

A putative mobile genetic element carrying a novel type IIF restriction-modification system (PluTI)

Feroz Khan^{1,2}, Yoshikazu Furuta^{1,2}, Mikihiro Kawai^{1,3}, Katarzyna H. Kaminska⁴, Ken Ishikawa^{1,2,6}, Janusz M. Bujnicki^{1,4,5} and Ichizo Kobayashi^{1,2,6,*}

¹Department of Medical Genome Sciences, Graduate School of Frontier Sciences, ²Institute of Medical Science, University of Tokyo, ³National Institute for Basic Biology, National Institutes of Natural Sciences, Japan, ⁴Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, ⁵Laboratory of Structural Bioinformatics, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Poznan, Poland and ⁶Graduate Program in Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Japan

Received October 7, 2009; Revised and Accepted December 18, 2009

ABSTRACT

Genome comparison and genome context analysis were used to find a putative mobile element in the genome of *Photorhabdus luminescens*, an entomopathogenic bacterium. The element is composed of 16-bp direct repeats in the terminal regions, which are identical to a part of insertion sequences (ISs), a DNA methyltransferase gene homolog, two genes of unknown functions and an open reading frame (ORF) (plu0599) encoding a protein with no detectable sequence similarity to any known protein. The ORF (plu0599) product showed DNA endonuclease activity, when expressed in a cell-free expression system. Subsequently, the protein, named R.PlTI, was expressed *in vivo*, purified and found to be a novel type IIF restriction enzyme that recognizes 5'-GGCG C/C-3' (/ indicates position of cleavage). R.PlTI cleaves a two-site supercoiled substrate at both the sites faster than a one-site supercoiled substrate. The modification enzyme homolog encoded by plu0600, named M.PlTI, was expressed in *Escherichia coli* and shown to protect DNA from R.PlTI cleavage *in vitro*, and to suppress the lethal effects of R.PlTI expression *in vivo*. These results suggested that they constitute a restriction-modification system, present on the putative mobile element. Our approach thus allowed detection of a previously uncharacterized family of DNA-interacting proteins.

INTRODUCTION

Restriction-modification (RM) systems, which are abundant among prokaryotic cells (1) attack invading DNAs to defend against invading genetic elements such as phages. However, many lines of evidence indicate that the RM systems themselves are mobile since they are found in mobile elements such as: plasmids, phages, integrons, transposons and conjugative transposons (2). The structure of some restriction gene complexes is related to that of transposons, and these may move as a part of composite transposons (3,4,5). In some naturally competent bacteria, restriction gene complexes move without other linked genes, for example by insertion involving a long variable target duplication (6). A comparison of closely related prokaryotic genomes revealed linkage of restriction gene homologs with various large genome polymorphisms and rearrangements (2,7). Similar to some pathogenic mobile elements, RM systems sometimes attack their hosts (8) and, like viral genomes and transposons, may be regarded as selfish mobile elements or opportunistic pathogens.

The extreme sequence diversity seen in the restriction (R) enzymes that form RM systems (9,10) could be related to their interaction with the host bacteria. RM systems consist of two enzymatic activities, a restriction endonuclease (R) that recognizes a specific DNA sequence and introduces a double-strand break, and a cognate modification enzyme (M) that protects the host genome from self-cleavage by methylating the cognate recognition sites (11). At least some type II RM systems, in which restriction enzymes cleave at or near a fixed sequence (10), have been observed to kill cells that have lost their genes in a process called post-segregational killing (7).

*To whom correspondence should be addressed. Fax: +81 3 5449 5422; Email: ikobaya@ims.u-tokyo.ac.jp

Transcriptome analysis further suggests that they trigger a bacterial programmed death pathway that is also activated by bactericidal antibiotics (12).

Some type II restriction enzymes, called Type IIE, interact with two copies of their recognition sequence but cleave only one (10,13); while others called Type IIF interact and cleave both the sites coordinately (13). The crystal structures of some Type IIF enzymes such as NgoMIV (14), Cfr10I (15) and Bse634I (16) show that two subunits form one of the equivalent DNA binding sites, and two such dimers pack back-to-back to constitute an active tetramer. These restriction endonucleases interact with two recognition sites *in cis* by forming a DNA loop (17,18). Their reaction mechanisms (19,20) and structures have been elucidated (14–16), but their biological significance is still not clear.

The restriction enzymes are also structurally diverse (21). Five restriction enzyme folds have been determined by structural and bioinformatics analyses: the PD-(D/E)XK nuclease fold, the HNH (= $\beta\beta\alpha$ -Me) fold, the GIY-YIG fold, the phospholipase D fold that includes R.BfiI, and the half-pipe fold that includes R.PabI. Many restriction enzymes remain unassigned to any of these, and may exhibit new architectures (9). A novel restriction enzyme structure would enlarge the structural universe of DNA-interacting proteins. We have developed a genome comparison method, based on the mobility of RM genes, to identify restriction enzymes with novel structure (22). We compared two closely related genomes and found a subgenomic region specific to one of the genomes and encompassing a DNA methyltransferase homolog and an open reading frame (ORF) lacking similarity to any structurally characterized protein (23). Restriction enzyme activity was observed for this ORF in an enzyme assay after *in vitro* protein expression using wheat germ extract (22), and subsequently the encoded enzyme, named R.PabI, was found to have a novel fold (24).

We have extended the genome comparison and genome context analysis to available prokaryotic genomes to find a restriction enzyme of novel structure. Here, we present a newly characterized restriction endonuclease, R.PlutI, that may well represent a novel endonuclease fold architecture. This enzyme forms an RM-system with its cognate methyltransferase, M.PlutI. The PlutI RM-system along with two other hypothetical genes is present on a putative mobile element.

MATERIALS AND METHODS

Materials

Escherichia coli K12 strain JM109 {*recA1*, *endA1*, *gyrA96*, *thi*, *hsdR17*, (r_{k}^- , m_{k}^+), *e14*⁻(*mcrA*⁻), *supE44*, *relA*, Δ (*lac-proAB*)/F' [*traD36*, *proAB*⁺, *lacI*^f, *lacZ* Δ *M15*]} was a gift from Dr Akio Nomoto (University of Tokyo) (25). *Escherichia coli* strain ER2566 (*fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10-Tet^S)2 [dcm] R(zgb-210::Tn10-Tet^S) endA1 Δ (*msrC-mrr*)114::IS10*) was from New England Biolabs, Inc, USA.

Escherichia coli strain BL21-DE3 F⁻ *dcm ompT hsdS*(r_{B}^- m_{B}^-) *gal λ (DE3)* was from Novagen.

Photobacterium luminescens subsp. *laumondii* TTO1 genomic DNA (26) was provided by Dr Alain Givaudan (Université Montpellier). The vector plasmid pTXB1, LITMUS38i, pBR322 and pUC19 were from New England Biolabs, Inc.; ϕ X174 DNA (replicative form) and λ -phage DNA were from Takara Bio Inc; pET26b and pET28a were from Novagen. The plasmid pBAD33 (27) was provided by Dr Miki Watanabe (University of Tokyo) and pEU3b (28) by Dr Yaeta Endo (Ehime University). Plasmids pFK1 and pFK2 were constructed from LITMUS38i (see Supplementary Methods for details).

Bioinformatics

Genome sequences were obtained from the National Center for Biotechnology Information (NCBI) Genome database (<http://www.ncbi.nlm.nih.gov/Genomes/>). Homologs of hypothetical and unknown protein genes that flanked a DNA methyltransferase gene homolog, as judged by genome comparison and genome context analysis to be present on the same variable unit, were searched using PSI-BLAST (29) and HHSEARCH (30). Protein secondary and tertiary structure predictions used the GeneSilico metaserver (31). Homology searches were done in October 2009. The details of systematic genome comparison will be presented elsewhere (M.K., K.K., J.M.B. and I.K., unpublished data).

Screening, expression and purification of R.PlutI

The *plutI* ORF was inserted into plasmid pEU3b (28), for *in vitro* screening and expression, and the protein was expressed using a wheat-germ expression kit WEPRO[®] (CellFree Sciences Co, Ltd, Japan). Restriction activity was monitored as the ability of a crude R.PlutI extract to digest ϕ X174 DNA (replicative form). *PhoI* (22) was used as control for restriction activity. For *in vivo* expression, *plutI* gene was inserted into the T7-promoter based pTXB1 expression vector. R.PlutI was purified by affinity and ion-exchange chromatography (see Supplementary Methods for details).

Characterization of R.PlutI

The recognition sequence was determined as described earlier (22) and cleavage position was determined by the primer extension method (32) (see Supplementary Methods for details). Optimal reaction conditions were determined with different buffers, salt concentrations and reaction temperatures as described in Supplementary Methods.

Cleavage kinetics

Substrate cleavage kinetics were monitored as time-dependent changes in the concentrations of substrate (supercoiled plasmid) and products: open circular (OC), full-length linear (FLL) and linear fragments (L1 and L2). The K_{m} and V_{max} were determined by non-linear fitting of the initial reaction velocity for the linear and supercoiled

substrates with one or two sites at different substrate concentrations (see Supplementary Methods for details).

***In vivo* activity of M.PluTI and toxicity of R.PluTI**

In vivo activity of M.PluTI was assessed as ability to protect a plasmid from *in vitro* cleavage with R.PluTI. *In vivo* toxicity of R.PluTI was demonstrated as loss of cell viability in cells induced to produce R.PluTI in the absence of active methylation. (see Supplementary Methods for details).

RESULTS

Genome comparison predicts a mobile unit with a restriction enzyme from a novel sequence family

A bioinformatic survey of systematic genome comparisons discovered, in the genome of a *Photorhabdus luminescence* strain, a subgenomic, potentially mobile unit containing a DNA methyltransferase homolog (plu0600) and three other genes, lying between plu0594 (4-hydroxy-3-methylbut-2-enyl diphosphate reductase) and plu0602 (dihydrodipicolinate reductase) (Figure 1). This rod-shaped bacterium (classification: *Bacteria*; *Proteobacteria*; *Gammaproteobacteria*; *Enterobacteriales*; *Enterobacteriaceae*; *Photorhabdus*; *Photorhabdus luminescens* subsp. *laumondii* TTO1) lives symbiotically in the gut of nematodes that attack insect larvae and eventually kill their host using various toxins (26,33).

The alignment of the *P. luminescence* genome with others revealed that the DNA methyltransferase-containing unit is substituted by diverse sequences. The observed 46 substitution cases were classified into six groups based on mutual homology (Table 1). Group A contained eight cases and had no annotated genes in the substitution region. In Group B, the unit was substituted with a hypothetical protein gene. Group C had seven cases all with an insertion of IS1541. Group D contained substitutions of the ribonucleoside hydrolase RihC gene only (Subgroup D-1) or RihC with other genes (Subgroup D-2). Group E contained a substitution of the LysR family transcriptional regulator gene only (Subgroup E1)

or LysR with other genes (Subgroup E2). Finally, Group F comprised one case containing several genes, including the AraC family transcriptional regulator gene.

In Subgroup E-2, the substituted region contained dihydrodipicolinate synthetase gene, whose product is known to act in the same pathway as dihydrodipicolinate reductase (product of the plu0602 gene) (<http://www.genome.jp/kegg/>) (34), which lies next to the putative mobile unit (Figure 1). Thus, we suggest that Subgroup E-2 may represent the ancient form of this locus, in which gene loss occurred resulting in an empty site as in Group A. This site experienced the insertion of various genes such as IS, the putative mobile unit identified here, and other genes.

Direct repeats flanking the PluTI unit

A BLASTN search for internal sequence similarity around the DNA methyltransferase homolog-containing unit (discussed above) showed 16-bp direct repeats flanking the unit (Figure 1). The repeats showed perfect (16/16) sequence matches and were not found elsewhere in the genome, suggesting a possible horizontal transfer of this unit, followed by its insertion into the genome by the site-specific recombination, similar to DNA phage genomes.

However, upon closer examination, the direct repeats appeared to be derived from IS elements (Supplementary Materials). The left repeat (Figure 1) is located in a region homologous to IS*Plu6* (26) (IS Finder, <http://www-is.biotoul.fr/>), an IS982 family member. The right repeat is located in a region homologous to another subfamily of IS982 family (35). These observations raise the other possibility of site-specific integration that the unit entered in the genome through some action of these terminal IS elements such as aberrant transposition and non-autonomous transposition.

Biochemical and biological experiments (described later) demonstrated that the plu0599 (= *pluTIR*) gene, which has no sequence similarity to any gene for any functionally or structurally characterized protein, encodes a restriction enzyme, designated R.PluTI, and that the

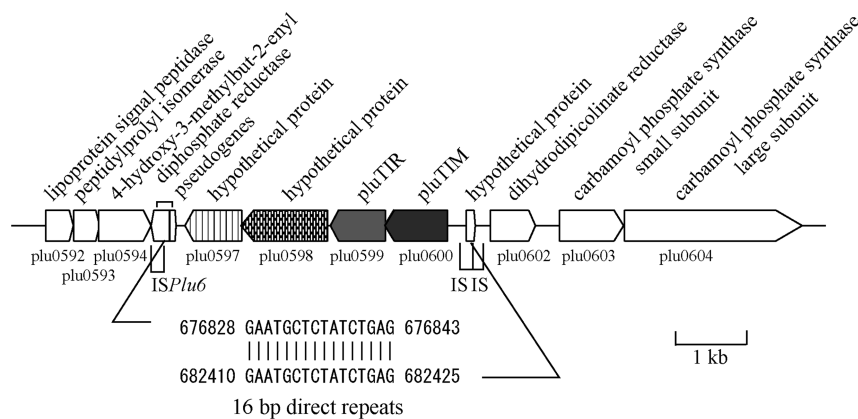


Figure 1. Genome organization of the PluTI RM complex. *pluTIR* and *pluTIM* genes are shown in solid boxes. The flanking 16-bp direct repeats are shown below. IS*Plu6* indicates nucleotide sequence similarity to the 3'-end of IS*Plu6* of IS982 family, while IS indicates homology to the 3'- and 5'-ends of IS copies from another subfamily of the IS982 family.

Table 1. Gene clusters substituting the PluTI unit

Group	Members found	Representative	Accession number	plu594 homolog	plu602 homolog	Insertion length (average of members)	Annotation of some of genes present
A	8	<i>Escherichia coli</i> ATCC 8739	NC_010468.1	EcolC_3626	EcolC_3624	1502	no gene
B	3	<i>Yersinia pseudotuberculosis</i> IP 31758	NC_009708.1	YpsIP31758_3457	YpsIP31758_3455	1621	hypothetical protein
C	7	<i>Yersinia pseudotuberculosis</i> IP 32953	NC_006155.1	YPTB0620	YPTB0622	2081	transposase of IS1541
D-1	21	<i>Escherichia coli</i> SMS-3-5	NC_010498.1	EcSMS35_0027	EcSMS35_0029	2610	ribonucleoside hydrolase RihC
D-2	4	<i>Salmonella enterica</i> subsp. enterica serovar Paratyphi B str. SPB7	NC_010102.1	SPAB_00059	SPAB_00076	16343	ribonucleoside hydrolase RihC, oxaloacetate decarboxylase
E-1	1	<i>Enterobacter</i> sp. 638	NC_009436.1	Ent638_0587	Ent638_0590	2735	LysR family transcriptional regulator
E-2	1	<i>Serratia proteamaculans</i> 568	NC_009832.1	Spro_0701	Spro_0712	13291	LysR family transcriptional regulator, dihydrodipicolinate synthetase, dihydroxy-acid dehydratase
F	1	Sequence 16855 from patent US 7319142	EA414747.1	N/A ^a	N/A ^a	12316	AraC-family transcriptional regulator ^b , major facilitator superfamily MFS_1 ^b

^aAnnotations not available.^bTop hit by BLASTX.

plu0600 gene (= *pluTIM*) encodes a modification methyltransferase, designated M.PluTI, which together constitute PluTI RM system (discussed below). The modification enzyme has also been annotated in REBASE.

Restriction enzyme activity detected by *in vitro* expression

The candidate ORF *pluTIR* (plu0599) was amplified from *P. luminescens* genomic DNA by PCR and placed in a vector for *in vitro* expression, pEU3b (28). DNA purified from transformed *E. coli* was transcribed *in vitro* with SP6 RNA polymerase, and proteins synthesized using wheat-germ extract. The product showed clear *in vitro* restriction endonuclease activity on a substrate DNA (ϕ X174 DNA, replicative form) (Figure 2).

Characterization of purified R.PluTI

The purified and refolded recombinant R.PluTI (see Supplementary Materials) was found to be active at pH 7 and 20–37°C temperature, and required Mg²⁺ ion for DNA cleavage. The monomeric molecular mass of tagged R.PluTI determined by SDS-PAGE (Supplementary Figure S1), and deduced from the amino acid sequence was 38.6 and 39.1 kDa, respectively. The native molecular mass was 140 (\pm 15) kDa as determined by gel filtration (Supplementary Figure S2), suggesting that R.PluTI forms a homotetramer (α_4) similar to other Type IIF enzymes (see below). R.PluTI is thermolabile and lost 100% activity after 30 min at 65°C. One unit of R.PluTI was defined as the amount of enzyme required for complete digestion of 1 μ g of ϕ X174 DNA (replicative form) in 1 h at 37°C in a 50 μ l reaction and specific activity of purified enzyme was 8000 U/mg.

The recognition sequence of R.PluTI was predicted to be 5'-GGCGCC-3' by REBpredictor (36) using the

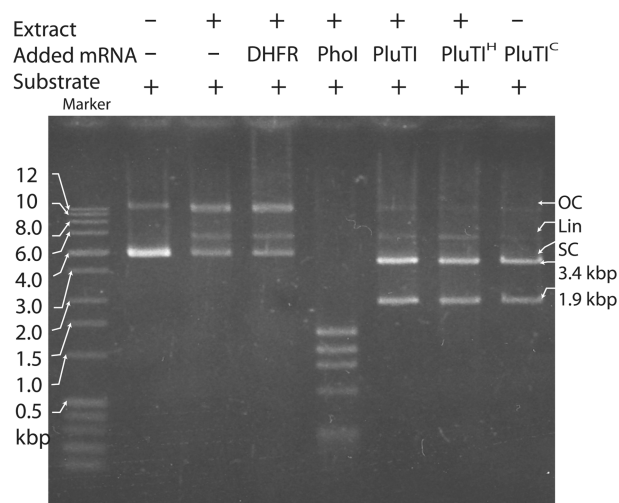


Figure 2. Detection of restriction activity of cell-free translation products. mRNA synthesized *in vitro* was incubated with a wheat-germ-based *in vitro* translation mixture. The extract was then incubated with ϕ X174 DNA (replicative form) for 1 h at 37°C (65°C for PhoI), and the digest resolved on 1% agarose gel and visualized with ethidium bromide staining. Marker, perfect DNA marker (Novagen); Extract, wheat-germ extract; Substrate, ϕ X174 DNA (replicative form); DHFR, dihydrofolate reductase (negative control for restriction activity); PhoI, restriction enzyme PhoI (positive control for restriction activity); PluTI, wild-type R.PluTI; PluTI^H, 6His-tagged R.PluTI; PluTI^C, cloned R.PluTI (with two extra amino acid residues at the N terminus) purified following *in vivo* expression in *E. coli*. SC, supercoiled; OC, open circle; Lin, linear.

cleavage pattern of pUC19 and ϕ X174 DNA (replicative form) (Supplementary Figure S3), similar to EheI, BbeI, NarI, DinI, EgeI, KasI, SfoI, SspDI and Mly113I. The cleavage position was determined to be 5'-GGCGC↓C-3' by comparison of a dideoxy sequencing ladder with

fragments generated by R.PluTI cleavage (Supplementary Figure S4).

Kinetic analysis

Efficient cleavage of a two-site plasmid in contrast to a one-site plasmid (Supplementary Figure S3) suggested the possibility that R.PluTI requires two copies of its recognition sequence for efficient cleavage. Many restriction enzymes such as Type I, IIB, IIE, IIF, most IIS, III and IV enzymes require two-sites for optimal activity (10). The mode of action can be deduced from reactions that occur in the presence of two supercoiled substrate DNAs: one with a single copy of the recognition sequence (one-site) and the other with two copies (two-site) (19,37). If an enzyme must bind two sites before cleavage, the cleavage rate *in cis* reaction for two-site substrates will be much higher than *in trans* reaction for one-site substrates.

This mechanism is distinguished by steady-state cleavage reactions when enzyme concentration is considerably lower than the substrate. Based on the one-site plasmid, LITMUS38i, we constructed a two-site plasmid, pFK1 with two R.PluTI sites placed 853 bp away from each other, with identical flanking sequences (Supplementary Figure S6). R.PluTI consumed initial 10% substrate at a rate of 0.011 nM/min for the one-site substrate, and 0.28 nM/min for the two-site substrate (Figure 3). In the reaction on a SC one-site plasmid, OC form that was cut only at one strand was generated, along with a relatively small fraction of FLL form cut on both strands (Figure 3a–b). In contrast, products from the two-site plasmid were mainly linear fragments L1 and L2, which were formed by cleavage of both strands at both sites (Figure 3c–d). Less amounts of the FLL and OC forms were obtained, indicating concerted cleavage mechanism working at both sites. These kinetics were very similar to those of NgoMIV, a Type IIF enzyme (19).

If a cleavage reaction requires two sites, the slow cleavage reaction against the one-site supercoiled plasmid may be enhanced by addition of an oligoduplex carrying the recognition sequence (38). Addition of such a duplex carrying the R.PluTI site initially enhanced cleavage of an SC plasmid (Figure 4) as reported for another Type IIF enzyme (39), but inhibited at higher concentrations, because of engagement of the enzyme molecule between two molecules of oligoduplex, rather than an oligoduplex and a plasmid (38).

The K_m and V_{max} were determined for one- or two-site linear and supercoiled DNA substrates (Table 2, Supplementary Figure S5); and the enzyme was found to obey Michaelis–Menten kinetics (Supplementary Figure S5) similar to other restriction enzymes (40,41). The two-site SC substrate showed a 27-fold higher V_{max} than the one-site SC substrate, while this difference was insignificant between the one- and two-site linear substrates. The V_{max} values for the one-site linear and the two-site linear substrate were between these values: [two-site SC] > [two-site linear] = [one-site linear] >> [one-site SC]. Superhelicity appeared to facilitate the reaction with

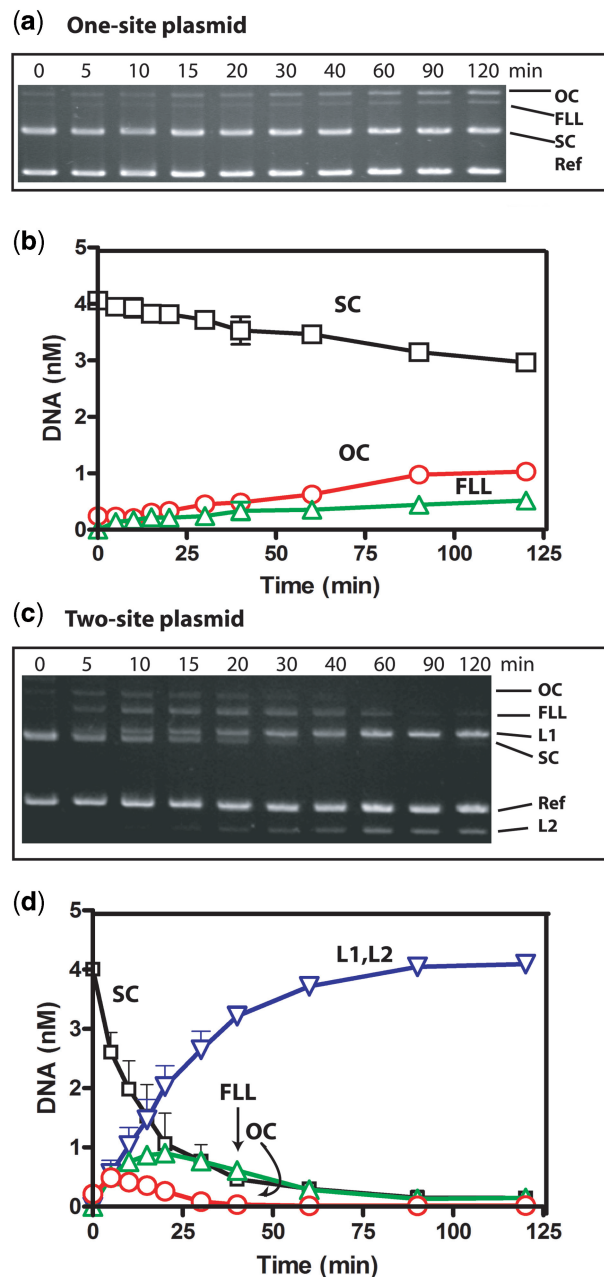


Figure 3. R.PluTI reactions on one- and two-site plasmids. Reaction contained 0.5 nM of purified R.PluTI and 4 nM of supercoiled (SC) DNA in NEB1* buffer (10 mM Tris–propane buffer, pH 7.0 supplemented with 10 mM MgCl₂, 1 mM dithiothreitol and 100 μg/ml BSA) at 37°C. DNA was (a and b), LITMUS 38i, with one R.PluTI site; (c and d), pFK1 with two R.PluTI sites. (a and c) Agarose gel electrophoresis of DNA taken from the reaction at the indicated time: electrophoretic mobilities of open-circle (OC), full-length linear (FLL), supercoiled (SC) and reference DNAs (1.3 kb linear fragment) are marked on the right, in (c) two linear products (L1 and L2) resulting from cutting at the both sites are also marked. (b and d) Concentrations of substrate (SC) and products (OC, FLL and average of L1, L2) over 120 min of reaction. Error bars (shown only one side, and in some cases masked by symbols) indicate standard deviation from three or more independent sets of experiments.

two-site substrates but suppress the reaction with one-site substrates, which could be attributed to steric hindrance, as the supercoiled nature of plasmid DNA reduces the efficiency of *trans* interactions (38).

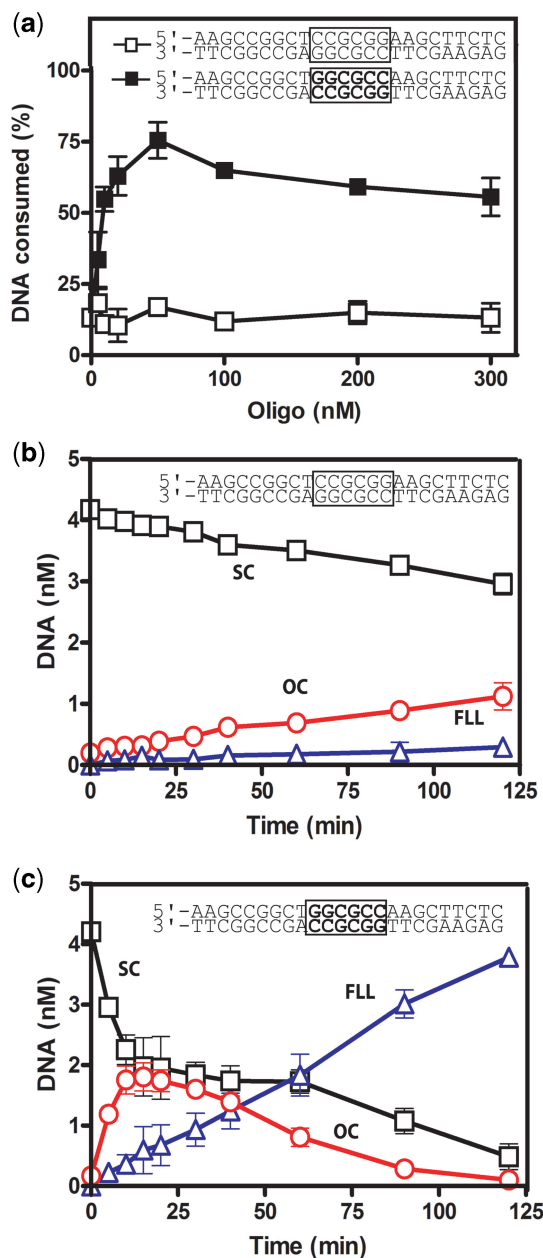


Figure 4. Effect of oligoduplexes on one-site plasmid cleavage. (a) Varying concentrations of oligoduplexes with fixed plasmid concentration. Reactions at 37°C contained 0.5 nM of purified R.Pluti and 4 nM of supercoiled (SC) LITMUS38i DNA in NEB1* buffer (10 mM Tris-propane buffer, pH 7.0 supplemented with 10 mM MgCl₂, 1 mM dithiothreitol, and 100 µg/ml BSA) and oligoduplexes with (solid squares) or without (open squares) the recognition sequence at concentrations indicated on the vertical axis. Reactions were stopped after 30 min; and samples were analyzed for the residual concentration of supercoiled LITMUS38i. (b and c) Varying concentrations of plasmid with a fixed concentration of oligoduplexes. Reaction, contained 0.5 nM of purified R.Pluti and varying concentration of supercoiled (SC) DNA (indicated on vertical axes) in NEB1* buffer (described above) and 100 nM oligoduplex. Following forms of the DNA were monitored over 120 min: SC substrate, OC DNA, FLL DNA. Oligoduplex without (b) and with (c) the recognition sequence were used. In all the three panels, error bars (masked by symbols in some cases) indicate standard deviation from three or more independent sets of experiments.

Table 2. V_{max} and K_m with various substrates

Substrate	V_{max} (nM/min)	K_m (nM)
One-site supercoiled	0.2	33
Two-site supercoiled	5.4	57
One-site linear	1.5	20
Two-site linear	1.4	12

See Supplementary Figure S6 for the substrates and Supplementary Figure S5 for the measurements.

M.Pluti prevents the cleavage of DNA by R.Pluti

The activity of the plu0600 product M.Pluti was examined *in vivo*. Plasmid pBAD33-*pluTIM* carries two R.Pluti sites in 20 bp proximity so reactions *in cis* are not feasible. Indeed, even unmethylated plasmid DNA was not efficiently cleaved by R.Pluti while it was cleaved by EhoI (data not shown). Therefore, we constructed pET26b-*pluTIM* having four R.Pluti sites, which was cleaved with R.Pluti or EhoI *in vitro* in uninduced cell (Supplementary Figure 7a), while in IPTG-induced cells, it was efficiently protected from *in vitro* cleavage by R.Pluti or EhoI.

In vivo toxicity of R.Pluti to host cells was suppressed by induction of M.Pluti (Supplementary Figure 7b). Taken together with the informatics data, these results support that M.Pluti is a modification enzyme that forms an RM system with R.Pluti.

R.Pluti sequence analysis and structure prediction

A PSI-BLAST search against the NCBI non-redundant database revealed 12 homologs of R.Pluti (Supplementary Table S2). Of these, nine were linked to an M.Pluti homolog, with the same relative order and orientation (M precedes R in the same orientation, Figure 1). The other three appeared truncated compared to the other homologs. The genes flanking *pluTIR* and *pluTIM* homologs were not homologous to each other, supporting the frequent movement of the RM system as a unit.

Analysis of residue conservation patterns in the R.Pluti family (Figure 5), and fold-recognition searches of the database of known structures revealed no overall similarity of these proteins to any of the known restriction enzyme superfamilies or folds, including PD-(D/E)XK, GIY-YIG, HNH, PLD-nuc and Half-pipe (PabI) (9,21,24,42–45). Secondary structure prediction suggested that R.Pluti exhibits α/β structure, with invariant or nearly invariant charged and polar residues concentrated in the middle of the sequence, mostly at the termini of predicted helices and strands, and in loops. This conservation pattern is typical for enzymes, including nucleases.

Regardless of the actual structure, and apart from the invariant or nearly invariant residues D211, H220 and T222, the residues R82, E149, R168, K184, D203, R208, T223, R232, N280, E283, Y303 and N304 are conserved (Figure 5), and may be involved in R.Pluti DNA-binding and cleavage activity. The residues E311, D313 and D319 are also conserved. Some of these conserved residues might be involved in tetramer formation.

The N-terminal region of the R.PluTI family appears divergent, but 9 out of the 11 members in the Figure 5 possess the conserved amino acid sequence EGGRTSRG (amino acid 87–94 in Figure 5), which might be involved in DNA binding, as suggested by protein–DNA binding residue prediction programs (46,47).

Genomic contexts of the gene pair linked to PluTI RM genes

The plu0597 and plu0598 genes, besides the PluTI RM genes, are also within the region flanked by the short direct repeats (Figure 1). The possibility that these two genes play some role in the life cycle of this putative

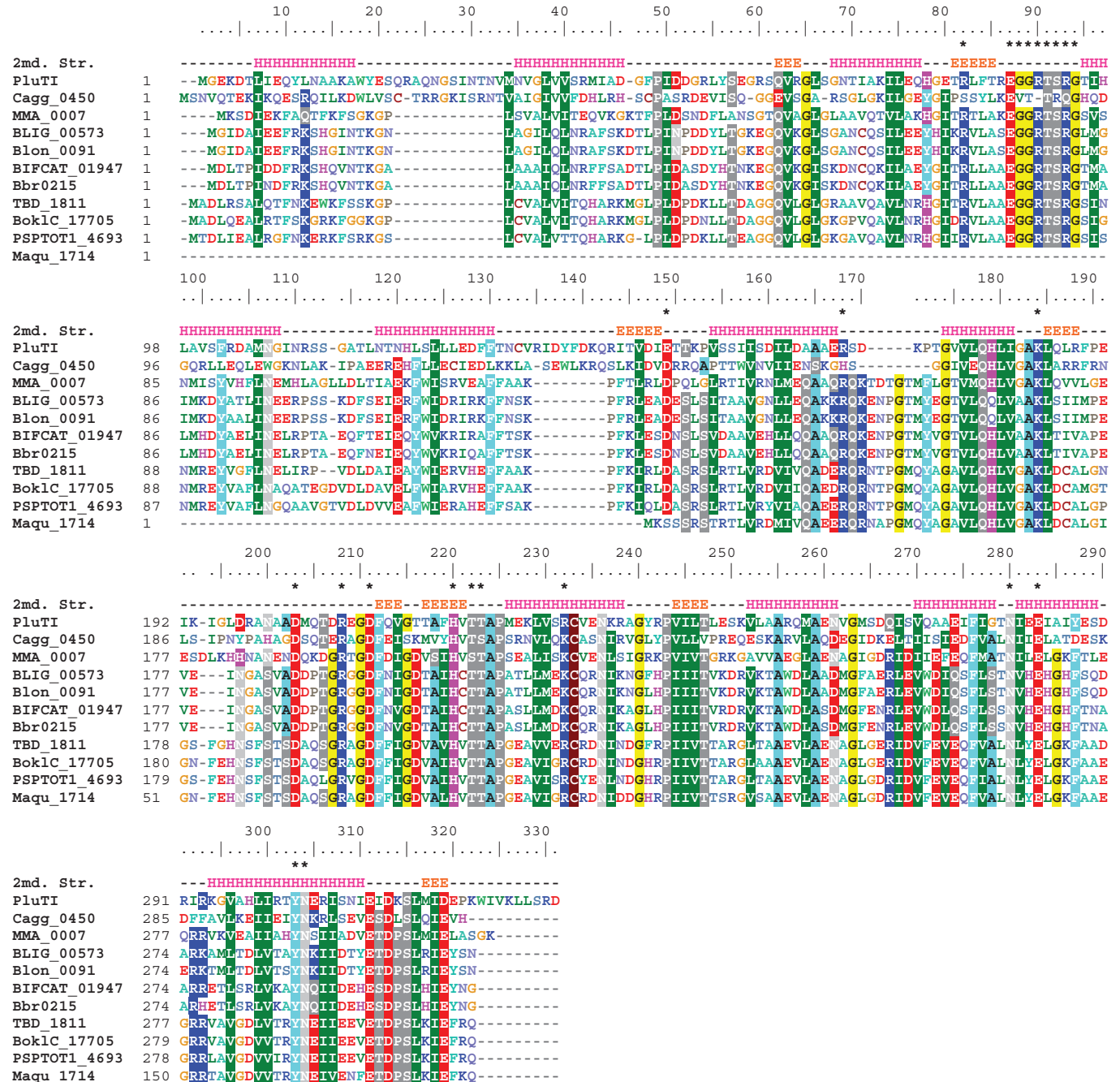


Figure 5. Amino acid sequence alignment of R.PluTI and its PSI-BLAST matches (see Supplementary Table S2 for detail). BL0974 and E4_15863 have been omitted from the alignment, as they are apparently products of degenerated genes (their sequences exhibit terminal truncations, in BL0974 the missing C-terminal part has been replaced with unrelated low-complexity sequence). A frameshift in Blon_0091 was corrected to restore the conserved C-terminal sequence. The alignment was prepared by PROMALS (56) and manually edited. The secondary structure prediction was carried out using the GeneSilico metaserver (31). The highlighted residues show >90% identity or similarity. E, extended (β); H, helical (α); asterisk, residues predicted to be involved in cleavage activity and/or DNA binding.

mobile element is suggested by their amino acid sequences and genome contexts. The products of both genes have a P-loop-containing domain that is conserved in nucleoside triphosphate hydrolases (cl09099), according to the Conserved Domain Database (CDD) (48). Proteins with this motif include DNA-interacting proteins such as helicases and putative endonucleases.

BLASTP and PSI-BLAST analyses showed that plu0597 product has sequence similarity to helicases with *e*-values of 1e-17 and 8e-91, respectively. RPS-BLAST searches against the Conserved Domain Database (*e*-value 8e-21) showed plu0598 was similar to the COG3593, OLD (*overcome lysogenization defect*)

nuclease family. The bacteriophage P2 OLD protein has DNase and RNase activities and consists of an N-terminal ABC-type ATPase domain and a C-terminal TOPRIM domain (49,50). These domains were also detected at the N- and C-terminus of plu0598, respectively.

We also found several homologs of plu0597 and plu0598 genes (*e*-value: 1e-110 to 2e-93 for plu0597 and 1e-141 to 1e-118 for plu0598) adjacent to each other on putative mobile elements including integrated plasmids, transposon-like structures or composite transposon-like structures, suggesting that these two genes might be involved in genome mobility (Figure 6). A homologous pair (Figure 6a) was found in a region annotated as a

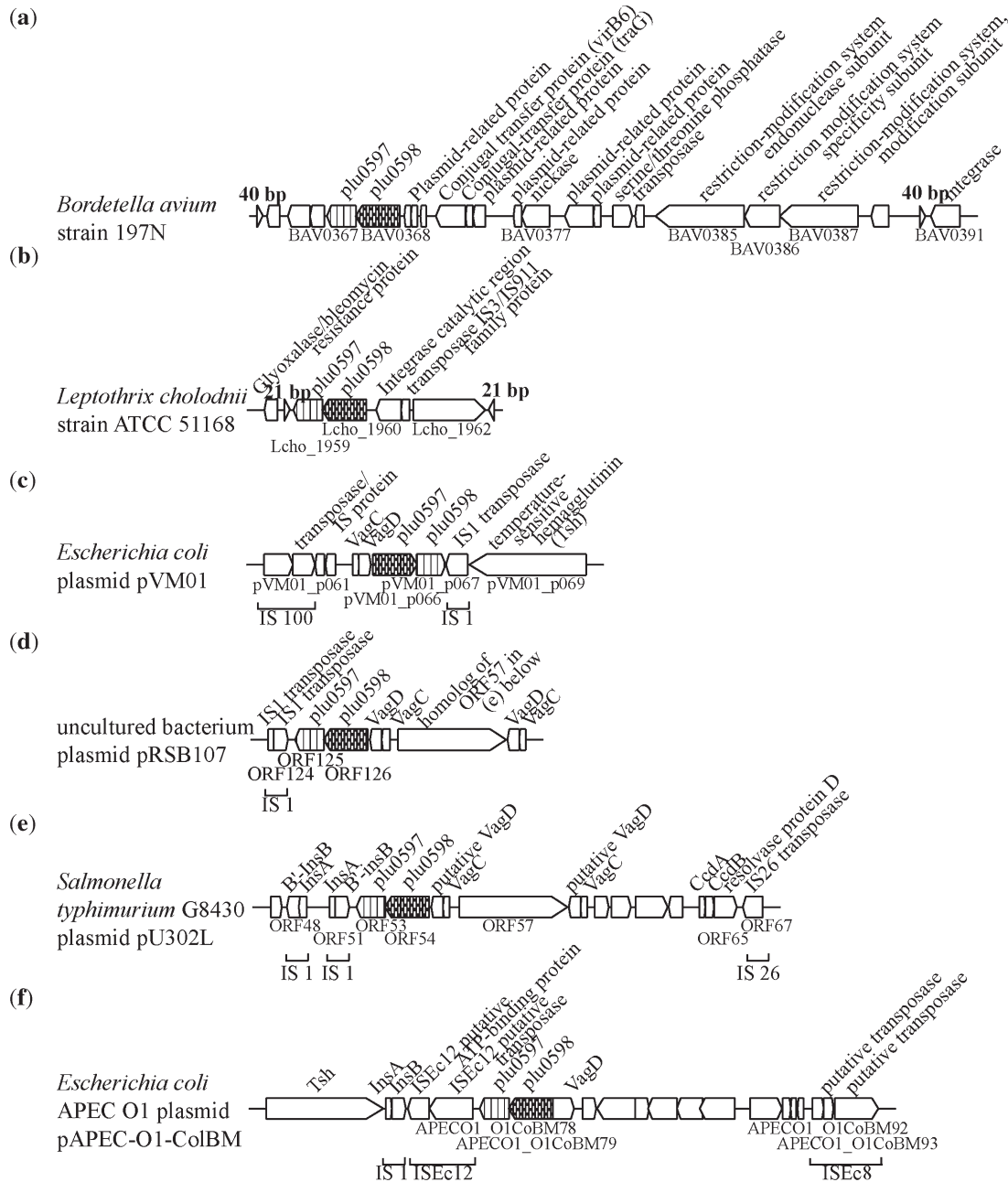


Figure 6. Genome contexts of plu0597 and plu0598 homologs (shaded boxes) in eubacterial genomes. Repeat sequences are drawn as triangles. The accession numbers of the sources are: (a) Refseq NC_010645; (b) Refseq NC_010524; (c) Refseq NC_010409; (d) Genbank AJ851089; (e) Refseq NC_006816; (f) Refseq NC_009837.

plasmid integrated into the 3'-terminus of a tRNA gene resulting in 40-bp long flanking direct repeats. Another homologous pair (Figure 6b) was flanked by 21-bp inverted repeats of 19/21 bp identity next to a transposase homolog. This structure is similar to those of transposons such as Tn3 in which a transposase and a resolvase catalyze site-specific recombination for cointegrated resolution (51). In three other cases (Figure 6c–e), the homologous pair is next to an *IS1* copy. The relative position and orientation of the homologous pair and *IS1* are the same in these three clusters, suggesting that they form a mobile unit or a composite transposon. In these three cases (Figure 6c–e), the homologous pair is also adjacent to the VagCD system, a toxin–antitoxin system (52). In another case (Figure 6f), the homologous pair lies next to a *vagD* homolog, but its linkage to an *IS1* copy is interrupted by insertion of *ISEc12*.

DISCUSSION

We have identified a novel putative mobile genetic element that is flanked by direct repeats using genome comparison and genome context analysis. This element encodes a DNA methyltransferase, a restriction nuclease lacking detectable sequence similarity to any characterized protein and showing only remote secondary structure similarity to PD-(D/E)XK enzymes, and two other genes of unknown functions. The restriction enzyme, designated R.Pluti, specifically cleaved the sequence 5'-GGCGC↓C-3'. Kinetic analysis and gel filtration suggested it is a tetrameric Type IIF enzyme. The DNA methyltransferase M.Pluti modified DNA to protect it from cleavage by R.Pluti. Thus, R.Pluti and M.Pluti form an RM system. In light of all the evidence, this unit likely represents a mobile element composed of an RM system with the plu0597/plu0598 gene pair that is integrated into the genome.

Reaction mechanism of R.Pluti

R.Pluti showed cleavage preference for a two-site SC substrate over one-site SC substrates. It cleaved both sites in a two-site SC substrate in a concerted manner (Figure 3d), which indicates that this is a Type IIF enzyme (13). R.Pluti cleaved only one bond in the one-site SC plasmid, liberating OC product at a higher rate than completely cleaved FLL product (Figure 3b). These kinetics are similar to those of NgoMIV, a Type IIF restriction enzyme (19). Moreover, addition of an oligoduplex carrying a recognition site to the reaction revealed some interesting features. Similar to other Type IIF enzymes (39), the oligoduplex enhanced cleavage of a one-site SC plasmid by R.Pluti (Figure 4a), but in the initial phase of the reaction, generation of OC unexpectedly superceded the linear form. This suggested that reactions *in trans*, even in the absence of any apparent steric hindrance, are not as efficient as reactions *in cis* that cleaved all the four bonds. This observation indicated that the reactions *in cis*, along with superhelicity, may provide something more than just close bridging of two sites.

The order of the V_{max} values obtained, two-site SC > two-site linear = one-site linear >> one-site SC (Table 2), supported a role for superhelicity and *cis–trans* effects in reactions catalyzed by this Type IIF enzyme. Superhelicity appeared to facilitate reaction at two sites *in cis*, but suppress reaction at two sites *in trans*.

Life cycle of the mobile element

Our genome comparison (Table 1) and genome context analysis (Figure 6) strongly suggested that the unit we describe represents a mobile element. The unit may have integrated into the genome by conservative site-specific recombination similar to bacteriophage integration or transposition. In the latter case, this unit may have integrated into the genome as a composite transposon composed of four genes in between two IS982 copies. The ISs appear from the sequence to be nonfunctional.

What might be the role of the neighboring two genes, plu0597 and plu0598? The proteins of plu0597 family have similarity to helicases, while the plu0598 family proteins have similarity to OLD family nucleases. The plu0597/plu0598 pair might encode a helicase/nuclease pair similar to what is found in the McrBC Type IV restriction system. An alternative, but not mutually exclusive, model is that plu0597 and plu0598 form a novel type of toxin–antitoxin systems that stabilizes the unit. The genome context of gene plu0598, and the similarity of its product to nucleases, together with the conservation of tight linkage of plu0597 and plu0598 and the frequent association with a mobile element (Figure 6), are compatible with the hypothesis that this gene pair represents a typical toxin–antitoxin system composed of a gene for an RNase toxin tightly linked with an antitoxin gene (53). Nonetheless, we cannot eliminate the possibility that plu0597/plu0598 might encode a recombinase involved in integrating the unit. The nuclease domain detected in plu0598 is associated with both DNase and RNase activities, so we cannot distinguish between these possibilities. Further experimental characterization is required to test these hypotheses.

No other genome sequences are available for other *P. luminescence* strain at present. In the future, comparison with more closely related genomes might reveal an empty site and provide more information about the evolutionary history of the Pluti RM system and the life cycle of the putative mobile element.

Further experiments are necessary to determine if this element can integrate into a genome by the action of the encoded proteins. We do not yet know whether the Pluti RM system is just a passenger in this vehicle or actively participates in its mobility and other phases of the life cycle of this element.

Search for novel structures in DNA-interacting proteins and novel mobile elements

This work employed a systematic genome analysis to detect variable subgenomic units that might contain restriction enzyme genes. Variable units carrying a DNA methyltransferase gene homolog were selected by genome comparison and genome context analysis and neighbor

genes screened for the restriction activity after *in vitro* expression. This method depended on the sequence similarity among the DNA methyltransferases rather than among the restriction endonucleases, so it was expected to be capable of detecting restriction enzymes not homologous to previously characterized enzymes, and possibly with a novel structure. The precedent of this procedure discovered a novel fold using pair-wise genome comparison (22–24). This approach should also allow discovery of many more novel mobile elements, which may or may not contain restriction enzymes, using the rapidly expanding availability of genome sequence information. The details of the bioinformatics procedure will be published elsewhere (M.K., J.M.B., K.H.K. and I.K., unpublished data).

R.PluTI is the first enzyme with the cleavage site 5'-GG CGC↓C-3' for which amino acid sequence information is publicly available. Sequence is also available for two neoschizomers, R.KasI (5'-G↓GCGCC-3') and R.SfoI (5'-GGC↓GCC-3'), opening the possibility for future detailed analysis of sequence–structure–function relationships (20).

The R.PluTI family includes the protein Maqu_1714 from *Marinobacter aquaeolei* VT8, which is shorter than other members. Its sequence corresponds to the C-terminal part of other family members (residues 147–322 of R.PluTI). Incidentally, the secondary structure of this region bears remote resemblance to the pattern characteristic of the PD-(D/E)XK fold (54), namely $\alpha 1 - \beta 2 - \beta 3 - \beta 4 - \alpha 2 - \beta 5 \dots$, with $\beta 3$ preceded by a conserved dipeptide 210-GD-211 and a semi-conserved motif 220-HX(T/S)-222 (numbering for R.PluTI) in positions corresponding to the PD and (D/E)XK half-motifs. While the D residue is almost invariant among PD-(D/E)XK enzymes, charged residues from the (D/E)XK motif are not, and can be substituted with many other residues (e.g. FTH in the Vsr nuclease), or 'migrate' to non-homologous positions in the sequence (55). However, while 'sequence gazing' might prompt us to speculate that R.PluTI might contain a C-terminal PD-(D/E)XK domain, possibly with a new type of active site, at this point, none of the state-of-the-art bioinformatics methods applied provided statistically significant support for this prediction. Thus, R.PluTI may be unrelated to PD-(D/E)XK nucleases and may exhibit a different novel fold.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Alain Givaudan (Université Montpellier) for generous gifts of genomic DNA. Computer resources were provided by the supercomputer system at the Human Genome Center, Institute of Medical Science, University of Tokyo.

FUNDING

The 21st century Center of Excellence (COE) program for 'Elucidation of Language Structure and Semantic behind Genome and Life System' and the global COE program for "Deciphering Biosphere from Genome Big Bang" from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) to I.K.; 'Grants-in-Aid for Scientific Research' from Japan Society for the Promotion of Science (JSPS) (19657002, 21370001 to I.K.; and 19·07420 to F.K.); 'Postdoctoral Fellowship for Foreign Researchers' from JSPS to F.K.; Postdoctoral Fellowships from the 21st century COE project of 'Elucidation of Language Structure and Semantic behind Genome and Life System' and Medical Genome Science Program in Support Program for Improving Graduate School Education of JSPS to M.K.; short-term fellowship from EMBO and by an exchange grant from JSPS and Polish Academy of Sciences (to K.H.K.); National Institutes of Health (1R01GM081680-01 to J.M.B.) in part. Funding for open access charge: The University of Tokyo.

Conflict of interest statement. None declared.

REFERENCES

- Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2009) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.*, **38**, D234–D236.
- Kobayashi,I. (2004) Restriction-modification systems as minimal forms of life. In Pingoud,A. (ed.), *Restriction Endonucleases*. Springer, Berlin, pp. 19–62.
- van Zyl,L.J., Deane,S.M., Louw,L.A. and Rawlings,D.E. (2008) Presence of a family of plasmids (29 to 65 kilobases) with a 26-kilobase common region in different strains of the sulfur-oxidizing bacterium *Acidithiobacillus caldus*. *Appl. Environ. Microbiol.*, **74**, 4300–4308.
- Stiens,M., Becker,A., Bekel,T., Godde,V., Goemann,A., Niehaus,K., Schneiker-Bekel,S., Selbitschka,W., Weidner,S., Schluter,A. *et al.* (2008) Comparative genomic hybridisation and ultrafast pyrosequencing revealed remarkable differences between the *Sinorhizobium meliloti* genomes of the model strain Rm1021 and the field isolate SM11. *J. Biotechnol.*, **136**, 31–37.
- Furuta,Y., Abe,K. and Kobayashi,I. (2010) Genome comparison and context analysis reveals putative mobile forms of restriction-modification systems and related rearrangements. *Nucleic Acids Res.*, doi:10.1093/nar/gkp1226.
- Nobusato,A., Uchiyama,I., Ohashi,S. and Kobayashi,I. (2000) Insertion with long target duplication: a mechanism for gene mobility suggested from comparison of two related bacterial genomes. *Gene*, **259**, 99–108.
- Kobayashi,I. (2001) Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res.*, **29**, 3742–3756.
- Naito,T., Kusano,K. and Kobayashi,I. (1995) Selfish behavior of restriction-modification systems. *Science*, **267**, 897–899.
- Orlowski,J. and Bujnicki,J.M. (2008) Structural and evolutionary classification of Type II restriction enzymes based on theoretical and experimental analyses. *Nucleic Acids Res.*, **36**, 3552–3569.
- Roberts,R.J., Belfort,M., Bestor,T., Bhagwat,A.S., Bickle,T.A., Bitinaite,J., Blumenthal,R.M., Degtyarev,S.K., Dryden,D.T.F., Dybvig,K. *et al.* (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.*, **31**, 1805–1812.
- Pingoud,A., Fuxreiter,M., Pingoud,V. and Wende,W. (2005) Type II restriction endonucleases: structure and mechanism. *Cell Mol. Life Sci.*, **62**, 685–707.

12. Asakura, Y. and Kobayashi, I. (2009) From damaged genome to cell surface: transcriptome changes during bacterial cell death triggered by loss of a restriction-modification gene complex. *Nucleic Acids Res.*, **37**, 3021–3031.
13. Mucke, M., Kruger, D.H. and Reuter, M. (2003) Diversity of type II restriction endonucleases that require two DNA recognition sites. *Nucleic Acids Res.*, **31**, 6079–6084.
14. Deibert, M., Grazulis, S., Sasnauskas, G., Siksnys, V. and Huber, R. (2000) Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat. Struct. Biol.*, **7**, 792–799.
15. Bozic, D., Grazulis, S., Siksnys, V. and Huber, R. (1996) Crystal structure of *Citrobacter freundii* restriction endonuclease Cfr10I at 2.15 Å resolution. *J. Mol. Biol.*, **255**, 176–186.
16. Grazulis, S., Deibert, M., Rimseliene, R., Skirgaila, R., Sasnauskas, G., Lagunavicius, A., Repin, V., Urbanke, C., Huber, R. and Siksnys, V. (2002) Crystal structure of the Bse634I restriction endonuclease: comparison of two enzymes recognizing the same DNA sequence. *Nucleic Acids Res.*, **30**, 876–885.
17. Siksnys, V., Skirgaila, R., Sasnauskas, G., Urbanke, C., Cherny, D., Grazulis, S. and Huber, R. (1999) The Cfr10I restriction enzyme is functional as a tetramer. *J. Mol. Biol.*, **291**, 1105–1118.
18. Wentzell, L.M. and Halford, S.E. (1998) DNA looping by the Sfi I restriction endonuclease. *J. Mol. Biol.*, **281**, 433–444.
19. Embleton, M.L., Siksnys, V. and Halford, S.E. (2001) DNA cleavage reactions by type II restriction enzymes that require two copies of their recognition sites. *J. Mol. Biol.*, **311**, 503–514.
20. Gowers, D.M., Bellamy, S.R. and Halford, S.E. (2004) One recognition sequence, seven restriction enzymes, five reaction mechanisms. *Nucleic Acids Res.*, **32**, 3469–3479.
21. Sokolowska, M., Czapinska, H. and Bochtler, M. (2009) Crystal structure of the $\beta\beta\alpha$ -Me type II restriction endonuclease Hpy99I with target DNA. *Nucleic Acids Res.*, **37**, 3799–3810.
22. Ishikawa, K., Watanabe, M., Kuroita, T., Uchiyama, I., Bujnicki, J.M., Kawakami, B., Tanokura, M. and Kobayashi, I. (2005) Discovery of a novel restriction endonuclease by genome comparison and application of a wheat-germ-based cell-free translation assay: PabI (5'-GTA/C) from the hyperthermophilic archaeon. *Pyrococcus abyssi*. *Nucleic Acids Res.*, **33**, e112.
23. Chinen, A., Uchiyama, I. and Kobayashi, I. (2000) Comparison between *Pyrococcus horikoshii* and *Pyrococcus abyssi* genome sequences reveals linkage of restriction-modification genes with large genome polymorphisms. *Gene*, **259**, 109–121.
24. Miyazono, K., Watanabe, M., Kosinski, J., Ishikawa, K., Kamo, M., Sawasaki, T., Nagata, K., Bujnicki, J.M., Endo, Y., Tanokura, M. et al. (2007) Novel protein fold discovered in the PabI family of restriction enzymes. *Nucleic Acids Res.*, **35**, 1908–1918.
25. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, **33**, 103–119.
26. Duchaud, E., Rusniok, C., Frangeul, L., Buchrieser, C., Givaudan, A., Taourit, S., Bocs, S., Boursaux-Eude, C., Chandler, M., Charles, J.F. et al. (2003) The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nat. Biotechnol.*, **21**, 1307–1313.
27. Guzman, L.M., Belin, D., Carson, M.J. and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.*, **177**, 4121–4130.
28. Sawasaki, T., Ogasawara, T., Morishita, R. and Endo, Y. (2002) A cell-free protein synthesis system for high-throughput proteomics. *Proc. Natl. Acad. Sci. USA*, **99**, 14652–14657.
29. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
30. Soding, J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics*, **21**, 951–960.
31. Kurowski, M.A. and Bujnicki, J.M. (2003) GeneSilico protein structure prediction meta-server. *Nucleic Acids Res.*, **31**, 3305–3307.
32. Komori, K., Fujita, N., Ichiyangi, K., Shinagawa, H., Morikawa, K. and Ishino, Y. (1999) PI-PfuI and PI-PfuII, intein-coded homing endonucleases from *Pyrococcus furiosus*. I. Purification and identification of the homing-type endonuclease activities. *Nucleic Acids Res.*, **27**, 4167–4174.
33. Williamson, V.M. and Kaya, H.K. (2003) Sequence of a symbiont. *Nat. Biotechnol.*, **21**, 1294–1295.
34. Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
35. Yu, W., Mierau, I., Mars, A., Johnson, E., Dunny, G. and McKay, L.L. (1995) Novel insertion sequence-like element IS982 in lactococci. *Plasmid*, **33**, 218–225.
36. Gingeras, T.R., Milazzo, J.P. and Roberts, R.J. (1978) A computer assisted method for the determination of restriction enzyme recognition sites. *Nucleic Acids Res.*, **5**, 4105–4127.
37. Taylor, J.D. and Halford, S.E. (1992) The activity of the EcoRV restriction endonuclease is influenced by flanking DNA sequences both inside and outside the DNA-protein complex. *Biochemistry*, **31**, 90–97.
38. Wentzell, L.M., Nobbs, T.J. and Halford, S.E. (1995) The SfiI restriction endonuclease makes a four-strand DNA break at two copies of its recognition sequence. *J. Mol. Biol.*, **248**, 581–595.
39. Gormley, N.A., Hillberg, A.L. and Halford, S.E. (2002) The type II restriction endonuclease BspMI is a tetramer that acts concertedly at two copies of an asymmetric DNA sequence. *J. Biol. Chem.*, **277**, 4034–4041.
40. Oller, A.R., Vanden Broek, W., Conrad, M. and Topal, M.D. (1991) Ability of DNA and spermidine to affect the activity of restriction endonucleases from several bacterial species. *Biochemistry*, **30**, 2543–2549.
41. Zebala, J.A., Choi, J. and Barany, F. (1992) Characterization of steady state, single-turnover, and binding kinetics of the TaqI restriction endonuclease. *J. Biol. Chem.*, **267**, 8097–8105.
42. Kaminska, K.H., Kawai, M., Boniecki, M., Kobayashi, I. and Bujnicki, J.M. (2008) Type II restriction endonuclease R.Hpy188I belongs to the GIY-YIG nuclease superfamily, but exhibits an unusual active site. *BMC Struct. Biol.*, **8**, 48.
43. Saravanan, M., Bujnicki, J.M., Cymerman, I.A., Rao, D.N. and Nagaraja, V. (2004) Type II restriction endonuclease R.KpnI is a member of the HNH nuclease superfamily. *Nucleic Acids Res.*, **32**, 6129–6135.
44. Bujnicki, J.M. and Rychlewski, L. (2001) Grouping together highly diverged PD-(D/E)XK nucleases and identification of novel superfamily members using structure-guided alignment of sequence profiles. *J. Mol. Microbiol. Biotechnol.*, **3**, 69–72.
45. Sapranaukas, R., Sasnauskas, G., Lagunavicius, A., Vilkaitis, G., Lubys, A. and Siksnys, V. (2000) Novel subtype of type II restriction enzymes. BfiI endonuclease exhibits similarities to the EDTA-resistant nuclease Nuc of *Salmonella typhimurium*. *J. Biol. Chem.*, **275**, 30878–30885.
46. Hwang, S., Gou, Z. and Kuznetsov, I.B. (2007) DP-Bind: a web server for sequence-based prediction of DNA-binding residues in DNA-binding proteins. *Bioinformatics*, **23**, 634–636.
47. Wang, L. and Brown, S.J. (2006) BindN: a web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. *Nucleic Acids Res.*, **34**, W243–W248.
48. Marchler-Bauer, A., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M. et al. (2009) CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res.*, **37**, D205–D210.
49. Koonin, E.V. and Gorbalenya, A.E. (1992) The superfamily of UvrA-related ATPases includes three more subunits of putative ATP-dependent nucleases. *Protein Seq. Data Anal.*, **5**, 43–45.
50. Myung, H. and Calendar, R. (1995) The old exonuclease of bacteriophage P2. *J. Bacteriol.*, **177**, 497–501.
51. Heffron, F., McCarthy, B.J., Ohtsubo, H. and Ohtsubo, E. (1979) DNA sequence analysis of the transposon Tn3: three genes and three sites involved in transposition of Tn3. *Cell*, **18**, 1153–1163.
52. Radnedge, L., Davis, M.A., Youngren, B. and Austin, S.J. (1997) Plasmid maintenance functions of the large virulence plasmid of *Shigella flexneri*. *J. Bacteriol.*, **179**, 3670–3675.
53. Zhang, Y., Zhang, J., Hoeflich, K.P., Ikura, M., Qing, G. and Inouye, M. (2003) MazF cleaves cellular mRNAs specifically at

- ACA to block protein synthesis in *Escherichia coli*. *Mol. Cell*, **12**, 913–923.
54. Bujnicki, J.M. (2003) Crystallographic and bioinformatic studies on restriction endonucleases: inference of evolutionary relationships in the "midnight zone" of homology. *Curr. Protein Pept. Sci.*, **4**, 327–337.
55. Kosinski, J., Feder, M. and Bujnicki, J.M. (2005) The PD-(D/E)XK superfamily revisited: identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function. *BMC Bioinformatics*, **6**, 172.
56. Pei, J. and Grishin, N.V. (2007) PROMALS: towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics*, **23**, 802–808.