3596–3609 *Nucleic Acids Research, 2012, Vol. 40, No. 8* *Published online 22 December 2011*
doi:10.1093/nar/gkr1198

Structuring the bacterial genome: Y1-transposases associated with REP-BIME sequences†

Bao Ton-Hoang*, Patricia Siguier, Yves Quentin, Séverine Onillon, Brigitte Marty, Gwennaele Fichant and Mick Chandler*

Laboratoire de Microbiologie et Génétique Moléculaires, Centre National de la Recherche Scientifique, 118, Route de Narbonne, 31062 Toulouse Cedex, France

Received September 26, 2011; Revised and Accepted November 16, 2011

ABSTRACT

REPs are highly repeated intergenic palindromic sequences often clustered into structures called BIMEs including two individual REPs separated by short linker of variable length. They play a variety of key roles in the cell. REPs also resemble the subterminal hairpins of the atypical IS*200/605* family of insertion sequences which encode Y1 transposases ($TnpA_{IS200/IS605}$). These belong to the HUH endonuclease family, carry a single catalytic tyrosine (Y) and promote single strand transposition. Recently, a new clade of Y1 transposases ($TnpA_{REP}$) was found associated with REP/BIME in structures called REPtrons. It has been suggested that $TnpA_{REP}$ is responsible for REP/BIME proliferation over genomes. We analysed and compared REP distribution and REPtron structure in numerous available *E. coli* and *Shigella* strains. Phylogenetic analysis clearly indicated that *tnp*A_{REP} was acquired early in the species radiation and was lost later in some strains. To understand REP/BIME behaviour within the host genome, we also studied *E. coli* K12 $TnpA_{REP}$ activity *in vitro* and demonstrated that it catalyses cleavage and recombination of BIMEs. While $TnpA_{REP}$ shared the same general organization and similar catalytic characteristics with $TnpA_{IS200/IS605}$ transposases, it exhibited distinct properties potentially important in the creation of BIME variability and in their amplification. $TnpA_{REP}$ may therefore be one of the first examples of transposase domestication in prokaryotes.

INTRODUCTION

Repeated extragenic palindrome (REP) or Palindromic unit (PU) sequences were identified nearly 30 years ago in the intergenic regions of enterobacterial genomes (1). They play a variety of key roles in the cell. They are involved in regulating gene expression (by functioning as transcription terminators, by stabilizing mRNA and by acting as topological insulators for transcription-induced positive supercoiling (2–5), and in structuring DNA (by binding proteins such as IHF, PolI and DNA gyrase) (6–9). They are also specific target sites for several bacterial insertion sequences (10–12).

REPs are between 20- and 40-nt long, often clustered in structures called bacterial interspersed mosaic element (BIMES) as two tandem inverted copies separated by linkers, and have now been identified in a large number of bacterial genera and species where they are often found in high copy number (12–18). There are about 600 copies in *Escherichia coli* representing ~1% of the genome (15,19) and over 1600 copies in *Stenotrophomonas maltophilia* (17). The ubiquitous nature of REPs and their multiplicity raises the important question of how they have expanded to populate their host genomes and have evolved their present multiple roles.

A clue to this may lie in members of a class of atypical bacterial insertion sequences (IS), the IS*200*/IS*605* family, whose ends strongly resemble REPs. These ISs carry REP-like subterminal hairpins or imperfect palindromes (IP) secondary structures which are recognized and bound by the IS-specific transposase. They use a transposase, TnpA, of the HUH endonuclease family with a single catalytic tyrosine (Y1) as an attacking nucleophile and transpose using obligatory single-stranded (ss) DNA intermediates (20–23). We (ISfinder: www-is.biotoul.fr) and others (24) have identified a group of proteins, $TnpA_{REP}$, closely related to IS*200*/IS*605* transposases associated with REP sequences but forming a distinct clade defining a separate Y1 family. $TnpA_{REP}$ occurs in a variety of bacterial species and genera and is always flanked by REP/BIME sequences. Its presence appeared to be correlated with an increased abundance of REPs in the corresponding genomes suggesting that $TnpA_{REP}$ may

*To whom correspondence should be addressed. Tel: +33561335882; Fax: +33561335886; Email: tonhoang@ibcg.biotoul.fr
Correspondence may also be addressed to Mick Chandler. Tel: +33561335858; Fax: +33561335886; Email: mike@ibcg.biotoul.fr
†We would like to dedicate this article to the late Maurice Hofnung.

be responsible for REP proliferation throughout their host genomes (24). The molecular mechanism generating these patterns is unknown.

In vitro and *in vivo* studies of two IS*200*/IS*605* family members, IS*608* and IS*Dra2*, have provided a detailed picture of their transposition. This family differs profoundly from 'classical' ISs: they do not include terminal inverted repeats (IRs) and do not generate direct flanking target repeats (DRs) on insertion. Cleavage occurs at some distance from the IPs (25,26) via a transient covalent 5′-phosphotyrosine linked intermediate with the substrate DNA, leaving a free 3′-OH group on the other side of the DNA break. DNA cleavage also requires a divalent metal ion coordinated by two histidine residues, constituting the HUH motif, together with a third residue located close to the catalytic tyrosine (23,27). Transposition is strand-specific: $TnpA_{IS608/ISDra2}$ recognises only the 'top' strand which undergoes strand cleavage and transfer to the target site. The 'bottom' strand is inactive. The cleavage site at both left and right ends is not recognized directly by TnpA but forms a set of hydrogen bonds with a short guide sequence located 5′ to the foot of the left and right subterminal IP (23,28,29). This recognition is essential for cleavage. Finally, excision and insertion occur preferentially at the lagging strand template in replication forks (30).

To address how BIMEs might invade and amplify within a genome, we first analysed BIME distribution and polymorphisms in the genomes of 44 assembled *E. coli* and *Shigella* strains. We also identified a single locus in a majority of the 110 available *E. coli* and *Shigella* genome sequences where a single $tnpA_{REP}$ gene is located. Phylogenetic analysis suggested that $tnpA_{REP}$ was acquired early in the radiation of the species into present-day strains. The gene is bordered by variable numbers of BIMEs in structures, similar to that of IS*200*/IS*605* family members, called REPtrons, However, REPtrons do not appear to transpose as a unit but the BIMEs themselves are likely to be mobile and may have spread in a two-step process: transposition/recombination, which generates the observed sequence diversity of BIMEs, followed by local amplification.

To determine whether $TnpA_{REP}$ might be involved in this process, we analysed its cleavage and recombination activity *in vitro*. While $TnpA_{REP}$ shares similar catalytic characteristics with $TnpA_{IS200/IS605}$ transposases, it exhibited distinct properties potentially important in the creation of BIME variability and in their amplification. In the light of these observations, we discuss the possible role of $TnpA_{REP}$ in generating variability and in proliferation of BIMEs throughout their host genomes.

To our knowledge, REP/BIME and $TnpA_{REP}$ probably represent the first example of bacterial transposable element domestication.

MATERIALS AND METHODS

Bioinformatic procedures

Transposase identification and analyses. The primary transposase sequence of representative elements was used as a query in a BLASTP (31) search among all complete and partial prokaryotic genome sequences available on the NCBI server. All apparently full-length transposases were retained. Recursive BLASTP searches were performed using the less conserved retrieved sequences, i.e. those with the lowest BLAST score. The procedure was terminated when the results converged to a final stable data set (no new transposase sequences were detected). BLASTP searches were performed on the NCBI BLAST online interface without the low complexity filter but with otherwise default parameters. Multiple alignments were carried out using either ClustalW (32) or MultAlin (33) and some displays were obtained using the Jalview alignment editor (34).

In a second step, we used the Markov Cluster Algorithm (MCL) (http://micans.org/mcl/) (35,36) to weigh relationships between protein clusters. An inflation factor (IF) of 1.2 was used and edges having BLASTP score values of <30 were filtered (score > 30).

REP identification. The GenBank files of the complete bacterial genomes used in this study were retrieved from the NCBI public repository (http://www.ncbi.nlm.nih.gov/genomes/ lproks.cgi).

REP identification was achieved with a combination of two methods: a method based on local alignments and a method based on sequence profiles.

For the first approach, consensus sequences were derived from Bachellier *et al*. (19) transformed to follow our sequence convention. They cover the three REP families y, z1 and z2 and each was used as a query for BLASTN (31) similarity searches on DNA sequences of the complete genomes. Since we are dealing with short and variable sequences, we set BLASTN parameters to permissive values to increase sensitivity (expectation value ≤ 10, word size = 4, reward for a nucleotide match = 1, penalty for a nucleotide mismatch = -1, cost to open a gap = -1). BLASTN is fast and selective but, since it produces a local alignment, the boundaries of predicted REPs can be shorter than expected. In addition, the observed nucleotide variation at each position is not included in the alignment scoring since it can decrease the sensitivity of the prediction. Thus, a second approach to predict motifs containing gaps was used based on the GLAM2 programme (37). Profiles for each REP family were built by applying the GLAM2 program on unambiguous full length REPs predicted by the previous BLASTN searches. The GLAM2SCAN program was further used to find occurrences of the GLAM2 motifs in target sequences.

To set the parameters for both approaches, we used the annotation of BIMEs in *E. coli* K12 (NC_000913.2) (available at: http:// www.pasteur.fr/recherche/unites/ pmtg/ repet/index.html) as a training set. The estimated number of identified REPs corresponds to the combined results of both approaches.

Plasmid construction and $TnpA_{REP}$ purification

Escherichia coli MG1655 $tnpA_{REP}$ was cloned with a 6-His extension under control of promoter p_{ara} in pBS176. Expression and purification were carried out in *E. coli*

K12 (Rosetta, DE3) (Novagen) on Ni-agarose (Qiagen) as described for $TnpA_{IS608}$ (25). Plasmid pBS180 and derivatives were constructed in several steps: the MG1655 REPtron region was first isolated by PCR directly from MG1655 genomic DNA and cloned into pBluescript, pSK. The $tnpA_{REP}$ gene was replaced by a Cm® cassette and the downstream BIMEs were then removed by iPCR. In pBS180mut, mutations of the conserved GTAG were introduced using the Quickchange Site-directed Mutagenesis Kit (Stratagene) and ssDNA was prepared using f1 helper phage as described by the supplier (Promega). Further details can be obtained on request.

Reactions *in vitro*

Oligonucleotides were 5′-end-labelled with [γ-^{33}P] ATP (Perkin Elmer) using T4 polynucleotide kinase (NEB Inc.) or, in experiments to identify a 5′-phosphotyrosine transposase-substrate intermediate, 3′-end-labelled with [α-^{32}P] dATP Cordycepin (Perkin Elmer) using Terminal Transferase (NEB Inc.). Labelled oligonucleotides were purified on a G25 column (GE Healthcare).

Double-stranded DNA was prepared by hybridization of complementary oligonucleotides. After 5-min denaturation at 95°C, the mixture was left to slowly cool to 25°C.

5′-Labelled oligonucleotide (0.02 µM) and unlabelled oligonucleotide (0.5 µM) were incubated with 2 and 4 µM $TnpA_{REP}$ (45 min, 37°C, final volume 10 µl) in 12.5 mM Tris (pH 7.5), 120 mM NaCl, 1 mM DTT, 20 µg/ml BSA, 0.5 µg of poly-dIdC and 7% glycerol in the presence or absence of 5 mM $MnCl_2$ or $MgCl_2$. Reactions were separated on an 8% native gel in TAE buffer, to detect retarded complexes, or on a 9% denaturing gel, to detect cleavage and recombination products, and analysed by phosphorimaging (Fuji). In reactions to detect covalent complex formation, 3′-labelled substrates were incubated with $TnpA_{REP}$ in the reaction mixture as described earlier and reactions were separated on 16% SDS–PAGE gel.

Cleavage sites were generally determined by comparing the band position in the sequencing gel with a sequence ladder. For certain small cleavage products, oligonucleotides of the presumed size and sequence were synthesized and used for comparison.

Primer extension

In vitro reaction mixtures were treated with Proteinase K, purified using the Promega PCR purification Kit and used as template for primer extension with 5′-end-labelled 'a' and 'b' primers with Taq polymerase (94°C 45 s, 52°C 45 s, 72°C 1 min) 35×.

a: GTAAAACGACGGCCAGT.
b: GCAGAACTGATCCGCTATGT.

RESULTS

REPs, BIMES and REPtrons: sequence, distribution and organization in *E. coli*

Figure 1 shows the sequence organization and previously defined nomenclature of *E. coli* REP and BIME elements

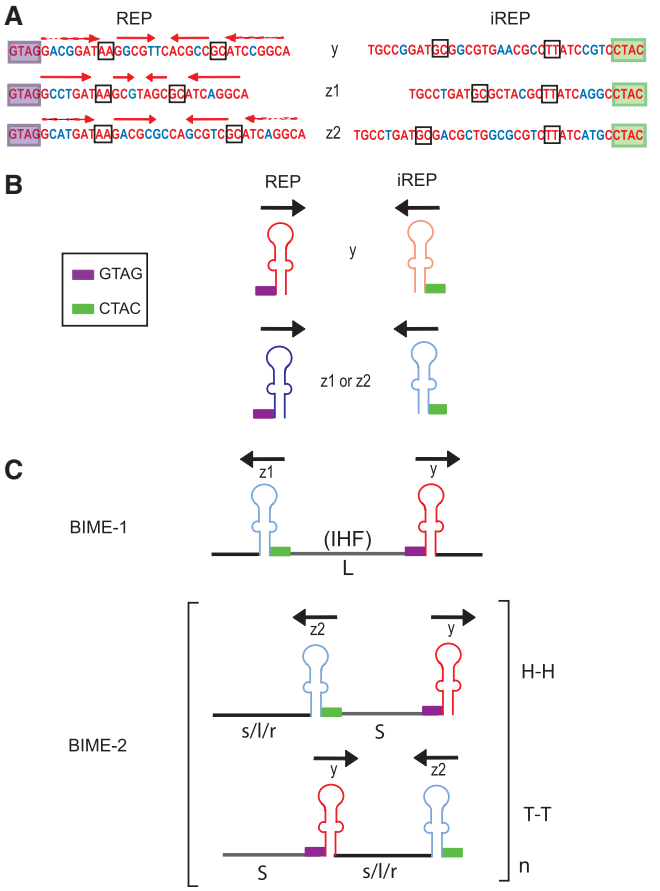

Figure 1. *Escherichia coli* REPs and BIMEs. (**A**) Consensus sequences of *E. coli* y, z1 and z2 REPs. The conserved tetranucleotide GTAG is boxed in violet, conserved positions are in red. Complementary sequences (iREP) corresponding to each category are presented on the right. The CTAC tetranucleotide, complementary to the conserved GTAG sequence, is boxed in green. Two base mismatches in the hairpin stem are boxed. The red horizontal arrows indicate complementary regions able to pair. Nucleotides in blue indicate bases that differ from one REP to another but nevertheless retain complementarity. (**B**) Structure of REP and iREP. Violet and green boxes represent the GTAG and CTAC, respectively. Black arrows indicate REP orientation. Red indicates a y REP and blue a z1 or z2 REP. Dark and light colours indicate REP and iREP, respectively. (**C**) Structures of BIME-1, and BIME-2. The reader is referred to Bachellier *et al.* (19). BIMEs are composed of a REP and an iREP separated by long (L) or short (S) linkers, H–H and T–T represent head-to-head and tail-to-tail configurations.

(19): they are 30- to 40-nt long and could fold into an imperfect palindrome (IP) with a highly conserved tetranucleotide, GTAG, localized 5′ to the IP foot (Figure 1A and B). There are three major types of *E. coli* REP sequence, y, z1 and z2 (Figure 1A). Only 84 REPs among 584 identified in *E. coli* K12 are single occurrences (19). Others are organized in pairs (BIME) including two REPs in inverse orientation separated by linker sequences (Figure 1B and C): one, in the orientation including the 5′ GTAG tetranucleotide, called REP, and a second inverted sequence called iREP (Figure 1). For functional reasons (see below), the sequence convention used here is inversed compared to Bachellier *et al.* (19). *Escherichia coli* BIMEs were classified into three families

(38). BIME-1 are composed of z1 and y (Figure 1) and occur as single copies in which the REP and iREP are separated by a long linker (L). BIME-2 are composed of z2 and y. They occur as multiple tandem copies with the REP and iREP components separated by a short linker, S, and with one of three types of flanking sequence, s, l or r. So-called atypical BIMEs are chimeras of BIME-1 and BIME-2, carrying different combinations of y, z1, z2, L, S, s, l and r. Like BIME-2, they also occur in multiple copies. The BIME-1 L linker is well conserved and frequently carries an IHF binding site (6) while those of BIME-2 and atypical BIMEs vary both in length and sequence. This can be seen among BIME-2 and atypical BIME copies carried by the MG1655 genome (Supplementary Figure S1). BIMEs also vary in number from locus to locus in a single strain. However, the sequence of tandem BIME copies at any one locus is well conserved (Supplementary Figure S2). Moreover, in different *E. coli* strains, the number of tandem BIMEs at a given locus is also variable [Supplementary Figure S3, see also (39)].

Identification of $tnpA_{REP}$ among members of the *E. coli*/*Shigella* group

We (ISfinder: www-is.biotoul.fr) and others (24) have independently identified a group of proteins ($TnpA_{REP}$), closely related to IS*200*/IS*605* transposases, associated with REP sequences. $TnpA_{REP}$ occurs in a variety of bacterial species and genera and is always flanked by REP/BIME sequences.

We analysed 110 *E. coli* and *Shigella* genomes available in the PATRIC database (40) for the presence of $tnpA_{REP}$ and focused on its immediately surrounding region (Supplementary Figure S4 and 'Discussion' section). Two-thirds (74/110) carried $tnpA_{REP}$ located at a unique position on the circular chromosome between *yaf*L and *fhiA*, even in strains (ATCC8739) (CP000946), DH1 (CP001637) and BL21 (AM946981) in which the entire region has undergone an inversion.

Figure 2A shows the distribution of REP and BIME elements in the $tnpA_{REP}$-carrying region from selected *E. coli* strains. While the left (5′) side invariably carried a single BIME, the right (3′) included a variable number of REPs and BIMEs (e.g. MG1655, APEC01, ED1a and O157:H7). In some strains, IS-mediated rearrangements had occurred (e.g. UMN026) (Figure 2A). Albeit more complex, these structures, called REPtrons, resemble members of the IS*200*/IS*605* family of bacterial insertion sequences. In 7 strains lacking $tnpA_{REP}$, all belonging to the B2 clade of the *E. coli*/*Shigella* group (41), the surrounding REP and BIME were still present but $tnpA_{REP}$ had been precisely excised (e.g. CFT073). In 24 strains, all except two belonging to the B1 clade (41), no trace of $tnpA_{REP}$ or associated BIMEs was found but, instead, these had acquired the toxin–antitoxin genes *hicA* and *hicB* between *yaf*L and *fhi*A (e.g. IAI1) (Figure 2A and Supplementary Figure S4). Mapping $tnpA_{REP}$ on the *E. coli*/*Shigella* phylogenetic tree (Supplementary Figure S4) suggested that the gene was acquired early in the species radiation, at least at the *E. fergusoni* and *E. coli*/*Shigella* separation, and was lost later in some strains by these two distinguishable events: either by replacement (together with its flanking BIMEs) with *hicA* and *hicB* or by precise deletion while retaining the flanking BIMEs.

We identified REP elements in 44 complete *E. coli* and *Shigella* genomes as well as in the available genome of *Escherichia fergusonii*. The estimated number of REPs in *E. coli* and *Shigella* genomes varied between 286 and 574 with an average number of 422 ± 74. This large dispersion reflects the presence of two subgroups of genomes showing extreme REP frequencies: the first group with a higher REP frequency (on an average 546) corresponds to strains from a same subtree including clade A; the second group, composed of two *E. coli* strains (SMS-3-5 and IAI39) from clade F and *Shigella dysenteriae* Sd197, displays less than 310 REPs. The other genomes have an estimated number of REPs correlated to the genome size and centred around 395 ± 32. This group includes strains with and without $tnpA_{REP}$. From these results, it is difficult to determine whether $tnpA_{REP}$ plays a role in REP amplification or maintenance or whether the loss of the gene is too recent to have had an observable effect on REP copy number. However, as only 216 REP elements have been identified in the *E. fergusonii* genome (which does not carry the REPtron), this distribution and the tree topology suggest that the large majority of REPs have arisen after the acquisition of the REPtron in the ancestor and before the divergence of *E. coli*/*Shigella* strains (Supplementary Figure S4).

$TnpA_{REP}$ and $TnpA_{IS200/605}$ form two different families

Figure 2B shows an alignment of a representative group of 10 $TnpA_{IS200/IS605}$ sequences (ISfinder) and a group of 10 $TnpA_{REP}$ sequences from the public databases. All retain a conserved tyrosine (Y) and the HUH amino acid triad (histidine–hydrophobic residue–histidine) typical of the Y1 transposase catalytic site. They also include a conserved asparagine (N), located four residues from the catalytic Y, replacing a glutamine (Q) residue of $TnpA_{IS200/IS605}$ involved in divalent metal ion coordination (27). However, $TnpA_{REP}$ group members are generally longer, include an additional C-terminal domain compared to $TnpA_{IS200/605}$ and exhibit specific conserved amino acid blocks throughout, in particular in the central and the C-terminal domains (24). MCL (Markov Clustering) analysis (35,36) of $TnpA_{IS200/605}$ and $TnpA_{REP}$ sequences also indicated that they represent two distinct Y1 families ('Materials and Methods' section). Clearly, $TnpA_{REP}$ has sequence features suggesting it may be involved in catalysis of REP invasion and dispersal within genomes.

$TnpA_{REP}$ activity *in vitro*

To gain insight into the potential role of $TnpA_{REP}$ in REP proliferation, *E. coli* $tnpA_{REP}$ was cloned with a His_6 carboxy-terminal tail under control of the p_{ara} promoter. The resulting protein, expressed from pBS176, was

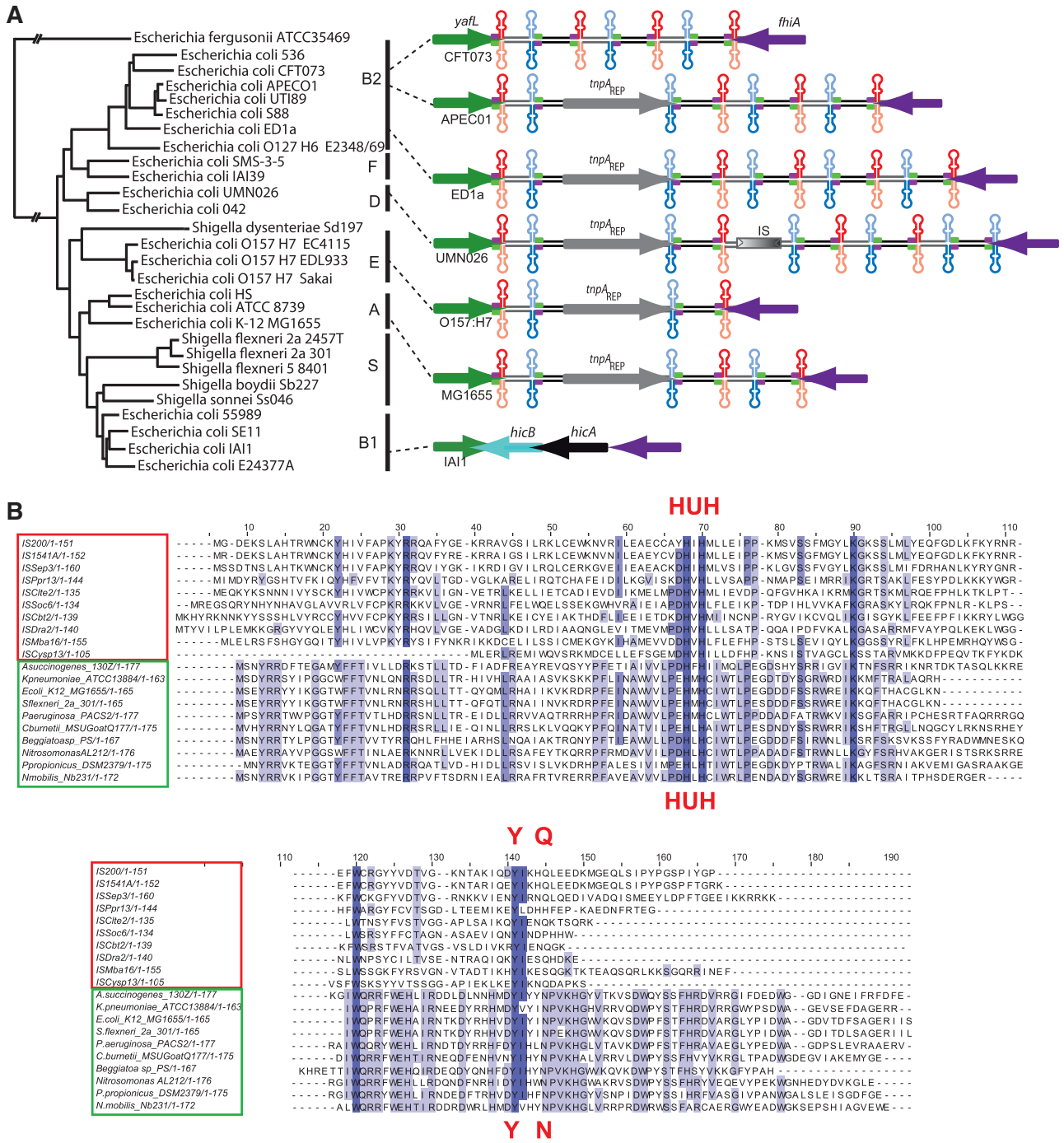

Figure 2. *Escherichia coli* phylogenetic tree, REPtron structure and $TnpA_{REP}$ alignment. (**A**) *E. coli* REPtron distribution and organization: the left of the figure represents, for clarity, a simplified phylogenetic tree of the *Escherichia coli* / *Shigella* group obtained by pruning the tree shown in Supplementary Figure S4 which was retrieved from the PATRIC database [http://www.patricbrc.org/portal/portal/patric/Home; (44)]. The clades (A, B1, B2, D, E, F and S) are from (41). The right of the figure shows examples of REPtrons from representative members of each clade. $tnpA_{REP}$ is shown in grey, the flanking genes *yafL* and *fhiA* in green and in violet, respectively. Arrows represent the direction of transcription. Flanking BIMEs are shown with the same convention as in Figure 1. The *hicA* and *hicB* genes are also indicated as black and blue arrows, respectively. (**B**) Alignment of $TnpA_{IS200/IS605}$ and $TnpA_{REP}$. TnpA from IS*200*/IS*605* family members are boxed in red while $TnpA_{REP}$ derivatives are boxed in green. Conserved positions are boxed as deep blue and less well-conserved residues in lighter shades of blue. The catalytic residues of $TnpA_{IS608}$ and the potential catalytic residues of *E. coli* MG1655 $TnpA_{REP}$ are shown in red above and below the alignment: histidine (H), hydrophobic (U) and tyrosine (Y) glutamine Q) and asparagine (N).

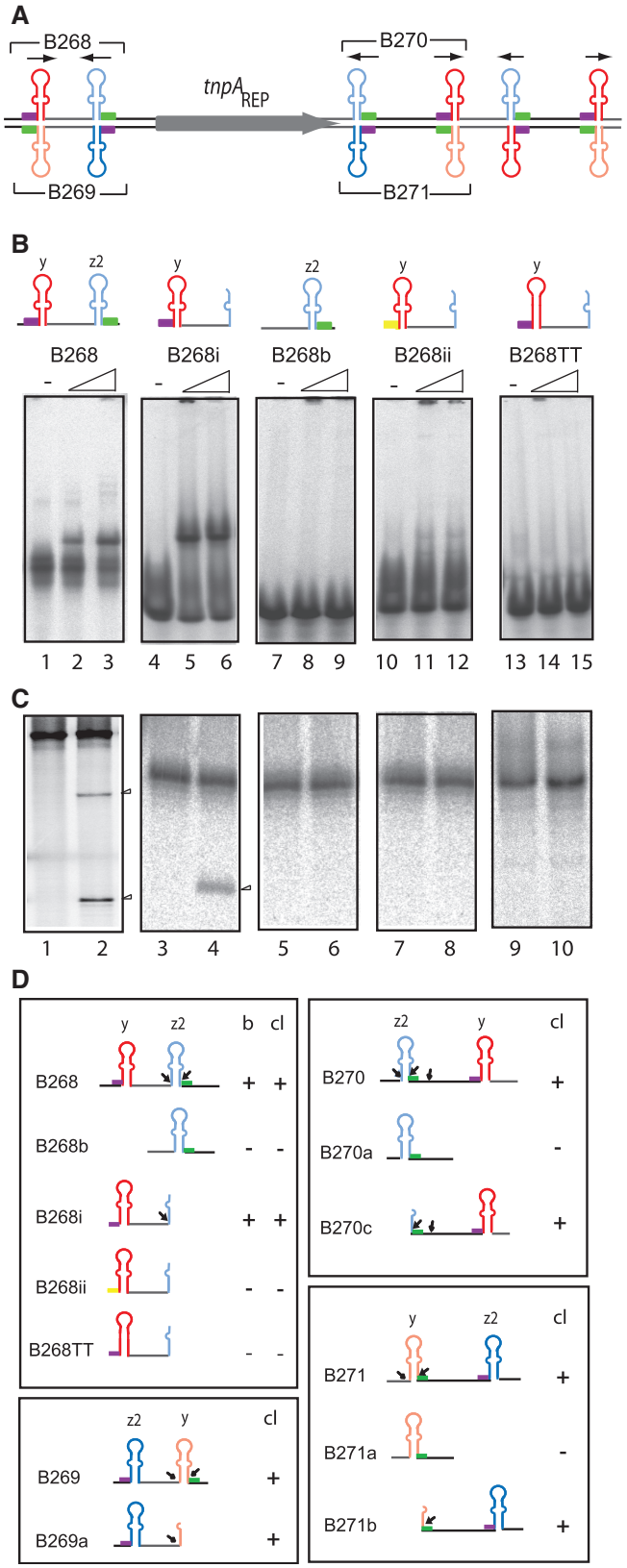

Figure 3. Binding and Cleavage activity on substrates derived from REPtron. (**A**) The *E. coli* MG1655 REPtron and oligonucleotides representing ssBIMEs used in this analysis. The oligonucleotides used are indicated with numbers above and below the cartoon. They are summarized in Supplementary Table S1. (**B**) Binding activity observed by EMSA. 5′-end-labelled oligonucleotides were incubated in the binding buffer ('Materials and Methods' section) in the absence or presence of 2 or 4 μM $TnpA_{REP}$ (shown as a triangle above the gels). '–' indicates no added $TnpA_{REP}$. The yellow box represents a GTAG tetranucleotide mutated to ACGA. (**C**) Cleavage of ssBIME derivatives. Arrow heads show the cleavage products. 5′-end-labelled oligonucleotides in the absence or presence of 4 μM $TnpA_{REP}$: B268 (116 nt), lanes 1–2; B268b (52 nt), lanes 3–4; B268i (61 nt), lanes 5–6; B268ii (61 nt, mutated for GTAG), lanes 7–8; B268TT (61 nt), lanes 9–10, respectively. (**D**) Structure and activity of other substrates derived from REPtron. In the right, binding and cleavage activity are summarized. (b = binding; cl = cleavage). Small black arrows indicate the position of cleavage sites.

purified on Ni-agarose resin ('Materials and Methods' section) and tested for binding and catalytic activities.

DNA binding. Initial DNA substrates were based on REPs and BIMEs (BIME II) from the REPtron present in MG1655 (Figure 3A and Supplementary Figure S5). Various 5′-end-labelled single- or double-strand substrates carrying REP or BIME sequences were incubated with purified $TnpA_{REP}$ in the absence of a divalent metal ion and analysed by EMSA (Figure 3B). No retarded band was observed with double-stranded substrates (Supplementary Figure S6A). However, B268, an ssBIME-carrying substrate from the 5′ REPtron end including a y REP and a z2 inverted REP sequence (iREP), showed a retarded band (Figure 3B, lanes 2 and 3) as did a substrate with half of the iREP (z2) (B268i; Figure 3B, lanes 5 and 6). Removal of the entire REP (y) eliminated binding (B268b, lanes 8 and 9). Mutation of the GTAG to ACGA (B268ii, lanes 11 and 12) or removal of the mismatch in the REP (y), in which mutation GC to TT should allow formation of a perfect REP palindrome sequence eliminated detectable $TnpA_{REP}$ binding (B268TT; lanes 14 and 15). Thus, both the GTAG tetranucleotide and the mismatch in the REP sequence are required for formation of robust $TnpA_{REP}$–DNA complexes.

Cleavage. To determine whether $TnpA_{REP}$ also has DNA cleavage activity, we incubated the 5′-end-labelled oligonucleotides used for EMSA studies with $TnpA_{REP}$ in reaction buffer containing Mn^{2+} or Mg^{2+} ('Materials and Methods' section). The products were separated in a denaturating sequencing gel. DsDNA was refractory to cleavage. $TnpA_{REP}$ was active only on ssDNA substrates and the reaction (using 3′-end-labelled substrates) generated a covalent DNA–protein intermediate as observed with $TnpA_{IS608}$ (Our unpublished data). Cleavage was generally more efficient with Mn^{2+} than with Mg^{2+} but no significant differences in cleavage specificity were observed (Supplementary Figure S6B). Except where stated, all assays presented here were performed with ssDNA substrates in the presence of Mn^{2+}.

B268, a 116-nt BIME-carrying substrate underwent two major cleavages at the 3′ z2 iREP to generate two labelled fragments of 85 and 55 nt (Figure 3C, lanes 1 and 2) whereas B268i, a 61-nt oligonucleotide carrying the entire y REP sequence but only part of the z2 iREP, shares the first cleavage site with B268, giving a 51-nt product (lanes 3 and 4). Even though the other B268 BIME derivatives showed no significant binding in

EMSA (Figure 3B), their activity was also examined since, in the case of $TnpA_{IS608}$, the fact that some DNA substrates do not form stable complexes visible by EMSA did not necessarily eliminate their capacity to undergo cleavage (18). B268b, a 52-nt substrate with only the z2 iREP was refractory to cleavage (lanes 5 and 6) as was B268ii, the 61-nt partial BIME carrying a mutated GTAG (Figure 3C, lanes 7 and 8). We confirmed this using additional GTAG mutants derived from B268 or other substrates. B268TT with a GC to TT mutation which allows formation of a perfect REP palindrome (Figure 3B, lanes 9 and 10) was also refractory to cleavage. We obtained a similar result with B268GC (mutation of AA to GC, not shown).

B269, an oligonucleotide complementary to B268 including a z2 REP and a y iREP sequence underwent two cleavages at the 3′ iREP (Figure 3D). B269a, an oligonucleotide derived from B269 carrying only part of the y iREP was also cleaved (Figure 3D). At first sight, this appears to contrast to the behaviour of the IS*200*/IS*605* family, where only the top strand is active (21,26). However, this is clearly due to different substrate configuration: the two component REP sequences in a BIME are inverted with respect to each other. Thus, each DNA strand carries a REP and an iREP, permitting cleavage on both strands.

We also examined the cleavage properties of the BIMEs located immediately 3′ of $tnpA_{REP}$. The results were similar to those obtained with the 5′ BIME. Cleavage occurred 5′ to the y REP on the top strand and to the z2 REP on the bottom strand (Figure 3D). Additional BIME variants from other *E. coli* chromosome regions showed similar behaviour (Supplementary Figure S5). $TnpA_{REP}$ catalysed cleavage of all three iREP variants, y, z1 or z2, and sometimes also cleaved sites within the linker sequence (Figure 3D, B270 and B270c) but only when the substrate also included a REP.

Thus, the data demonstrate that a REP structure is indispensable for BIME cleavage, presumably by providing a binding site for $TnpA_{REP}$. Moreover, they show that $TnpA_{REP}$ recognises the REP with its 5′ conserved GTAG tetranucleotide and requires the non-complementary base(s) in the stem for binding and for activity but cleaves at the inverted sequence of the 5′ or 3′ REP (iREP) and at the linker sequence with the expected polarity (each cleavage resulting in a 5′ phosphotyrosine intermediate; unpublished data). The results also demonstrate that cleavage can be either 5′ or 3′ to the essential REP sequence.

Defining the cleavage sites. Although REP and iREP portions of BIMEs are relatively well conserved, the linkers are highly variable (Supplementary Figure S1). Like REPtron-derived substrates (Figure 3D), several of these proved to contain cleavage sites in the linker region although the efficiency of cleavage at these sites is variable. Examples are B315 a partial BIME located at *araD-A* in MG1655 (one site, Figure 4A) and B319, located at *gltP-yjcO* (four sites, Figure 4A).

In summary, we observed two categories of cleavage site: those on either side of iREP and those in the linker regions. We refer to the first category as 'iREP' and to the second as 'linker' cleavage sites. The pattern of cleavage was similar in the presence of either Mn^{2+} or Mg^{2+} (Supplementary Figure S6).

Cleavage sites from each category obtained from a set of naturally occurring BIME sequences are aligned in Figure 4B. 'iREP' sites (Figure 4B left) are situated at both sides of the iREP palindrome in relatively conserved regions: the first type (I) occurred at $T_{-4}G_{-3}C_{-2}C_{-1}|T_1G_2A_3T_4/A_4$ (where '|' represents the site of cleavage) and the second (II) at $G_{-4}/T_{-4}G_{-3}/T_{-3}/A_{-3}C_{-2}\underline{C}_{-1}|\underline{T}_1\underline{A}_2\underline{C}_3A_4/C_4/G_4$, within CTAC, the complement of the conserved GTAG tetranucleotide. Note that type I sites are present in z1 and z2 iREPs but are absent in y iREP sequences.

The 'linker' sites (Figure 4B right) appear less conserved. The deduced consensus reveals conserved CC|T (coordinates $C_{-2}C_{-1}|T_1$) for 'iREP' and C|T ($C_{-1}|T_1$) sequences for 'linker' sites.

The importance of the C|T for cleavage. To understand the rules governing BIME processing in more detail, we analysed the cleavage pattern of a panel of mutant sites introduced at the iREPI site in the 5′ partial BIME (B268i) from the MG1655 REPtron (Figure 4C) and at the iREPII and linker sites in a second partial BIME (B315) (Figure 4D and E) present at the *araD-A* intergenic region. Mutations were introduced either in a block or individually in each site. Although this is not an exhaustive analysis, the results obtained show that the central C|T sequence was indispensable for cleavage as all substitutions at these positions prevented cleavage while mutations in some other positions were tolerated.

Although we have not systematically measured the efficiency of cleavage at each site, sequences flanking the CT dinucleotide may influence cleavage efficiency. For example, the weak site B319(1) includes GACC|TACA compared to the stronger B315(1) with GGCC|TACA: G_{-3} may therefore influence cleavage efficiency (Figure 4A and B). In addition, mutation of T_{-4} to any base (B268i, iREPI site, Figure 4C) reduced cleavage in B268ib, B268ic and B268id (Supplementary Figure S6).

Thus, 'iREP' and 'linker' sites appear to share similar sequence requirements indicating relatively limited cleavage specificity.

Cleavage of single strand circular DNA. To confirm the requirement of C|T for cleavage and to assess the distance over which $TnpA_{REP}$ might act, we examined cleavage of a significantly longer DNA substrate, a single strand DNA circle derived from the 4.1-kb bacteriophage f1-derived phasmid, pBluescript II SK, into which a BIME had been cloned (pBS180, 'Materials and Methods' section). Following cleavage with $TnpA_{REP}$, DNA was deproteinized and the cleavage sites were mapped by primer extension using two different primers ('Materials and Methods' section). The results are presented in Figure 5. In addition to bands resulting from two natural polymerase stalling sites which depend on the presence of a BIME (lanes 1 and 5), the substrate contains many cleavage sites stretching both upstream

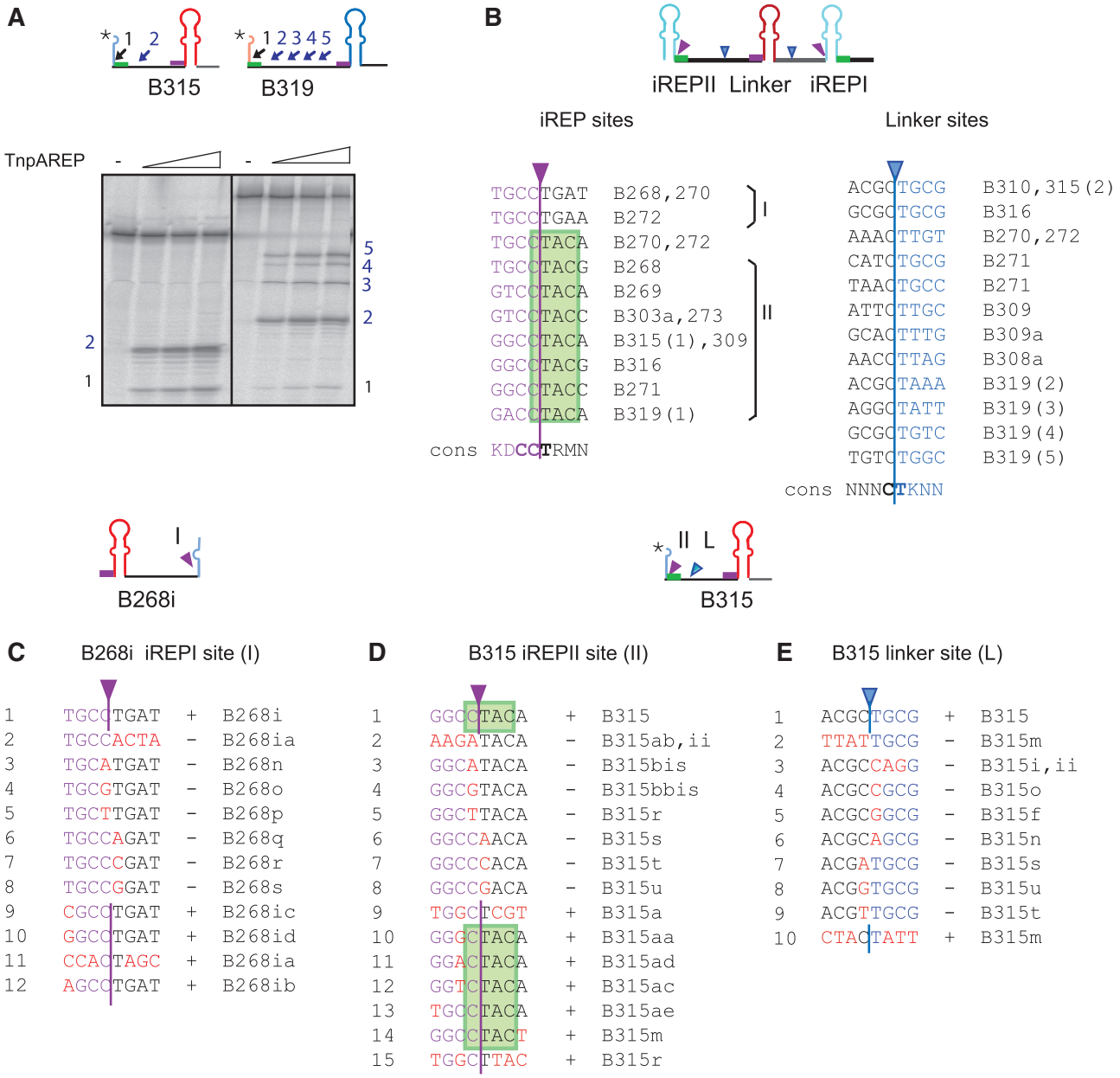

Figure 4. iREP and linker cleavage sites. (**A**) Examples of linker cleavage sites on B315 and B319. B315 and B319 are partial BIMEs derived from MG1655 *araD-A* and *gltP-yjcO* regions. Linker cleavage sites are indicated by blue small arrows and numbered. 'asterisks' indicates the position of labelling. (**B**) Alignment of observed *in vitro* 'iREP' and 'linker' sites. 'iREP' and 'linker' sites are indicated by purple or blue arrows. Their sequence is shown with the consensus sequence (cons) for each category below in bold. iREPI and iREPII sites occurred on both sides of the iREP palindrome, the second overlapped the CTAC tetranucleotide (in green), the complement of the conserved GTAG tetranucleotide. The oligonucleotides used are shown to the right of the sequences. (**C**) Characterization of 'iREPI' (I) site using B268i as an example. Wild-type bases to the left of the cleavage site are shown in purple. Those to the right are shown in black. Mutated bases are shown in red. '+' indicates cleavage, '−' indicates no cleavage. Nucleotide designations are shown to the right. (**D** and **E**) Characterization of 'iREPII' (II) and 'linker' (L) using B315. Colour codes are the same as for (C). [For the B315, II and L correspond to sites B315(1) and B315(2) in Figure 4A]. For the B315 linker site, nucleotides to the left of the cleavage site are shown in black and those to the right in blue. Mutated nucleotides are shown in red.

and downstream of the resident BIME over the entire region (>400 nt) analysed (lanes 6 and 8). Moreover, cleavage is absolutely dependent on the presence of the functional BIME since no products are observed with a substrate carrying a BIME mutated for the conserved GTAG (compare lanes 2 and 4). Mapping these sites on the DNA sequence indicated that they occurred at a C$^{\downarrow}$T dinucleotide.

Thus, BIME-directed cleavage can occur at a considerable distance from the $TnpA_{REP}$ binding site.

Recombination. A striking characteristic of BIME-2 and atypical BIMEs is the variation in the linker sequence and in copy number at a given locus in different strains suggesting that they may undergo recombination and amplification. To investigate this *in vitro*, we examined

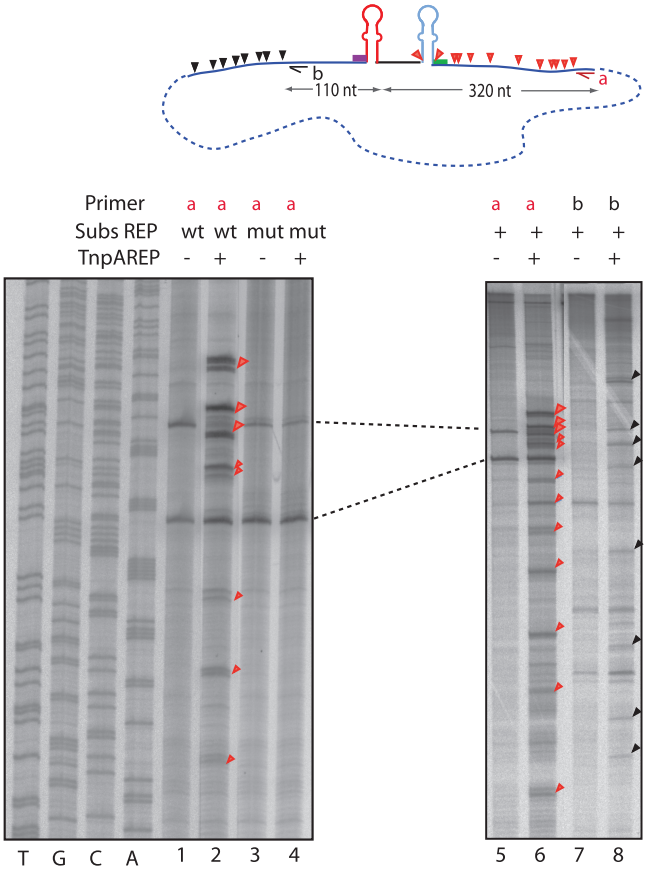

Figure 5. Cleavage of circular ssDNA. The substrate and primers used for analysis are shown. ssDNA circles derived from pBS180 (substrate with functional REP, lanes 1–2; 5–8) and pBS180mut (substrate with REP mutated for GTAG, lanes 3–4) were incubated in the absence or presence of $TnpA_{REP}$ in buffer containing $MnCl_2$ and used for primer extension with 5′-end labelled oligonucleotide 'a' (lanes 1–6) and 'b' (7–8). Lanes 1–2 and 5–6 correspond to the same samples separated under two different migration conditions on a 6% sequencing gel. Distances in nucleotides show the distance of the complementary oligonucleotide primer from the foot of the functional REP. Cleavages revealed by primers 'a' and 'b' are shown by red and black arrowheads, respectively.

the capacity of $TnpA_{REP}$ to promote BIME recombination with two sets of substrates. The first included ssBIMEs from the REPtron region (B268, B268i and B268ii; Figures 3 and 6A). In these experiments, we used a 20-fold molar excess of the unlabelled partner oligonucleotide to facilitate recombination. When 5′ labelled 116 nt B268 was incubated with $TnpA_{REP}$, it was cleaved at two 'iREP' sites generating 85 and 55 nt products (Figure 6A, lanes 1 and 2). Addition of unlabelled 61 nt B268i generated a 65-nt recombination product (lane 3). This species no longer appeared in the reaction with the inactive substrate B268ii mutated for GTAG (Figure 6A, lane 4). These results demonstrate an exchange of sequences between the partner DNA molecules demonstrating recombination between two 'iREP' sites.

The second substrate set included derivatives of an ssBIME from the *araD-A* region (B316, Figure 6B). 5′-end-labelled 88 nt B316 was cleaved at a 'iREP II' and a linker site, generating 76 and 69 nt products (Figure 6B, lanes 2–5). In the presence of unlabelled B316, a product of 95 nt resulting from recombination between the 'iREP' and the 'linker' sites was generated (Figure 6B, lane 2). We no longer observed this species on addition of unlabelled B316a, mutated for GTAG or of unlabelled B316b, mutated at the 'iREPII' site, respectively (Supplementary Table S1; Figure 6B, lanes 3 and 4). The recombination product was still observed with cold B316c, mutated at the 'linker' site but keeping the 'iREP' site intact (Figure 6B, lane 5). However, this recombination product was not generated when we used 5′-end-labelled B316c in the presence of unlabelled B316c (Figure 6B, lanes 6 and 7), confirming the nature of this reaction. We did not observe recombination products from certain reactions since they reconstitute a fragment having the length of the original labelled substrate (e.g. Figure 6B). We also used the 5′-end-labelled 88 nt B316 and oligonucleotide B315 (Figure 6C). *In vitro*, in addition to the two cleavage products of B316, these substrates generated a series of DNA species which migrated high in the gel and had sizes consistent with recombination products which include an additional REP together with various combinations of linker sequences.

These results demonstrate recombination between iREP and linker sites which might result in BIME variability, multiplication and amplification.

DISCUSSION

$tnpA_{REP}$, coding for a protein related to the Y1 transposases, was identified in association with REP/BIME sequences in structures called REPtrons found in a number of bacterial genomes Here we compared the REPtron structure and REP/BIME distribution in available *E. coli* and *Shigella* genomes. We also analysed *E. coli* K12 $TnpA_{REP}$ activity including cleavage and recombination *in vitro*. While $TnpA_{REP}$ shared the same general organization and similar catalytic characteristics as the $TnpA_{IS200/IS605}$ Y1 transposases, it exhibited distinct properties potentially important in creation of BIME variability and amplification. The presumed importance *of* $tnpA_{REP}$ in REP/BIME evolution and dispersion within genomes and multiple roles assumed by REPs and BIMEs themselves in cell physiology could be interpreted as domestication. Although many cases of domestication of eukaryotic transposable elements have been documented (42,43), such domestication has not yet been described for classical bacterial elements.

REPtron evolution in *E. coli* and *Shigella*

The $tnpA_{REP}$ gene was present in 74 of the 110 *E. coli* and *Shigella* genomes available in the PATRIC database (http://www.patricbrc.org/portal/portal/patric/Home); (40,44) always as a single copy at the same genetic locus. This genetic conservation implies that these sequences had been acquired in a single event early in the last common ancestor of the species which gave rise to present-day strains. The phylogenetic analysis of *E. coli*/*Shigella* strains lacking $tnpA_{REP}$ indicates that $tnpA_{REP}$ had been

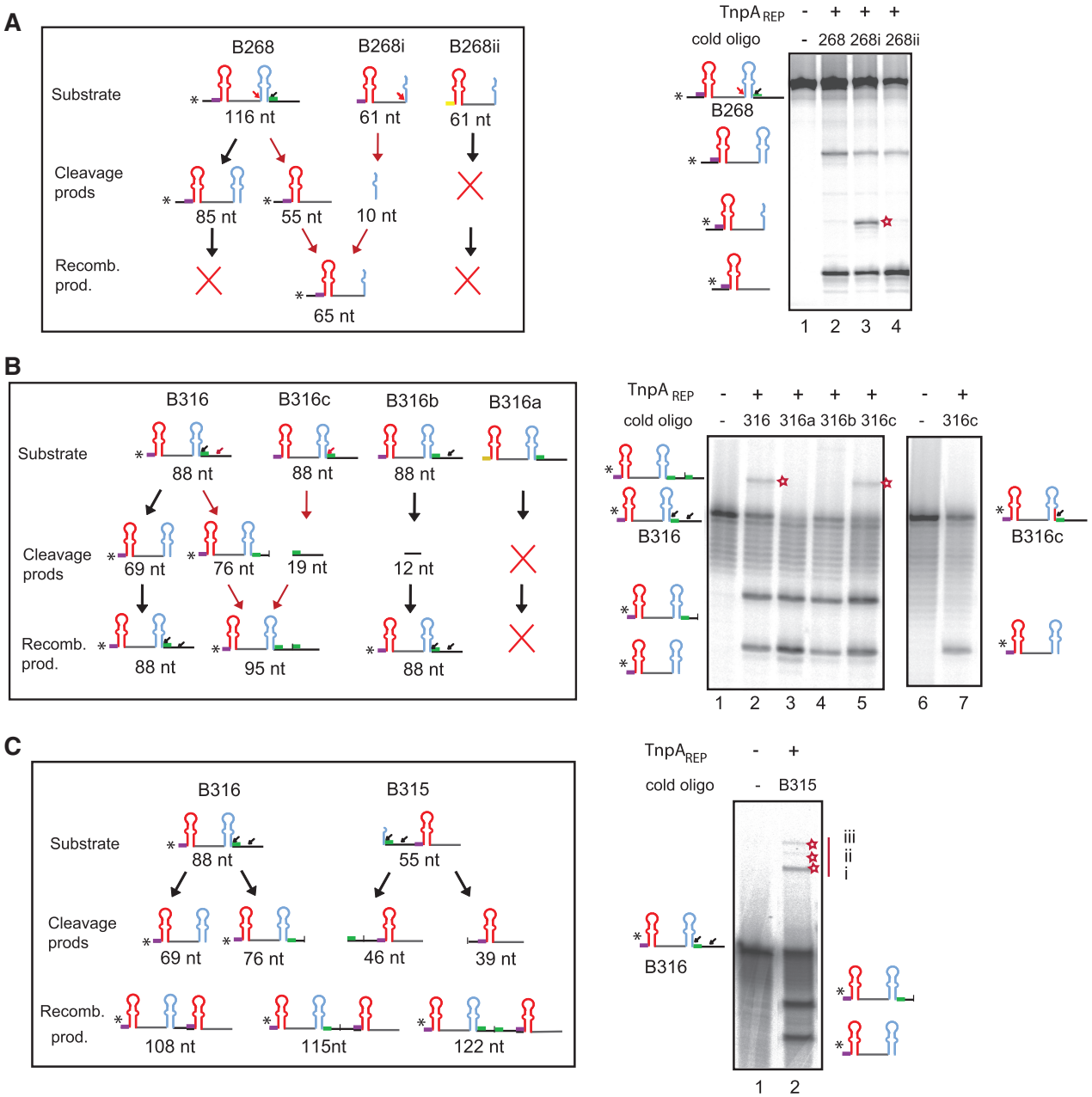

Figure 6. TnpA$_{REP}$ catalysed strand exchange *in vitro*. DNA substrates and recombination products for each reaction are indicated, '*' in the cartoons indicates radioisotope position; colour codes are as in Figure 1. Positions of recombination products are indicated by a red star, small red arrow indicate cleavage sites giving rise to detectable recombination. (**A**) Strand exchange between 'iREP' sites. 5′-end labelled oligonucleotide: B268 (116 nt): lane 1: no protein, lanes 2–4: in presence of TnpA$_{REP}$ and unlabelled 61 nt B268, B268i and B268ii (mutated at GTAG), respectively. (**B**) Recombination between 'iREP' and 'linker' sites. 5′-labelled oligonucleotide: B316 (88 nt) was incubated with: no protein (lane 1), in the presence of TnpA$_{REP}$ and unlabelled oligonucleotides: B316, B316a, B316b, B316c (lanes 2–5), respectively. In the reaction with 5′-labelled B316c: no protein or in the presence of protein and unlabelled B316c (lanes 6–7). (**C**) Recombination leading to assembly of multiple REPs. 5′-labelled B316 was incubated with: no protein (lane 1), in the presence of TnpA$_{REP}$ and unlabelled oligonucleotide B315 (lane 2). The products of strand transfer (marked with a red star) migrate as expected for the structures i, ii and iii shown in the left hand cartoon.

present but had subsequently been replaced or deleted (Supplementary Figure S4).

Although REPtron organization resembles that of members of the IS*200*/IS*605* family, we do not believe that it represents a true transposable genetic element. Its unique genetic location indicates that it has not undergone subsequent rounds of transposition. This suggests that if the spread of REPs within a genome is catalysed by TnpA$_{REP}$ it must occur by mobilization *in trans*.

We observed only minor differences in REPs/BIMEs copy number and distribution in strains with or without *tnpA*$_{REP}$ (Supplementary Figure S4), in agreement with the idea of REPtron/*tnpA*$_{REP}$ loss occurring late in the radiation of these strains. This is in contrast to the

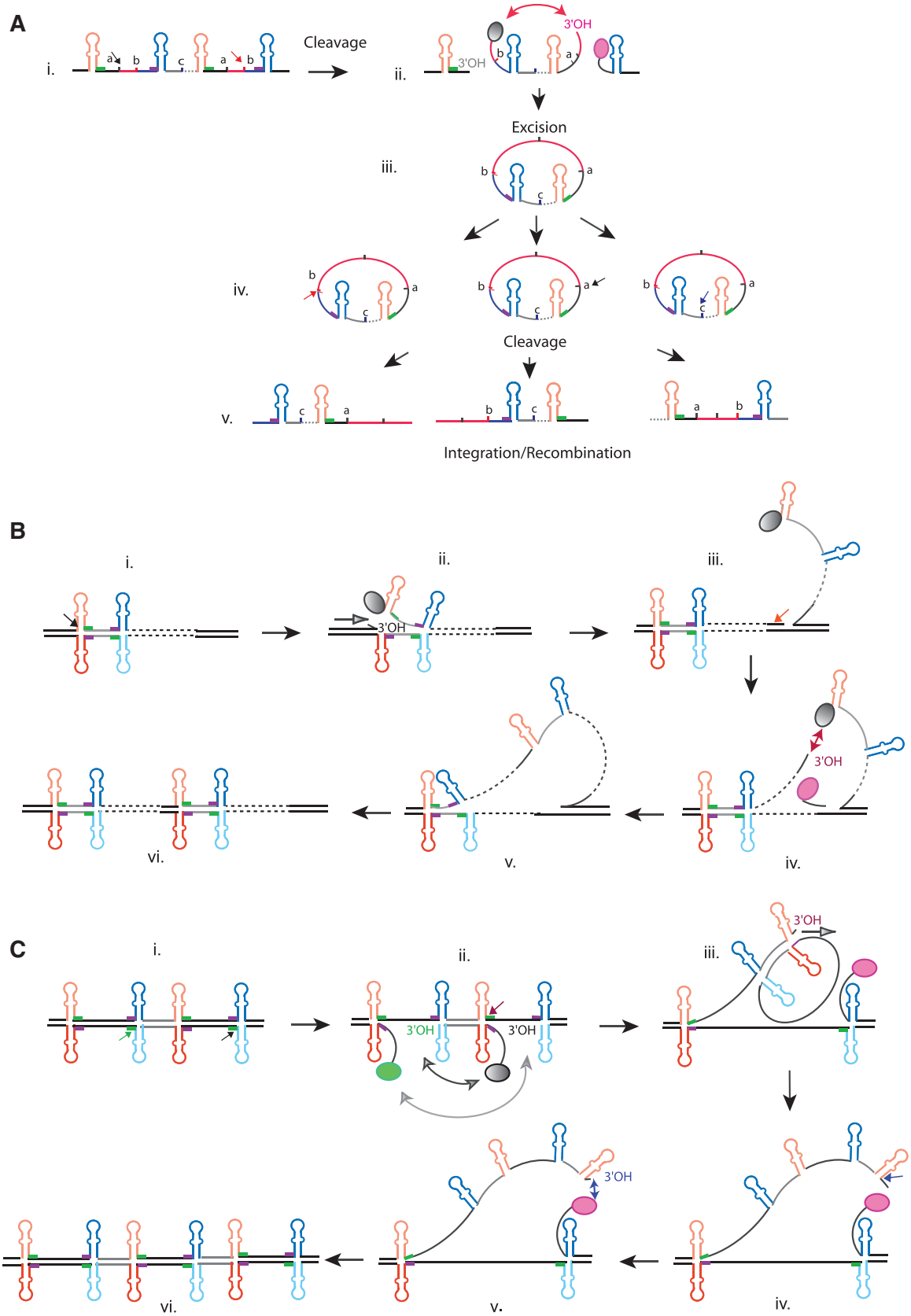

Figure 7. Model for BIME diversification and amplification. (**A**) BIME excision and integration model: From a BIME array carrying several 'linker' cleavage sites [A(i)], cleavages at two distinct sites ('iREP' or 'linker' types) and strand exchange between Tyr-5′P (in grey) generated from the first and the 3′-OH from the second (in red) would lead to excision of a ssBIME circle [A(ii,iii)]. Cleaved again (by $TnpA_{REP}$) at alternative cleavage sites and/or processed by unidentified host factors [A(iv)] before integration, this could give rise to several BIME variants [A(v)]. (**B**) BIME amplification

(continued)

previous observation in some other bacterial species of a strong correlation between the presence of $tnpA_{REP}$ and the increased number of REPs/BIMEs in corresponding genomes (24), and may reflect a more recent REPtron/$tnpA_{REP}$ acquisition event in these strains.

The REPtron distribution therefore clearly argues for an origin which predates the radiation of *E. coli*/*Shigella* and, since *E. alberti* appears to carry a similar (but not identical) REPtron, it may predate the separation of this species.

$TnpA_{REP}$ binding, cleavage and strand transfer activity *in vitro*

To evaluate whether $TnpA_{REP}$ might be involved in the recombination events leading to REP invasion and spread, we examined its activities *in vitro*. We demonstrated that it binds ssREP sequences and requires the conserved 5′ GATG tetranucleotide for this. It also catalyses BIME cleavage both upstream and downstream of the REP sequence. The data suggest that the functional unit on which the enzyme acts is a (complete or partial) BIME rather than an individual REP or iREP. Cleavage requires an entire REP sequence and occurs at the iREP and in BIME linkers. Since BIMEs are composed of two inverted REP copies, this implies that $TnpA_{REP}$ can cleave both DNA strands, contrary to $TnpA_{IS608/ISDra2}$ which functions in a strand specific manner. Moreover, the two base pair mismatch located in the REP stem is also essential for binding and activity.

Although $TnpA_{REP}$ forms a distinct clade within the Y1 transposase family, it appears to share with $TnpA_{IS608}$ the same absolute requirement for ss substrates (Supplementary Figure S6) and the formation of a covalent protein–DNA intermediate (25) (unpublished data). However, unlike $TnpA_{IS200/IS605}$, the sequence specificity for cleavage appears relatively low, requiring only the dinucleotide CT while tolerating substitutions in other surrounding positions. This limited cleavage specificity may be responsible for BIME diversification. Cleavage occurs in the presence of Mg^{2+} but is much more pronounced with Mn^{2+}. However, similar cleavage patterns are observed with both metal ions (Supplementary Figure S6).

We also observed strand transfer activity *in vitro*. Strand transfer can therefore create sequence variability and assemble tandem BIME copies resembling the tandem amplification observed *in vivo* (see below).

Model of BIME variability and amplification

In light of the observed distribution and sequence variability in the collection of *E. coli* strains, BIME colonization and expansion throughout genomes may occur as a two-step process involving diversification followed by amplification (Figure 7). There are several ways in which this might be accomplished.

Diversity could be generated in an excision and insertion process by using different cleavage sites for excision and for insertion as shown in Figure 7A. From a BIME array carrying several 'linker' cleavage sites [Figure 7A(i)], cleavages at two distinct sites ('iREP' or 'linker' types) and joining between Tyr-5′P generated from the first and the 3′-OH from the second would lead to excision of a ssBIME circle [Figure 7A(ii,iii)] in the same way as been shown for IS*608* (25). Clearly, like IS*608* and IS*Dra2*, this could involve ssDNA on the lagging strand template of a replication fork but might also use ssDNA generated during R-loop formation, triggered by transcription-induced negative supercoiling (45), repair or by supercoil-driven extrusion of the REP secondary structure element (46). If processed at different cleavage sites [Figure 7A(iv)] before integration, this intermediate could give rise to several BIME variants [Figure 7A(v)]. For BIMEs, this process can occur on both DNA strands which in principle could carry different 'linker' cleavage sites thus increasing the potential BIME sequence diversity. Degradation of 3′-OH end by host nucleases might also contribute to BIME variation. In this model, variation would be coupled to integration.

Amplification could occur following insertion of the excised BIME into a suitable target as that shown in Figure 7B. This uses a 'rolling circle' like recombination mechanism on an inserted BIME (either in the head to head or tail to tail orientation) which does not necessarily involve rolling circle replication of an excised circular product (see below). Using an example of an H–H BIME, an initial cleavage at an 'iREP' site [Figure 7B(i)] would generate a 3′-OH which could act as a primer and be extended by host DNA polymerases [Figure 7B(ii)]. A second cleavage occurring at a 'linker' site 3′ of the REP on the newly synthesized DNA strand [Figure 7B(iii)] would liberate another 3′-OH that attacks the first Tyr-5′P complex [Figure 7B(iv)]. This would lead to addition of a supplementary BIME unit [Figure 7B(v,vi)].

Alternatively, amplification could take place by a mechanism involving rolling circle replication from two BIMEs in tandem such as that proposed previously [(24); Figure 7C]. Although this model is attractive and would explain the amplification process, it requires four $TnpA_{REP}$-directed cleavages and might appear complex. However, the model implies cleavages on BIMEs in both strands [Figure 7C(ii,iii)], might therefore explain the necessity to maintain BIME as unit.

Figure 7. Continued
model: The first cleavage at 'iREP' site [B(i)] would generate a 3′-OH which could act as a primer and be extended by host DNA polymerases [B(ii)]. A second cleavage occurring at a 'linker' site 3′ of the REP on the newly synthesized DNA strand [B(iii)] liberates another 3′-OH (in red) that attacks the first Tyr-5′P complex [B(iv), in grey]. This leads to addition of a supplementary BIME unit [B(v,vi)]. (**C**) Alternative model of BIME amplification [adapted from (24)]: From two BIMEs in tandem, two cleavages on the bottom strand, followed by reciprocal strand exchange [C(ii)] would lead to excision of the bottom central BIME. A 3′-OH resulting from a third cleavage on the top strand [C(iii)] would be used as primer for a 'rolling replication amplification' of the excised circular BIME. A fourth cleavage on the newly synthesized DNA strand [C(iv)] liberates another 3′-OH (in blue) that attacks the third Tyr-5′P complex (in pink). This leads to addition of one (or numerous) supplementary BIME(s) [C(vi)].

Note that the recombination we observe *in vitro* between two 'iREP' sites (Figure 6A) and between a 'iREP' and a 'linker' site (Figure 6B), is equivalent to steps Aii and Biv. We believe that integration step (Aiv) corresponds to a 'recombination' of excised BIME with a target carrying a REP/BIME. The ability of $TnpA_{REP}$ to cleave at a significant distance upstream or downstream of a resident BIME (Figure 5) would indeed enable a BIME insertion/recombination at a distance from an existent BIME, therefore disseminating them on the chromosome. This capacity would also allow the unit to sequester additional flanking sequences including entire genes or gene fragments as have been observed for *codA–cynR* from *E. coli* and $tnpA_{REP}$ from *Pseudomonas putida*, *Pseudomonas fluorescens*, *Stenotrophomonas maltophilia*, *Mannheimia succinicproducens* (24) (Supplementary Figure S7). Acquisition of neighbouring genes is also characteristic of rolling circle (RC) transposition of the IS*91*, ISCR and Helitron elements (47,48)

While these data underline the potential plasticity conferred on the host genome by the $tnpA_{REP}$/BIME system, neither the exact mechanism nor the inherent frequency of $TnpA_{REP}$-mediated BIME recombination are at present known. Further bioinformatic approaches are expected to reveal more details concerning BIME spread through genomes. Moreover, mechanistic studies would be greatly aided by knowledge of the $TnpA_{REP}$ structure with and without its DNA substrate. Additionally, it will be essential in the future to develop an *in vivo* system to observe the activity of the $tnpA_{REP}$/BIME system within its host genome since it is possible that $TnpA_{REP}$ activity requires host proteins and may be coupled to cell physiology such as replication, transcription or supercoiling. Such studies are underway.

This class of enzyme is widespread. It includes transposases of the IS*200*/IS*605* and IS*91*/ISCR families of insertion sequences, $TnpA_{REP}$, relaxases of conjugative plasmids and proteins involved in the replication of rolling circle plasmids, phage and eukaryotic viruses. These studies raise important questions concerning the evolutionary relationship between transposable elements and their domestication in cell function.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1 and Supplementary Figures 1–7.

ACKNOWLEDGEMENTS

We would like to thank D. Lane, L. Lavatine, A. Hickman and F. Dyda as well as members of the Mobile Genetic Elements group for reading the manuscript and for discussions. We would also like to thank the two anonymous referees for their very constructive suggestions.

FUNDING

Centre National de Recherche Scientifique (CNRS, France) (intramural). Funding for open access charge: Intramural CNRS funding.

Conflict of interest statement. None declared.

REFERENCES

1. Higgins,C.F., Ames,G.F., Barnes,W.M., Clement,J.M. and Hofnung,M. (1982) A novel intercistronic regulatory element of prokaryotic operons. *Nature*, **298**, 760–762.
2. Espeli,O., Moulin,L. and Boccard,F. (2001) Transcription attenuation associated with bacterial repetitive extragenic BIME elements. *J. Mol. Biol.*, **314**, 375–386.
3. Khemici,V. and Carpousis,A.J. (2004) The RNA degradosome and poly(A) polymerase of *Escherichia coli* are required in vivo for the degradation of small mRNA decay intermediates containing REP-stabilizers. *Mol. Microbiol.*, **51**, 777–790.
4. Aguena,M., Ferreira,G.M. and Spira,B. (2009) Stability of the pstS transcript of *Escherichia coli*. *Arch. Microbiol.*, **191**, 105–112.
5. Moulin,L., Rahmouni,A.R. and Boccard,F. (2005) Topological insulators inhibit diffusion of transcription-induced positive supercoils in the chromosome of *Escherichia coli*. *Mol. Microbiol.*, **55**, 601–610.
6. Boccard,F. and Prentki,P. (1993) Specific interaction of IHF with RIBs, a class of bacterial repetitive DNA elements located at the 3′ end of transcription units. *EMBO J.*, **12**, 5019–5027.
7. Gilson,E., Bachellier,S., Perrin,S., Perrin,D., Grimont,P.A., Grimont,F. and Hofnung,M. (1990) Palindromic unit highly repetitive DNA sequences exhibit species specificity within *Enterobacteriaceae*. *Res. Microbiol.*, **141**, 1103–1116.
8. Gilson,E., Perrin,D. and Hofnung,M. (1990) DNA polymerase I and a protein complex bind specifically to E. coli palindromic unit highly repetitive DNA: implications for bacterial chromosome organization. *Nucleic Acids Res.*, **18**, 3941–3952.
9. Espeli,O. and Boccard,F. (1997) In vivo cleavage of *Escherichia coli* BIME-2 repeats by DNA gyrase: genetic characterization of the target and identification of the cut site. *Mol. Microbiol.*, **26**, 767–777.
10. Clement,J.-M., Wilde,C., Bachellier,S., Lambert,P. and Hofnung,M. (1999) IS1397 is active for transposition into the chromosome of *Escherichia coli* K-12 and inserts specifically into palindromic units of bacterial interspersed mosaic elements. *J. Bacteriol.*, **181**, 6929–6936.
11. Wilde,C., Bachellier,S., Hofnung,M. and Clement,J.-M. (2001) Transposition of IS1397 in the family *Enterobacteriaceae* and first characterization of ISKpn1, a new insertion sequence associated with *Klebsiella pneumoniae* palindromic units. *J. Bacteriol.*, **183**, 4395–4404.
12. Tobes,R. and Pareja,E. (2006) Bacterial repetitive extragenic palindromic sequences are DNA targets for insertion sequence elements. *BMC Genomics*, **7**, 62.
13. Stern,M.J., Ames,G.F.-L., Smith,N.H., Clare Robinson,E. and Higgins,C.F. (1984) Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell*, **37**, 1015–1026.
14. Gilson,E., Saurin,W., Perrin,D., Bachellier,S. and Hofnung,M. (1991) Palindromic units are part of a new bacterial interspersed mosaic element (BIME). *Nucl. Acids Res.*, **19**, 1375–1383.
15. Gilson,E., Saurin,W., Perrin,D., Bachellier,S. and Hofnung,M. (1991) The BIME family of bacterial highly repetitive sequences. *Res. Microbiol.*, **142**, 217–222.
16. Aranda-Olmedo,I., Tobes,R., Manzanera,M., Ramos,J.L. and Marques,S. (2002) Species-specific repetitive extragenic palindromic (REP) sequences in *Pseudomonas putida*. *Nucleic Acids Res.*, **30**, 1826–1833.
17. Rocco,F., De Gregorio,E. and Di Nocera,P.P. (2010) A giant family of short palindromic sequences in *Stenotrophomonas maltophilia*. *FEMS Microbiol. Lett.*, **308**, 185–192.

18. Bertels,F. and Rainey,P.B. (2011) Within-genome evolution of REPINs: a new family of miniature mobile DNA in bacteria. *PLoS Genet*., **7**, e1002132.
19. Bachellier,S., Clément,J.-M. and Hofnung,M. (1999) Short palindromic repetitive DNA elements in enterobacteria: a survey. *Res. Microbiol*., **150**, 627–639.
20. Kersulyte,D., Velapatino,B., Dailide,G., Mukhopadhyay,A.K., Ito,Y., Cahuayme,L., Parkinson,A.J., Gilman,R.H. and Berg,D.E. (2002) Transposable element ISHp608 of *Helicobacter pylori*: nonrandom geographic distribution, functional organization, and insertion specificity. *J. Bacteriol*., **184**, 992–1002.
21. Ton-Hoang,B., Guynet,C., Ronning,D.R., Cointin-Marty,B., Dyda,F. and Chandler,M. (2005) Transposition of ISHp608, member of an unusual family of bacterial insertion sequences. *EMBO J*., **24**, 3325–3338.
22. Ronning,D.R., Guynet,C., Ton-Hoang,B., Perez,Z.N., Ghirlando,R., Chandler,M. and Dyda,F. (2005) Active site sharing and subterminal hairpin recognition in a new class of DNA transposases. *Mol. Cell*, **20**, 143–154.
23. Barabas,O., Ronning,D.R., Guynet,C., Hickman,A.B., Ton-Hoang,B., Chandler,M. and Dyda,F. (2008) Mechanism of IS200/IS605 family DNA transposases: activation and transposon-directed target site selection. *Cell*, **132**, 208–220.
24. Nunvar,J., Huckova,T. and Licha,I. (2010) Identification and characterization of repetitive extragenic palindromes (REP)-associated tyrosine transposases: implications for REP evolution and dynamics in bacterial genomes. *BMC Genomics*, **11**, 44.
25. Guynet,C., Hickman,A.B., Barabas,O., Dyda,F., Chandler,M. and Ton-Hoang,B. (2008) In vitro reconstitution of a single-stranded transposition mechanism of IS608. *Mol. Cell*, **29**, 302–312.
26. Pasternak,C., Ton-Hoang,B., Coste,G., Bailone,A., Chandler,M. and Sommer,S. (2010) Irradiation-induced *Deinococcus radiodurans* genome fragmentation triggers transposition of a single resident insertion sequence. *PLoS Genet*., **6**, e1000799.
27. Hickman,A.B., James,J.A., Barabas,O., Pasternak,C., Ton-Hoang,B., Chandler,M., Sommer,S. and Dyda,F. (2010) DNA recognition and the precleavage state during single-stranded DNA transposition in D. radiodurans. *EMBO J*., **29**, 3840–3852.
28. Guynet,C., Achard,A., Hoang,B.T., Barabas,O., Hickman,A.B., Dyda,F. and Chandler,M. (2009) Resetting the site: redirecting integration of an insertion sequence in a predictable way. *Mol. Cell*, **34**, 612–619.
29. He,S., Hickman,A.B., Dyda,F., Johnson,N.P., Chandler,M. and Ton-Hoang,B. (2011) Reconstitution of a functional IS608 single-strand transpososome: role of non-canonical base pairing. *Nucleic Acids Res*., **39**, 8503–8512.
30. Ton-Hoang,B., Pasternak,C., Siguier,P., Guynet,C., Hickman,A.B., Dyda,F., Sommer,S. and Chandler,M. (2010) Single-stranded DNA transposition is coupled to host replication. *Cell*, **142**, 398–408.
31. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol*., **215**, 403–410.
32. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*., **22**, 4673–4680.
33. Corpet,F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res*., **16**, 10881–10890.
34. Waterhouse,A.M., Procter,J.B., Martin,D.M., Clamp,M. and Barton,G.J. (2009) Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.
35. Van Dongen,S. (2000) A cluster algorithm for graphs. *Technical Report INS-R0010, National Research Institute for Mathematics and Computer Science. in the Netherlands*, Amsterdam.
36. Enright,A.J., Van Dongen,S. and Ouzounis,C.A. (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res*., **30**, 1575–1584.
37. Frith,M.C., Saunders,N.F., Kobe,B. and Bailey,T.L. (2008) Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput. Biol*., **4**, e1000071.
38. Bachellier,S., Saurin,W., Perrin,D., Hofnung,M. and Gilson,E. (1994) Structural and functional diversity among bacterial interspersed mosaic elements (BIMEs). *Mol. Microbiol*., **12**, 61–70.
39. Bachellier,S., Clement,J.M., Hofnung,M. and Gilson,E. (1997) Bacterial interspersed mosaic elements (BIMEs) are a major source of sequence polymorphism in *Escherichia coli* intergenic regions including specific associations with a new insertion sequence. *Genetics*, **145**, 551–562.
40. Gillespie,J.J., Wattam,A.R., Cammer,S.A., Gabbard,J.L., Shukla,M.P., Dalay,O., Driscoll,T., Hix,D., Mane,S.P., Mao,C. *et al.* (2011) PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. *Infect. Immun*., **79**, 4286–4298.
41. Touchon,M., Hoede,C., Tenaillon,O., Barbe,V., Baeriswyl,S., Bidet,P., Bingen,E., Bonacorsi,S., Bouchier,C., Bouvet,O. *et al.* (2009) Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet*., **5**, e1000344.
42. Chalker,D.L. and Yao,M.C. (2010) DNA elimination in ciliates: transposon domestication and genome surveillance. *Annu. Rev. Genet*., **45**, 227–246.
43. Sinzelle,L., Izsvak,Z. and Ivics,Z. (2009) Molecular domestication of transposable elements: from detrimental parasites to useful host genes. *Cell. Mol. Life Sci*., **66**, 1073–1093.
44. Snyder,E.E., Kampanya,N., Lu,J., Nordberg,E.K., Karur,H.R., Shukla,M., Soneja,J., Tian,Y., Xue,T., Yoo,H. *et al.* (2007) PATRIC: the VBI PathoSystems Resource Integration Center. *Nucleic Acids Res*., **35**, D401–D406.
45. Drolet,M. (2006) Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. *Mol. Microbiol*., **59**, 723–730.
46. Bikard,D., Loot,C., Baharoglu,Z. and Mazel,D. (2011) Folded DNA in action: hairpin formation and biological functions in prokaryotes. *Microbiol. Mol. Biol. Rev*., **74**, 570–588.
47. Toleman,M.A., Bennett,P.M. and Walsh,T.R. (2006) ISCR elements: novel gene-capturing systems of the 21st century? *Microbiol. Mol. Biol. Rev*., **70**, 296–316.
48. Kapitonov,V.V. and Jurka,J. (2001) Rolling-circle transposons in eukaryotes. *Proc. Natl Acad. Sci. USA*, **98**, 8714–8719.