GM37949 and Welch Foundation grant F-1607.

Author contributions: SMC and PJE performed the GFP and DEBS-TE insertions, which were planned by PJE, SMC, ATK-C, and ADE and analyzed by SMC and PJE. JP designed the introns and planned the *S. aureus* deletion. MND constructed most of the introns, and designed, performed, and analyzed the simultaneous triple deletion. JY, JTW, and AML designed and tested the *B. subtilis* introns, and EMQ designed and tested the *S. oneidensis* rDNA intron. PJE and ADE designed the rest of the experiments, and PJE performed and analyzed all other experiments. All authors participated in writing, revising, and editing the paper.

Conflict of interest

Targetron technology is subject to issued US and foreign patents and patent applications that are licensed by the Ohio State University and the University of Texas to InGex, LLC, which sublicenses the technology to others for commercial applications. JP, 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. JP and AML may receive royalty payments for commercial use of the technology. JP is the founder of a company, Targetronics, which sublicenses targetron technology from InGex, LLC.

References

Akhtar P, Khan SA (2012) Two independent replicons can support replication of the anthrax toxin-encoding plasmid pXO1 of *Bacillus anthracis*. *Plasmid* **67**: 111–117

Albert H, Dale EC, Lee E, Ow DW (1995) Site-specific integration of DNA into wild-type and mutant *lox* sites placed in the plant genome. *Plant J* **7**: 649–659

Alonzo 3rd F, Port GC, Cao M, Freitag NE (2009) The posttranslocation chaperone PrsA2 contributes to multiple facets of *Listeria monocytogenes* pathogenesis. *Infect Immun* **77**: 2612–2623

Aravind L, Koonin EV (2001) Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res* **11**: 1365–1374

Beloin C, Deighan P, Doyle M, Dorman CJ (2003) *Shigella flexneri* 2a strain 2457T expresses three members of the H-NS-like protein family: characterization of the Sfh protein. *Mol Genet Genomics* **270**: 66–77

Bikard D, Hatoum-Aslan A, Mucida D, Marraffini LA (2012) CRISPR interference can prevent natural transformation and virulence acquisition during *in vivo* bacterial infection. *Cell Host Microbe* **12**: 177–186

Biswas I, Gruss A, Ehrlich SD, Maguin E (1993) High-efficiency gene inactivation and replacement system for gram-positive bacteria. *J Bacteriol* **175**: 3628–3635

Campo N, Dias MJ, Daveran-Mingot ML, Ritzenthaler P, Le Bourgeois P (2004) Chromosomal constraints in Gram-positive bacteria revealed by artificial inversions. *Mol Microbiol* **51**: 511–522

Chambers SP, Prior SE, Barstow DA, Minton NP (1988) The pMTL nic- cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. *Gene* **68**: 139–149

Chang S, Cohen SN (1977) *In vivo* site-specific genetic recombination promoted by the EcoRI restriction endonuclease. *Proc Natl Acad Sci USA* **74**: 4811–4815

Cheng C, Nair AD, Indukuri VV, Gong S, Felsheim RF, Jaworski D, Munderloh UG, Ganta RR (2013) Targeted and Random Mutagenesis of *Ehrlichia chaffeensis* for the Identification of Genes Required for *In vivo* Infection. *PLoS Pathog* **9**: e1003171

Cho SW, Kim S, Kim JM, Kim JS (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31**: 230–232

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819–823

Cooper S, Helmstetter CE (1968) Chromosome replication and the division cycle of *Escherichia coli* B/r. *J Mol Biol* **31**: 519–540

Corvaglia AR, Francois P, Hernandez D, Perron K, Linder P, Schrenzel J (2010) 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* **107**: 11954–11958

Costantino N, Court DL (2003) Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc Natl Acad Sci USA* **100**: 15748–15753

Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* **97**: 6640–6645

Datta S, Costantino N, Zhou X, Court DL (2008) Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc Natl Acad Sci USA* **105**: 1626–1631

Derbise A, Lesic B, Dacheux D, Ghigo JM, Carniel E (2003) A rapid and simple method for inactivating chromosomal genes in *Yersinia*. *FEMS Immunol Med Microbiol* **38**: 113–116

Dymond JS, Richardson SM, Coombes CE, Babatz T, Muller H, Annaluru N, Blake WJ, Schwerzmann JW, Dai J, Lindstrom DL, Boeke AC, Gottschling DE, Chandrasegaran S, Bader JS, Boeke JD (2011) Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**: 471–476

Earl AM, Losick R, Kolter R (2008) Ecology and genomics of *Bacillus subtilis*. *Trends Microbiol* **16**: 269–275

Enyeart PJ, Ellington AD (2011) Synthetic biology: a yeast for all reasons. *Nature* **477**: 413–414

Esnault E, Valens M, Espeli O, Boccard F (2007) Chromosome structuring limits genome plasticity in *Escherichia coli*. *PLoS Genet* **3**: e226

Fredrickson JK, Romine MF, Beliaev AS, Auchtung JM, Driscoll ME, Gardner TS, Nealon KH, Osterman AL, Pinchuk G, Reed JL, Rodionov DA, Rodrigues JL, Saffarini DA, Serres MH, Spormann AM, Zhulin IB, Tiedje JM (2008) Towards environmental systems biology of *Shewanella*. *Nat Rev Microbiol* **6**: 592–603

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*

Fukiya S, Mizoguchi H, Mori H (2004) An improved method for deleting large regions of *Escherichia coli* K-12 chromosome using a combination of Cre/*loxP* and lambda Red. *FEMS Microbiol Lett* **234**: 325–331

Garcia-Russell N, Harmon TG, Le TQ, Amaladas NH, Mathewson RD, Segall AM (2004) Unequal access of chromosomal regions to each other in *Salmonella*: probing chromosome structure with phage lambda integrase-mediated long-range rearrangements. *Mol Microbiol* **52**: 329–344

Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA, Merryman C, Young L, Noskov VN, Glass JI, Venter JC, Hutchison 3rd CA, Smith HO (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**: 1215–1220

Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C,

- Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH *et al* (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**: 52–56
- Green MR, Sambrook J (2012) *Molecular Cloning: A Laboratory Manual*. 4th edn. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y.
- Guijo MI, Patte J, del Mar Campos M, Louarn JM, Rebollo JE (2001) Localized remodeling of the *Escherichia coli* chromosome: the patchwork of segments refractory and tolerant to inversion near the replication terminus. *Genetics* **157**: 1413–1423
- Hamilton CM, Aldea M, Washburn BK, Babitzke P, Kushner SR (1989) New method for generating deletions and gene replacements in *Escherichia coli*. *J Bacteriol* **171**: 4617–4622
- Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP (2007) The Clostron: a universal gene knock-out system for the genus *Clostridium*. *J Microbiol Methods* **70**: 452–464
- Hefferin ML, Tomkinson AE (2005) Mechanism of DNA double-strand break repair by non-homologous end joining. *DNA Repair (Amst)* **4**: 639–648
- Higgins D, Dworkin J (2012) Recent progress in *Bacillus subtilis* sporulation. *FEMS Microbiol Rev* **36**: 131–148
- Hill CW, Gray JA (1988) Effects of chromosomal inversion on cell fitness in *Escherichia coli* K-12. *Genetics* **119**: 771–778
- Hoess RH, Abremski K (1984) Interaction of the bacteriophage P1 recombinase Cre with the recombining site loxP. *Proc Natl Acad Sci USA* **81**: 1026–1029
- Hughes KT, Maloy SR (2007) *Advanced Bacterial Genetics: Use of Transposons and Phage for Genomic Engineering* Academic Press, San Diego, CA
- Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* **31**: 227–229
- Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppas NB, Emig CJ, Bang D, Hwang SJ, Jewett MC, Jacobson JM, Church GM (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**: 348–353
- Jia B, Yang JK, Liu WS, Li X, Yan YJ (2010) Homologous overexpression of a lipase from *Burkholderia cepacia* using the lambda Red recombinase system. *Biotechnol Lett* **32**: 521–526
- Jia K, Zhu Y, Zhang Y, Li Y (2011) Group II intron-anchored gene deletion in *Clostridium*. *PLoS One* **6**: e16693
- Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* **31**: 233–239
- Kaern M, Blake WJ, Collins JJ (2003) The engineering of gene regulatory networks. *Annu Rev Biomed Eng* **5**: 179–206
- Katashkina JI, Hara Y, Golubeva LI, Andreeva IG, Kuvaeva TM, Mashko SV (2009) Use of the lambda Red-recombineering method for genetic engineering of *Pantoea ananatis*. *BMC Mol Biol* **10**: 34
- Kodumal SJ, Patel KG, Reid R, Menzella HG, Welch M, Santi DV (2004) Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc Natl Acad Sci USA* **101**: 15573–15578
- Kolisnychenko V, Plunkett 3rd G, Herring CD, Feher T, Posfai J, Blattner FR, Posfai G (2002) Engineering a reduced *Escherichia coli* genome. *Genome Res* **12**: 640–647
- Kolter R, Inuzuka M, Helinski DR (1978) Trans-complementation-dependent replication of a low molecular weight origin fragment from plasmid R6K. *Cell* **15**: 1199–1208
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* **19**: 1639–1645
- Kumar S, Smith KP, Floyd JL, Varela MF (2011) Cloning and molecular analysis of a mannitol operon of phosphoenolpyruvate-dependent phosphotransferase (PTS) type from *Vibrio cholerae* O395. *Arch Microbiol* **193**: 201–208
- Lambowitz AM, Zimmerly S (2004) Mobile group II introns. *Annu Rev Genet* **38**: 1–35
- Langer SJ, Ghafoori AP, Byrd M, Leinwand L (2002) A genetic screen identifies novel non-compatible loxP sites. *Nucleic Acids Res* **30**: 3067–3077
- Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, Hutchison 3rd CA, Smith HO, Venter JC (2007) Genome transplantation in bacteria: changing one species to another. *Science* **317**: 632–638
- Lartigue C, Vashee S, Algire MA, Chuang RY, Benders GA, Ma L, Noskov VN, Denisova EA, Gibson DG, Assad-Garcia N, Alperovich N, Thomas DW, Merryman C, Hutchison 3rd CA, Smith HO, Venter JC, Glass JI (2009) Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science* **325**: 1693–1696
- Lathe 3rd WC, Snel B, Bork P (2000) Gene context conservation of a higher order than operons. *Trends Biochem Sci* **25**: 474–479
- Leenhouts K, Buist G, Bolhuis A, ten Berge A, Kiel J, Mierau I, Dabrowska M, Venema G, Kok J (1996) A general system for generating unlabelled gene replacements in bacterial chromosomes. *Mol Gen Genet* **253**: 217–224
- Leibig M, Krismer B, Kolb M, Friede A, Gotz F, Bertram R (2008) Marker removal in staphylococci via Cre recombinase and different lox sites. *Appl Environ Microbiol* **74**: 1316–1323
- Lestic B, Rahme LG (2008) Use of the lambda Red recombinase system to rapidly generate mutants in *Pseudomonas aeruginosa*. *BMC Mol Biol* **9**: 20
- Link AJ, Phillips D, Church GM (1997) Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: application to open reading frame characterization. *J Bacteriol* **179**: 6228–6237
- Liu Q, Li MZ, Leibham D, Cortez D, Elledge SJ (1998) The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. *Curr Biol* **8**: 1300–1309
- Livet J, Weissman TA, Kang H, Draft RW, Lu J, Bennis RA, Sanes JR, Lichtman JW (2007) Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**: 56–62
- Lu TK, Khalil AS, Collins JJ (2009) Next-generation synthetic gene networks. *Nat Biotechnol* **27**: 1139–1150
- Luchansky JB, Benson AK, Atherly AG (1989) Construction, transfer and properties of a novel temperature-sensitive integrable plasmid for genomic analysis of *Staphylococcus aureus*. *Mol Microbiol* **3**: 65–78
- Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. *Science* **339**: 823–826
- Mehdi Q, Hockaday TD, Newlands E, el-Kabir DJ (1971) LATS actions of bacterial permeability and protein synthesis. *Proc R Soc Med* **64**: 1268–1269
- Miller JH (1991) *Bacterial Genetic Systems*. San Diego: Academic Press
- Miller VL, Mekalanos JJ (1988) A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires toxR. *J Bacteriol* **170**: 2575–2583
- Murli S, Kennedy J, Dayem LC, Carney JR, Kealey JT (2003) Metabolic engineering of *Escherichia coli* for improved 6-deoxyerythronolide B production. *J Ind Microbiol Biotechnol* **30**: 500–509
- Nagy A (2000) Cre recombinase: the universal reagent for genome tailoring. *Genesis* **26**: 99–109
- Otto M (2012) MRSA virulence and spread. *Cell Microbiol* **14**: 1513–1521
- Palonen E, Lindstrom M, Karttunen R, Somervuo P, Korkeala H (2011) Expression of signal transduction system encoding genes of *Yersinia pseudotuberculosis* IP32953 at 28 degrees C and 3 degrees C. *PLoS One* **6**: e25063
- Park JM, Jang YS, Kim TY, Lee SY (2010) 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* **309**: 193–200
- Perutka J, Wang W, Goerlitz D, Lambowitz AM (2004) Use of computer-designed group II introns to disrupt *Escherichia coli* DEXH/D-box protein and DNA helicase genes. *J Mol Biol* **336**: 421–439

- Pomerantsev AP, Sitaraman R, Galloway CR, Kivovich V, Leppla SH (2006) Genome engineering in *Bacillus anthracis* using Cre recombinase. *Infect Immun* **74**: 682–693
- Rebollo JE, Francois V, Louarn JM (1988) Detection and possible role of two large nondivisible zones on the *Escherichia coli* chromosome. *Proc Natl Acad Sci USA* **85**: 9391–9395
- Rocha EP (2008) The organization of the bacterial genome. *Annu Rev Genet* **42**: 211–233
- Rodriguez SA, Davis G, Klose KE (2009) Targeted gene disruption in *Francisella tularensis* by group II introns. *Methods* **49**: 270–274
- Rogozin IB, Makarova KS, Murvai J, Czabarka E, Wolf YI, Tatusov RL, Szekely LA, Koonin EV (2002) Connected gene neighborhoods in prokaryotic genomes. *Nucleic Acids Res* **30**: 2212–2223
- Rosenberg AH, Lade BN, Chui DS, Lin SW, Dunn JJ, Studier FW (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* **56**: 125–135
- Rossi MS, Paquelin A, Ghigo JM, Wandersman C (2003) Haemophore-mediated signal transduction across the bacterial cell envelope in *Serratia marcescens*: the inducer and the transported substrate are different molecules. *Mol Microbiol* **48**: 1467–1480
- Segall A, Mahan MJ, Roth JR (1988) Rearrangement of the bacterial chromosome: forbidden inversions. *Science* **241**: 1314–1318
- Siegel RW, Jain R, Bradbury A (2001) Using an in vivo phagemid system to identify non-compatible loxP sequences. *FEBS Lett* **505**: 467–473
- Smith CL, Weiss BL, Aksoy S, Runyen-Janecky LJ (2013) Characterization of the achromobactin iron acquisition operon in *Sodalis glossinidius*. *Appl Environ Microbiol* **79**: 2872–2881
- Steen JA, Harrison P, Seemann T, Wilkie I, Harper M, Adler B, Boyce JD (2010) Fis is essential for capsule production in *Pasteurella multocida* and regulates expression of other important virulence factors. *PLoS Pathog* **6**: e1000750
- Sternberg N, Hamilton D (1981) Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. *J Mol Biol* **150**: 467–486
- Swingle B, Bao Z, Markel E, Chambers A, Cartinhour S (2010) Recombineering using RecTE from *Pseudomonas syringae*. *Appl Environ Microbiol* **76**: 4960–4968
- Ubeda C, Olivarez NP, Barry P, Wang H, Kong X, Matthews A, Tallent SM, Christie GE, Novick RP (2009) Specificity of staphylococcal phage and SaPI DNA packaging as revealed by integrase and terminase mutations. *Mol Microbiol* **72**: 98–108
- Valens M, Penaud S, Rossignol M, Cornet F, Boccard F (2004) Macrodomain organization of the *Escherichia coli* chromosome. *EMBO J* **23**: 4330–4341
- van Kessel JC, Hatfull GF (2007) Recombineering in *Mycobacterium tuberculosis*. *Nat Methods* **4**: 147–152
- van Kessel JC, Hatfull GF (2008) Efficient point mutagenesis in mycobacteria using single-stranded DNA recombineering: characterization of antimycobacterial drug targets. *Mol Microbiol* **67**: 1094–1107
- van Pijkeren JP, Britton RA (2012) High efficiency recombineering in lactic acid bacteria. *Nucleic Acids Res* **40**: e76
- Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**: 894–898
- Wang Y, Weng J, Waseem R, Yin X, Zhang R, Shen Q (2012) *Bacillus subtilis* genome editing using ssDNA with short homology regions. *Nucleic Acids Res* **40**: e91
- Whitt JT (2011) *DNA Target Site Recognition by the LL.LtrB Group II Intron RNP* PhD Thesis. University of Texas at Austin
- Wiesmann KE, Cortes J, Brown MJ, Cutter AL, Staunton J, Leadlay PF (1995) Polyketide synthesis *in vitro* on a modular polyketide synthase. *Chem Biol* **2**: 583–589
- Yao J (2008) *Bacterial Gene Targeting Using Group II Intron LL.LtrB and the Role of Bacterial Host Factors in LL.LtrB Splicing and Retrohoming* PhD Thesis University of Texas at Austin
- Yao J, Zhong J, Fang Y, Geisinger E, Novick RP, Lambowitz AM (2006) Use of targetrons to disrupt essential and nonessential genes in *Staphylococcus aureus* reveals temperature sensitivity of LL.LtrB group II intron splicing. *RNA* **12**: 1271–1281
- Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci USA* **97**: 5978–5983
- Zarschler K, Janesch B, Zayni S, Schaffer C, Messner P (2009) Construction of a gene knockout system for application in *Paenibacillus alvei* CCM 2051T, exemplified by the S-layer glycan biosynthesis initiation enzyme WsfP. *Appl Environ Microbiol* **75**: 3077–3085
- Zhuang F, Karberg M, Perutka J, Lambowitz AM (2009) Ecl5, a group IIB intron with high retrohoming frequency: DNA target site recognition and use in gene targeting. *RNA* **15**: 432–449
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415



Molecular Systems Biology is an open-access journal published by the **European Molecular Biology Organization** and **Nature Publishing Group**. This work is licensed under a **Creative Commons Attribution 3.0 Unported Licence**. To view a copy of this licence visit <http://creativecommons.org/licenses/by/3.0/>.