The *Helicobacter pylori* HpyAXII restriction–modification system limits exogenous DNA uptake by targeting GTAC sites but shows asymmetric conservation of the DNA methyltransferase and restriction endonuclease components

Olivier Humbert[1,2] and Nina R. Salama[1,*]

[1]Division of Human Biology, Fred Hutchinson Cancer Research Center and [2]Program in Molecular and Cellular Biology, University of Washington, Seattle, WA, USA

ABSTRACT

The naturally competent organism *Helicobacter pylori* encodes a large number of restriction–modification (R–M) systems that consist of a restriction endonuclease and a DNA methyltransferase. R–M systems are not only believed to limit DNA exchange among bacteria but may also have other cellular functions. We report a previously uncharacterized *H. pylori* type II R–M system, M.HpyAXII/ R.HpyAXII. We show that this system targets GTAC sites, which are rare in the *H. pylori* chromosome but numerous in ribosomal RNA genes. As predicted, this type II R–M system showed attributes of a selfish element. Deletion of the methyltransferase M.HpyAXII is lethal when associated with an active endonuclease R.HpyAXII unless compensated by adaptive mutation or gene amplification. R.HpyAXII effectively restricted both unmethylated plasmid and chromosomal DNA during natural transformation and was predicted to belong to the novel 'half pipe' structural family of endonucleases. Analysis of a panel of clinical isolates revealed that R.HpyAXII was functional in a small number of *H. pylori* strains (18.9%, *n* = 37), whereas the activity of M.HpyAXII was highly conserved (92%, *n* = 50), suggesting that GTAC methylation confers a selective advantage to *H. pylori*. However, M.HpyAXII activity did not enhance *H. pylori* fitness during stomach colonization of a mouse infection model.

INTRODUCTION

Helicobacter pylori, a Gram-negative human pathogen, colonizes the stomach and is associated with several gastrointestinal disorders including peptic ulcers and gastric cancers (1). Over half of the world's population is estimated to harbor *H. pylori*, but only 15–20% of those infected develop severe disease. *Helicobacter pylori* is known for its remarkably high level of genetic diversity. A comparison of the fully sequenced genomes of three clinical isolates revealed substantial differences in chromosomal structure and a large number of genes that are variably present or absent among strains (up to 25% of the total coding regions) (2,3). These observations have raised many questions, including how genetic diversity is generated in *H. pylori* and how variable genes contribute to disease outcome.

One class of highly variable genes are restriction–modification (R–M) systems (2), which are typically composed of both a restriction endonuclease (REase) and a cognate DNA methyltransferase (MTase). Generally, the REase cleaves a specific DNA sequence unless it is methylated by the cognate MTase. R–M systems have been divided in type I, II or III, according to their cofactor requirement, molecular structure, mode of action and structure of the recognition sequence site (4). Typical type II R–M systems consist of two independent enzymes that act on the same sequence, usually a 4- to 8-bp palindrome. The recent sequencing of more than 400 bacterial and archaeal genomes has revealed that ∼90% of the genomes contain at least one R–M system and over 80% contain multiple R–M systems (5), but only a small fraction have been

*To whom correspondence should be addressed. Tel: +1 206 667 1540; Fax: +1 206 667 6524; Email: nsalama@fhcrc.org

comprehensively studied. The *H. pylori* chromosome encodes a large number of active and inactive R–M systems. Analysis of the sequenced genomes of *H. pylori* strain 26695 and J99 predicted 26 putative R–M systems including 14 and 16 type II R–M systems, respectively. Comparative genomics of these systems suggests they have a high degree of sequence heterogeneity and that only a small fraction retained full activity (6,7). Interestingly, for some R–M systems, MTase activity may be preserved despite inactivation of the respective REase.

Although still under debate, R–M systems may carry out several biological roles. They have traditionally been implicated in bacterial defense from bacteriophage infection, which have in turn evolved a variety of anti-restriction defense mechanisms (8). R–M systems also regulate genetic exchange among bacteria, particularly relevant for bacteria like *H. pylori* that are naturally competent for DNA uptake (9). By constituting an effective barrier against DNA transformation in a bacterial population, R–M systems may provide the genetic isolation required for adaptation of the organisms to a special ecological niche (10). Paradoxically, R–M systems were also suggested to promote homologous recombination from the small DNA fragments produced by the incomplete digestion of incoming DNA by REases (11). Kobayashi and colleagues have suggested that R–M systems have attributes of selfish genes, and may act as invasive elements independent of any selective advantage to their host (12,13). This hypothesis emerged from the findings that several R–M systems are horizontally transferred and that a fully active type II R–M system cannot easily be lost from its host because of postsegregational killing. Finally, DNA methylation may have important biological properties independent of the endonuclease function by regulating the interactions between proteins and DNA, in a similar manner to the stand-alone DNA methyltransferase Dam in both α- and γ-Proteobacteria (14,15).

In *H. pylori*, R–M systems may carry out other biological roles. Several R–M systems have been classified as 'chronic atrophic gastritis (ChAG) associated genes' by comparing the genome content of eight ChAG strains to a *H. pylori* 'pangenome' defined previously (3,16). Other R–M genes were acid-regulated (3), undergo phase variation (17,18) and several R–M systems were identified as candidate colonization factors using a mouse model of infection (19). Expression of the REase encoded by gene *iceA1* was upregulated upon contact with gastric epithelial cells and its presence was associated with peptidic ulcer disease (20). A mutant in the endonuclease R.HpyC1I, isolated from *H. pylori* clinical isolate NTUH-C1, exhibited cell elongation and decreased adherence to a gastric cell line (21). In another report, DNA methylation by M.HpyAIV was shown to alter transcription of a selected number of genes including catalase (*katA*) (22).

Here, we annotate a previously uncharacterized *H. pylori* R–M system M.HpyAXII/R.HpyAXII encoded by the locus HP0502/0503/0504/0505. We present biochemical and genetic analyses of this R–M system. We additionally explored its conservation among clinical isolates and its role during establishment of infection to query alternative biological functions of M.HpyAXII/R.HpyAXII.

MATERIALS AND METHODS

Bacterial strains and growth conditions

Helicobacter pylori strains (Supplementary Table S1) were grown on solid horse blood agar (HB) plates containing 4% Columbia agar base (Oxoid, Hampshire, UK), 5% defibrinated horse blood (HemoStat Labs, Dixon, CA), 0.2% β-cyclodextrin (Sigma, St Louis, MO), 10 μg/ml vancomycin (Sigma), 5 μg/ml cefsulodin (Sigma), 2.5 U/ml polymyxin B (Sigma), 5 μg/ml trimethoprim (Sigma) and 8 μg/ml amphotericin B (Sigma) at 37°C either under a microaerobic atmosphere generated using a CampyGen sachet (Oxoid) in a gas pack jar or in an incubator equilibrated with 14% CO_2 and 86% air. For liquid culture, *H. pylori* was grown in Brucella broth (Difco, Franklin Lakes, NJ) containing 10% fetal bovine serum (BB10; Invitrogen, Carlsbad, CA) with shaking in a gas pack jar containing a CampyGen sachet. For resistance marker selections, bacterial media were additionally supplemented with 15 μg/ml chloramphenicol (Cm, Sigma), 25 μg/ml kanamycin (Kan, Fisher Scientific, Pittsburgh, PA) or 60 mg/ml sucrose (Fisher Scientific). When culturing bacteria from mouse stomachs, 200 μg/ml bacitracin (Sigma) was added to eliminate normal mouse microbiota contamination.

DNA manipulations

DNA manipulations, such as restriction digestion, PCR and agarose gel electrophoresis, were performed according to standard procedures (23). *Helicobacter pylori* genomic DNA was prepared by Wizard genomic DNA preparation kits (Promega, Madison, WI). Primers used for PCR and sequencing are in Supplementary Table S2. Plasmid DNA (Supplementary Table S3) was isolated and prepared from *Escherichia coli* or *H. pylori* using Qiagen Midiprep kit (Qiagen, Valencia, CA). Sequencing of plasmid DNA and PCR products was performed by the FHCRC Genomics Shared Resource and the resulting sequences analyzed using Sequencher (Gene Codes Corporation, Ann Harbor, MI).

Generation of *H. pylori* knockout isogenic mutants

Knockout alleles were constructed using a vector-free allelic replacement strategy to generate alleles in which a chloramphenicol acetyl transferase (*cat*) resistance cassette (24), a nonpolar kanamycin resistance (*aphA3*) cassette (25), or a *cat* cassette fused to a sucrose sensitivity marker (*sacB*) (26) replaced 80–90% of the coding sequence of the gene while preserving the start and stop codons. The primers used for this procedure are designated as 1 through 4 and are given in Supplementary Table S2. The *catsacB* cassette was generated by fusing *cat* to *sacB* with elimination of the terminator sequence using PCR (primers C2SA4 and SB1). After natural transformation (27) with the appropriate PCR product and selection on Cm- or Kan- containing media, four clones

were evaluated by PCR to confirm replacement of the wild-type allele with the null allele. Urease activity and flagella-based motility were also confirmed for all the clones generated. Single clones were used for infection experiments and phenotypic characterization.

Generation of *H. pylori* knockin isogenic mutants

Null alleles generated by replacement with the *catsacB* cassette were subsequently replaced with alternate alleles by sucrose counter-selection. *H. pylori* strain NSH57 $\Delta hpyAXIIM/R::catsacB$ was naturally transformed with the appropriate plasmid or genomic DNA, and recombinants were selected on sucrose-containing HB plates, screened on Cm-containing media, and evaluated by PCR to confirm they lost the *catsacB* cassette.

Site-directed mutagenesis

To generate a catalytically inactive DNA MTase in *H. pylori* strain NSH57, an asparagine in conserved motif IV was replaced by alanine (N87A). A DNA fragment including *hpyAXIIM/R* was amplified from NSH57 by PCR and cloned into TA vector pCR-TOPO (TOPO cloning, Invitrogen). pCR-TOPO/*hpyAXII* was digested with EcoRI (NEB, Ipswich, MA) and subcloned into plasmid Bluescript SK + (Stratagene, La Jolla, CA). Site-directed mutagenesis was performed on *hpyAXIIM* using the QuikChange kit (Stratagene) following the manufacturer instructions and confirmed by sequencing.

Natural transformation

H. pylori bacteria were freshly grown for 24–32 h on HB plates, transferred as patches onto fresh plates and grown for an additional 6–8 h. Donor DNA (plasmid, genomic or PCR product) was diluted as appropriate in distilled water and 10 µl was added to each patch and incubated overnight. The transformation mixture was harvested from the plate surface and resuspended in 350 µl phosphate buffer saline (PBS). Serial 10-fold dilutions were inoculated onto selective or nonselective HB plates. Transformation frequency was calculated as the number of Kan or Cm resistant colonies per recipient cell per nanogram of donor DNA.

Expression of M.HpyAXII and R.HpyAXII proteins

hpyAXIIM was amplified from DNA of wild-type NSH57 and NSH79 *H. pylori* by PCR and cloned into a pET-15b vector (Novagen, Gibbstown, NJ) in frame with the N-terminal His-Tag sequence using NdeI and XhoI (New England Biolabs). The resulting pET15b/*hpyAXIIM* plasmids were transformed into *E. coli* strain ER2925(DE3), derived from strain ER2925 [NEB and (28)] and engineered to become a λDE3 lysogen using a lysogenization kit (Novagen). Plasmid pET15b/*hpyAXIIR* was transformed into *E. coli* strain BL21CodonPlus(DE3)-RIL. Protein expression was induced using 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) induction for 3 h at 37°C for M.HpyAXII and overnight at 16°C for R.HpyAXII to minimize endonuclease activity. Bacterial pellets were resuspended in lysis buffer (50 mM Tris pH 7.9, 100 mM NaCl, 1 mM DTT, 2% glycerol), incubated with 1 mg/ml lysozyme (Sigma) for 30 min on ice, and sonicated. For R.HpyAXII, soluble and insoluble fractions were separated by centrifugation at 18 000 r.p.m. for 20 min at 4°C.

Antibody-based M.HpyAXII methylation activity assays

M. HpyAXII DNA MTase activity was assessed in a dot-blot assay using a rabbit primary antibody raised against DNA with N6-methyladenine (m6A) (29). *Escherichia coli* ER2925(DE3) transformed with pET15b and pET15b/ *hpyAXIIM* was grown to log phase and induced with 1 mM IPTG for 3 h at 37°C. Genomic DNA of the cultures was isolated and 3 µl cross-linked to a nylon membrane (Hybond-XL, Amersham Biosciences, Piscataway, NJ) by UV (1.2 mJ/cm^2 for 30 s). The membrane was blocked in a 3% nonfat dry milk solution and incubated with the m6A polyclonal rabbit antibody (1:10 000 dilution) overnight at 4°C. After a series of washes in PBST (PBS and 0.1% Tween-20), anti-rabbit IgG antibody conjugated to horseradish peroxidase (1:1000 dilution) was added to the membrane and detected using ECL Western Blotting reagent (Amersham Biosciences).

Biochemical R.HpyAXII endonuclease activity assay

One microliter of freshly prepared soluble fraction obtained from *E. coli* BL21(DE3) expressing R.HpyAXII was incubated for 3 h at 37°C with 500 ng *E. coli* ER2925 genomic DNA or pET15b plasmid DNA isolated from *E. coli* DH10B in presence of NEB buffer 1 (New England Biolabs). The product DNA was separated on a 0.8% agarose gel.

Genetic R.HpyAXII endonuclease activity assay

To assess R.HpyAXII activity in a large number of *H. pylori* strains, we replaced the null allele in *H. pylori* strain NSH57$\Delta hpyAXIIM/R::catsacB$ with that from other strains as described above (Generation of *H. pylori* knockin isogenic mutants section). R.HpyAXII activity was determined by transforming these *H. pylori* clones with PCR constructs $\Delta hpyXIIM::cat$ and $\Delta HP0368::cat$ and enumerating recombinants recovered on Cm- containing HB plates.

Mouse infections

Female C57BL/6 mice 24- to 28-days old were obtained from Charles River Laboratories and certified free of endogenous *Helicobacter* infection by the vendor. The mice were housed in sterilized microisolator cages with irradiated PMI 5053 rodent chow, autoclaved corn cob bedding, and acidified, reverse-osmosis purified water provided ad libitum. All studies were done under practices and procedures of Animal Biosafety Level 2. The facility is fully accredited by the Association for Assessment and Accreditation of Laboratory Animal Care, International, and all activities were approved by the FHCRC Institutional Animal Care and Use Committee. Competition experiments were performed as described previously (24) with 5.0×10^8 bacteria of each strain in the inocula.

In silico genomic analysis

Helicobacter pylori sequences were retrieved from the Comprehensive Microbial Resource (CMR) (30). Homologs of M. and R.HpyAXII and were sought using REBASE Blast, a tool of REBASE (5). Protein structure prediction analysis for M.HpyAXII and R.HpyAXII was carried out using the Protein Homology/Analogy Recognition Engine PHYRE (31). Sequence alignments were carried out with CLUSTALW2 (32) and 3D molecular structures were visualized and modified using PyMOL (33). The genome-wide distribution and frequency of GTAC sites was determined using genome-scale DNA pattern in Regulatory Sequence Analysis (RSA) Tools (http://rsat.ulb.ac.be/rsat/). For *H. pylori* strain NSH57, the sequence of the derivative strain G27 was used (D.A. Baltrus, M.R. Amieva, A. Covacci, D.S. Merrell, K. M. Ottemann, M. Stein, N.R. Salama, K. Guillemin, personal communication) and genomic GTAC sites were analyzed with the genome tools/restriction digest option of the CMR website.

Sequences are available from the GenBank/EMBL/DDBJ databases with accession numbers FJ233107 to FJ233133.

RESULTS

Structure-based prediction uncovers the existence of a novel *H. pylori* R–M system

We set out to further characterize the HP0502/0503/0504/0505 locus that we previously identified in our genome-wide screen for candidate colonization factors (19). BLAST search of DNA sequences of these genes from *H. pylori* strain 26695 failed to identify any homologs. However, several lines of evidence suggested that HP0502/0503 encodes a DNA MTase and that it is organized in an R–M system along with the downstream gene, HP0504/0505. Nobusato *et al.* (7) previously predicted that this locus encodes a R–M system based on sequence analysis. In agreement with this finding, we used structure prediction algorithms of the program PHYRE (31) and found that the N-terminal domain of the HP0502/0503 polypeptides could be threaded onto a tertiary structure of the M.TaqI DNA MTase, crystallized from *Thermus aquaticus* (Supplementary Figure S1, *E*-value 6.1×10^{-15}). Despite weak protein sequence homology, several amino acid motifs were shared between HP0502/0503 and M.TaqI (Figure 1A) and mapped to the catalytic domain of M.TaqI (Supplementary Figure S1), further supporting the function of HP0502/0503 as an adenine MTase (34). More recently, the REase R.PabI from *Pyrococcus abyssi* was characterized and HP0504/0505 was identified as a homolog (28% protein sequence identity) and suggested to recognize the same target sequence, 5′-GTAC-3′ (35). Based on the proposed nomenclature for R–M systems (4), we have named this novel *H. pylori* R–M system M.HpyAXII/R.HpyAXII, encoded by genes *hpyAXIIM* and *hpyAXIIR*, respectively. Since R–M systems are highly polymorphic among *H. pylori* strains (6), we evaluated the genomic organization of

M.HpyAXII/R.HpyAXII in several isolates. As illustrated in Figure 1B, gene *hpyAXIIM* encodes a polypeptide of 331 amino acids and is predicted to be active in strains J99, HPAG1 and NSH57 but is likely inactive in 26695 and NSH79 where the expected protein products are truncated and fused with R.HpyAXII, respectively. Gene *hpyAXIIR* encodes a protein of 252 amino acids and is only expected to be functional in strains J99 and HPAG1 based on DNA sequences.

M.HpyAXII functions as a γ-adenine MTase

To validate the function of M.HpyAXII as a DNA MTase, we inducibly expressed Histidine-tagged M.HpyAXII from *H. pylori* strain NSH57 and Histidine-tagged M.HpyAXII fused to R.HpyAXII from strain NSH79 in *E. coli* (Figure 2A). M.HpyAXII activity was studied using *E. coli* strain ER2925(DE3) (Supplementary Table S1), defective for all known R–M systems, and presumably harboring chromosomal DNA devoid of adenine or cytosine methylation (Materials and methods section). Genomic DNA from the *E. coli* cultures expressing M.HpyAXII was isolated and adenine methylation assessed using antibodies against N6-methyladenine. Adenine MTase activity could be detected when M.HpyAXII from NSH57 but not NSH79 was expressed in *E. coli* (Figure 2B). The lack of MTase activity by NSH79 M.HpyAXII probably reflects the genetic rearrangements observed at the NSH79 locus (Figure 1B). To further confirm that M.HpyAXII acts as an adenine MTase, we used site-directed mutagenesis on the predicted catalytic site of this enzyme. We identified conserved motif IV (D/N/S)PP(Y/F) [Figure 1A and (34)] and the substitution of asparagine with alanine (point mutation designated N87A) in NSH57 M.HpyAXII completely abolished adenine MTase activity without affecting the expression levels of this protein (Supplementary Figure S2). Along with our genomic data, this finding supports the re-annotation of this *H. pylori* protein as a γ-adenine MTase.

The HpyAXII R-M system targets the tetramer 5′-GTAC-3′

Two independent approaches defined the target sequence of this novel R–M system. First, using an array of methylation-sensitive endonucleases, we sought to identify an enzyme that could digest *H. pylori* NSH57 chromosomal DNA isolated from a Δ*hpyAXIIM* mutant but not from wild-type. Only the enzyme RsaI digested DNA from the Δ*hpyAXIIM::cat* mutant but not wild-type (Figure 2C), indicating that the DNA MTase M.HpyAXII and the endonuclease RsaI act on the same target sequence, 5′-GTAC-3′, where the adenine residue is methylated by M.HpyAXII. Using this result, we investigated M.HpyAXII activity in the *H. pylori* strains where the HpyAXII R-M system was sequenced (Figure 1B). Genomic DNA from *H. pylori* strains 26695 and NSH79 could be digested with RsaI, whereas J99, NSH57 and HPAG1 DNA were protected from RsaI digestion (data not shown), indicating that M.HpyAXII is only active in these three strains and agreeing with our predictions from the genomic sequences. RsaI digestion of DNA isolated from the NSH57*hpyAXIIM*^N87A^ mutant showed
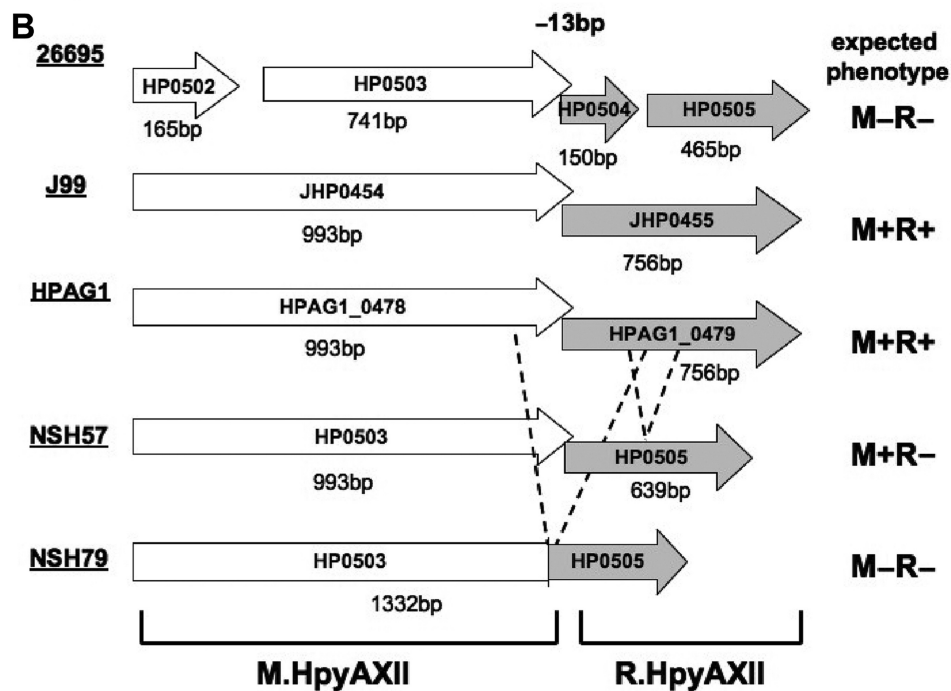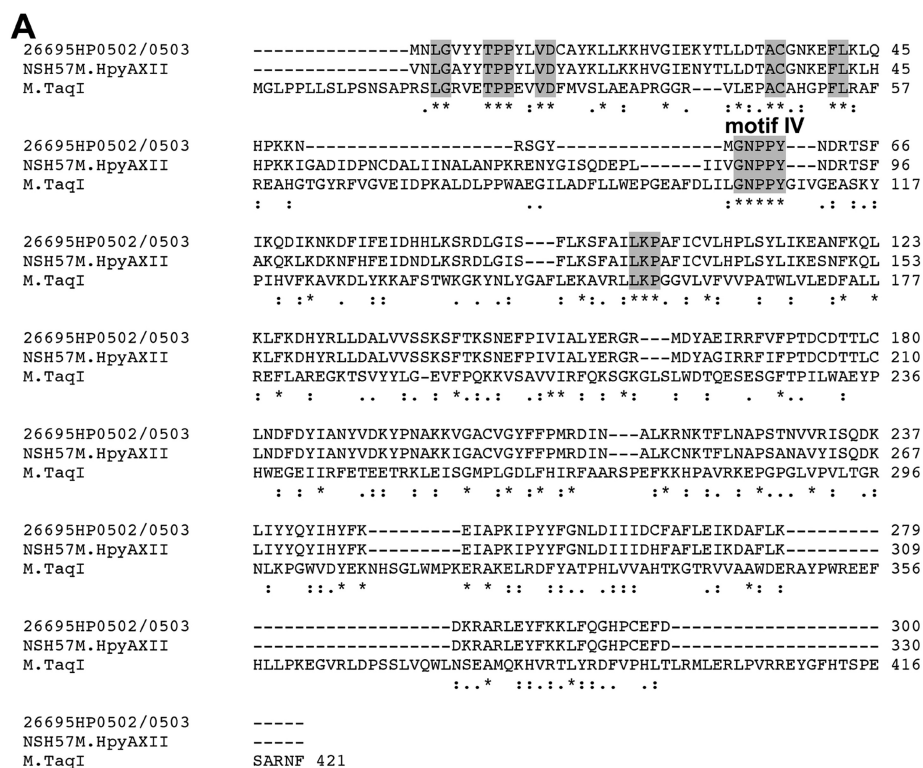
A

```
26695HP0502/0503   ---------------MNLGVYYTPPYLVDCAYKLLKKHVGIEKYTLLDTACGNKEFLKLQ  45
NSH57M.HpyAXII     ---------------VNLGAYYTPPYLVDYAYKLLKKHVGIENYTLLDTACGNKEFLKLH  45
M.TaqI             MGLPPLLSLPSNSAPRSLGRVETPPEVVDFMVSLAEAPRGGR---VLEPACAHGPFLRAF  57
                                  .**    ***.:**    .*  :   * .   .*:..**.:  **:
                                                                          motif IV
26695HP0502/0503   HPKKN-------------------RSGY----------------MGNPPY---NDRTSF  66
NSH57M.HpyAXII     HPKKIGADIDPNCDALIINALANPKRENYGISQDEPL------IIVGNPPY---NDRTSF  96
M.TaqI             REAHGTGYRFVGVEIDPKALDLPPWAEGILADFLLWEPGEAFDLILGNPPYGIVGEASKY  117
                   : :                       ..              :*****  .: :.:

26695HP0502/0503   IKQDIKNKDFIFEIDHHLKSRDLGIS---FLKSFAILKPAFICVLHPLSYLIKEANFKQL  123
NSH57M.HpyAXII     AKQKLKDKNFHFEIDNDLKSRDLGIS---FLKSFAILKPAFICVLHPLSYLIKESNFKQL  153
M.TaqI             PIHVFKAVKDLYKKAFSTWKGKYNLYGAFLEKAVRLLKPGGVLVFVVPATWLVLEDFALL  177
                   : :*  .  ::       . . .:    : *:. :***.: *:    :  :  :*  *

26695HP0502/0503   KLFKDHYRLLDALVVSSKSFTKSNEFPIVIALYERGR---MDYAEIRRFVFPTDCDTTLC  180
NSH57M.HpyAXII     KLFKDHYRLLDALVVSSKSFTKSNIEFPIVIALYERGR---MDYAGIRRFIFPTDCDTTLC  210
M.TaqI             REFLAREGKTSVYYLG-EVFPQKKVSAVVIRFQKSGKGLSLWDTQESESGFTPILWAEYP  236
                   : *  :      ..  :.: *.:.: .:** : : *:   :    : ..   :

26695HP0502/0503   LNDFDYIANYVDKYPNAKKVGACVGYFFPMRDIN---ALKRNKTFLNAPSTNVVRISQDK  237
NSH57M.HpyAXII     LNDFDYIANYVDKYPNAKKIGACVGYFFPMRDIN---ALKCNKTFLNAPSANAVYISQDK  267
M.TaqI             HWEGEIIRFETEETRKLEISGMPLGDLFHIRFAARSPEFKKHPAVRKEPGPGLVPVLTGR  296
                   : : *    .:: : :   * :* :* :*     :* : :. : *... * : .:

26695HP0502/0503   LIYYQYIHYFK---------EIAPKIPYYFGNLDIIIDCFAFLEIKDAFLK---------  279
NSH57M.HpyAXII     LIYYQYIHYFK---------EIAPKIPYYFGNLDIIIDHFAFLEIKDAFLK---------  309
M.TaqI             NLKPGWVDYEKNHSGLWMPKERAKELRDFYATPHLVVAHTKGTRVVAAWDERAYPWREEF  356
                   :   ::.* *          * * ::  ::.. .:::     .:  *:  :

26695HP0502/0503   -------------------DKRARLEYFKKLFQGHPCEFD--------------------  300
NSH57M.HpyAXII     -------------------DKRARLEYFKKLFQGHPCEFD--------------------  330
M.TaqI             HLLPKEGVRLDPSSLVQWLNSEAMQKHVRTLYRDFVPHLTLRMLERLPVRREYGFHTSPE  416
                                      :..*  ::.:.*::..  .:

26695HP0502/0503   -----
NSH57M.HpyAXII     -----
M.TaqI             SARNF  421
```

B

Figure 1. Genomic organization of the HpyAXII R–M system. (**A**) Amino acid sequence alignment of 26695 HP0502/0503, NSH57 M.HpyAXII and M.TaqI from *Thermus aquaticus* reveals conserved motifs (highlighted in gray) in the N-terminal portion of the protein. (*) - identical residues, (:) conserved substitution, (.) semi-conserved substitution. (**B**) Genomic organization of genes *hpyAXIIM* and *hpyAXIIR* in *H. pylori* strains 26695, J99, HPAG1 and two mouse adapted strains NSH57 and NSH79. The expected phenotype for this R-M system in each strain is given based on DNA sequences. M, DNA MTase; R, REase; +, active; –, inactive. M.HpyAXII is truncated in 26695 and fused to R.HpyAXII in NSH79. R.HpyAXII is truncated in 26695, partially deleted in NSH57 (dashed lines), and fused to M.HpyAXII in NSH79 (in-frame deletion represented with dashed lines).

fragmentation when separated on agarose gel (data not shown), confirming that this single amino acid substitution abolishes MTase catalytic activity. In the second approach, we cloned and expressed the putative endonuclease gene *hpyAXIIR* in *E. coli* under the control of an inducible promoter (Figure 2D). We chose *hpyAXIIR* from *H. pylori* strain HPAG1 because this protein is expected to be functional based on sequence
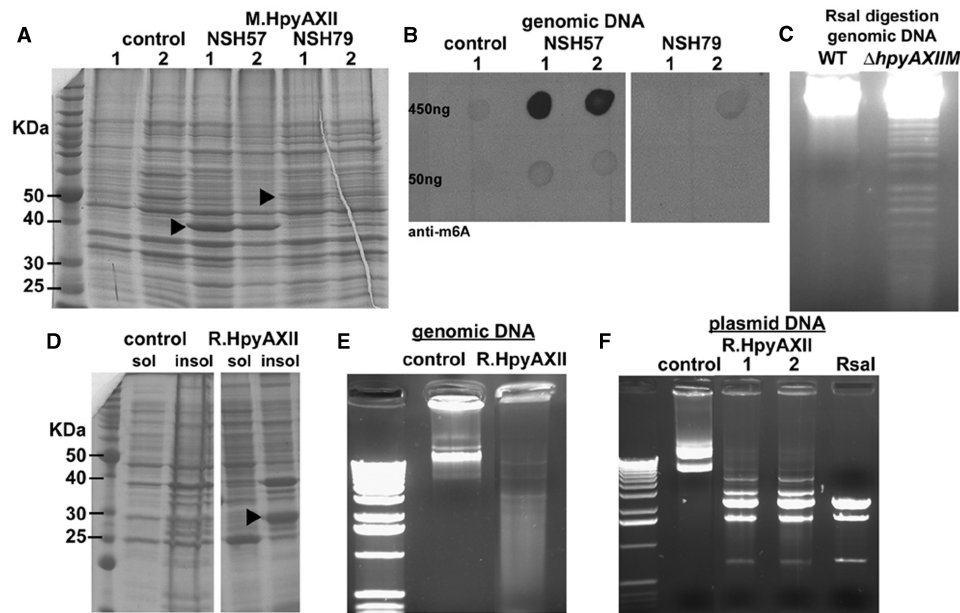
Figure 2. Biochemical demonstration that HypAXII functions as an R–M system targeting GTAC sites. (**A** and **B**) NSH57 M.HpyAXII methylates genomic DNA on adenines when expressed in *E. coli*. (A) *Escherichia coli* ER2925(DE3) extracts expressing NSH57 or NSH79 M.HpyAXII (arrowheads) separated by SDS–PAGE gel and stained with Coomassie blue. Control - empty plasmid. Two separate *E. coli* clones (1 and 2) were analyzed. (B) The indicated amount of genomic DNA from the respective *E. coli* cultures was cross-linked to a nitrocellulose membrane. DNA N-6 adenine methylation (m6A) was detected using anti-m6A antibodies and showed a signal during NSH57 M.HpyAXII but not NSH79 M.HpyAXII expression. (**C–F**) The HpyAXII R–M system targets the tetramer 5′-GTAC-3′. (C) *Helicobacter pylori* strain NSH57 wild-type and Δ*hpyAXIIM* genomic DNA was isolated and digested with RsaI. Agarose gel separation showed fragmentation for Δ*hpyAXIIM* genomic DNA but not wild-type. (D) *hpyAXIIR* from *H. pylori* strain HPAG1 was cloned and expressed *E. coli* BL21(DE3) under inducible conditions. Soluble (sol) and insoluble (insol) fractions were separated by SDS–PAGE gel and stained with Coomassie blue. R.HpyAXII was mostly insoluble (arrowhead). Control, empty plasmid. The soluble fractions shown in C were incubated with unmethylated genomic DNA (E) or plasmid DNA (F) and separated by agarose gel. DNA fragmentation was only observed when R.HpyAXII was expressed. Similar DNA restriction patterns (F) indicate that these enzymes cleave the same DNA sequence. Bands of higher sizes seen for with R.HpyAXII suggest incomplete digestion. Two separate *E. coli* clones are shown (1 and 2).

analysis (Figure 1). Incubation of the *E. coli* soluble fraction expressing HPAG1 R.HpyAXII with genomic DNA unmethylated for GTAC sites showed endonuclease activity as indicated by the resulting smear on agarose gel (Figure 2E). To determine R.HpyAXII cleavage site, the same experiment was carried out with unmethylated plasmid DNA. R.HpyAXII-containing extract cleaved plasmid DNA and gave an identical fragmentation pattern as RsaI (Figure 2F) confirming that R.HpyAXII acts on GTAC sites. In contrast to HPAG1 R.HpyAXII, no endonuclease activity was detected when R.HpyAXII from *H. pylori* strain J99 was expressed in *E. coli* (data not shown). While the exact cutting site of R.HpyAXII within the tetramer GTAC remains to be defined, our demonstration that *hpyAXIIM* and *hpyAXIIR* encode two independent enzymes suggests that they together constitute a class II R-M system (4).

GTAC sites are scarce in the *H. pylori* chromosome but are found at high frequency in the rDNA genes

We determined the frequency and distribution of GTAC sites in four sequenced strains of *Helicobacter pylori*: 26695 (122 sites), J99 (184 sites), HPAG1 (114 sites), G27 (115 sites). This total number of sites is >50-fold lower than the expected number based on frequency of nucleotides and confirms previous observations (35).

Interestingly, a significant fraction of the GTAC sites are located within the ribosomal RNA (rRNA) genes. The 23SrRNA contains eight or nine sites depending on the strain analyzed and the 16SrRNA contains two sites (Table 1). Since two copies of each gene exist in the *H. pylori* chromosome, a total of 20 or 22 GTAC sites are located within the rRNA genes. We were particularly interested in GTAC sites that are strictly conserved among *H. pylori* strains since they may be biologically important. Of all intragenic sites, ten are conserved in all four *H. pylori* strains and can be mapped to a total of 10 genes that belong to various functional categories (Table 1). Interestingly, three of these genes belong to the cag pathogenicity island (*cagT*, *cagL* and *cagH*). Some of these sites are located in the promoter region of genes and may influence transcription efficiency.

The HpyAXII R-M system effectively restricts incoming plasmid DNA

Bacterial R–M systems protect host genomes from invading DNA such as viral, plasmid and other foreign DNA by cleaving exogenous DNA at specific sites when unmethylated. We determined whether the endonuclease R.HpyAXII could restrict incoming plasmid DNA during natural transformation of *H. pylori*. As shown above, R.HpyAXII is biochemically active in *H. pylori* HPAG1,

Table 1. Distribution of all conserved GTAC sites in four *H. pylori* strains

Gene number[a]	Gene name[a]	Gene description	Characteristics[b]
	23S rDNA	Ribosomal DNA	8 sites
	16S rDNA	Ribosomal DNA	2 sites
HP0200	*rpl32*	Predicted ribosomal protein L32	8 bp after start of HP0200 and 200 bp from start of HP0201
HP0486	*hofC*	Outer membrane protein	
HP0499	*pldA*	Phospholipase A1 precursor	
HP0532	*cagT*	cag pathogenicity island protein T	
HP0539	*cagL*	cag pathogenicity island protein L	200 bp from start of cagI
HP0541	*cagH*	cag pathogenicity island protein H	
HP0613		Predicted ABC transporter	87 bp after start of HP0613
HP0630	*mda66*	Modulator of drug activity	11 bp after start of HP0630 and 170 bp from start of HP0631
HP1197	*rps12*	Predicted ribosomal protein S12	190 bp from start of HP1198
HP1228	*invA*	Diadenosine polyphosphate hydrolase	50 bp from start of HP1229

[a]Based on 26695 annotation (61).
[b]The number of conserved sites is given when higher than one. The exact position of the site is given only when within 200bp from the start codon of a gene.

a strain difficult to transform (data not shown). We thus replaced the endogenous locus encoding M.HpyAXII/R.HpyAXII in the easily transformable NSH57 strain with that from HPAG1 (Materials and methods section). This engineered strain, NSH57*hpyAXIIM/R*$_{HPAG1}$, encodes a fully active HpyAXII R-M system (R + M + phenotype). For comparison, we deleted the endonuclease component R.HpyAXII in the same strain and replaced it with an antibiotic resistance cassette to create a null allele ($\Delta hpyAXIIR::aphA3$, R–M + phenotype). As donor DNA, we used plasmid pTM113-*cat* (Materials and methods and Supplementary TableS3) that we isolated from *H. pylori* strain NSH57 wild-type (fully methylated, denoted 'M +'), from the catalytically inactive mutant NSH57*hpyAXIIM*^{N87A} (unmethylated for GTAC sites, denoted 'M−'), or from *E. coli* (unmethylated at all sites, denoted 'all M−'). We first verified that the numerous R–M systems encoded by *H. pylori* strain NSH57 could effectively restrict unmethylated (all M−) plasmid DNA as compared to methylated (M +) plasmid. Transformation frequency of wild-type *H. pylori* NSH57 decreased over 200-fold when completely unmethylated plasmid DNA prepared from *E. coli* was used as compared to methylated (M +) plasmid (data not shown), in agreement with previous work (37). We similarly found that transformation frequency involving M− plasmid decreased over 60-fold in the bacteria harboring an active R.HpyAXII endonuclease (R +) (ratio R−/R +, Table 2), demonstrating that this enzyme constitutes an effective barrier against incoming plasmid DNA *in vivo*.

The HpyAXII R–M system effectively restricts incoming chromosomal DNA

To investigate the role of the endonuclease R.HpyAXII in restricting incoming chromosomal DNA *in vivo*, we assessed transformation using donor genomic DNA harboring a kanamycin resistance marker (*aphA3*) at the 23S rDNA locus. This marker contains a GTAC site and is susceptible to cleavage by R.HpyAXII. As described for plasmid DNA, donor chromosomal DNA was isolated

Table 2. R.HpyAXII effectively restricts incoming plasmid DNA

Donor DNA[a]	Transformation frequency (# transformants/recipient cell/ng $\times 10^{-8}$)[b], mean (SEM)		Ratio
	Recipient strain[c]		
	R +	R−	R−/R +
M−	12 (8.8)	760 (120)	63
M +	1900 (400)	2500 (180)	1.3

[a]Donor plasmid pTM113-*cat* contains a total of five GTAC sites and was constructed by substituting *cat* for *aphA3* in pTM113 (Supplementary Table S3). pTM113-*cat* was isolated from *H. pylori* strain NSH57 wild-type (M +) or NSH57*hpyAXIIM*^{N87A} (M−). A total of 3.3 ng of plasmid DNA was used for each transformation.
[b]Transformation frequency was calculated by dividing the number of transformants recovered by the total number of recipient bacteria and was based on the average of three replicates.
[c]Strain NSH57*hpyAXII*$_{HPAG1}$ (R +) encodes an active R.HpyAXII and strain NSH57*hpyAXIIM/R*$_{HPAG1}$ $\Delta hpyAXIIR::aphA3$ (R−) lacks this enzyme.

from bacteria encoding an active (M +) or inactive (M−) M.HpyAXII MTase. We determined if the endonuclease R.HpyAXII could restrict incoming chromosomal DNA by comparing transformation efficiency in NSH57 harboring either an active (R +) or inactive (R−) restriction endonuclease R.HpyAXII. When genomic DNA isolated from *H. pylori* lacking M.HpyAXII MTase activity (M−) was transformed into the bacteria harboring an active R.HpyAXII endonuclease (R +), a 240-fold decrease in transformation frequency was observed (R−/R +, Table 3). We also digested the same donor genomic DNA *in vitro* with RsaI prior to transformation, which lowered the transformation frequency further to undetectable levels (Table 3). We suggest that cleavage of the GTAC site in gene *aphA3* by the endonucleases RsaI and R.HpyAXII inactivates this marker and results in the subsequent decrease in transformation frequency. Overall, we demonstrated that active R.HpyAXII endonuclease effectively limits the incorporation of

Table 3. R.HpyAXII effectively restricts incoming genomic DNA

Donor DNA[a]	Transformation frequency (# transformants/recipient cell/ng $\times 10^{-10}$)[b], mean (SEM)		Ratio
	Recipient strain[c]		
	R+	R−	R−/R+
M+	18 (0.16)	44 (17)	2.4
M+ RsaI	49 (9.3)	95 (30)	1.9
M−	0.42 (0.26)	100 (11)	240
M− RsaI	0	1.4 (0.60)	inf [d]

[a]Donor genomic DNA was constructed by inserting the *aphA3* cassette at position 2586 bp (from start codon) in one of the two copies of 23S rDNA genes. One GTAC site is present in *aphA3* at position #94 bp and three flanking GTAC sites are found in close proximity to the marker at positions #2527, #2660 and #2745. Chromosomal DNA was isolated from *H. pylori* strain NSH57 wild-type (M+) or NSH57Δ*hpyAXIIM::cat* (M−), and predigested or not with RsaI. A total of 10 ng genomic DNA per tranformation was used for M− and 7.9 ng for M+.
[b]Transformation frequency was based on the average of three replicates.
[c]Strain NSH57*hpyAXIIM*/R_{HPAG1} (R+) encodes an active R.HpyAXII and strain NSH57*hpyAXIIM*/R_{NSH57} (R−) encodes an inactive enzyme.
[d]Ratio is infinite because no transformants were recovered for R+ recipient strain

Table 4. The HpyAXII R-M system follows the selfish hypothesis

Donor DNA	Transformation frequency (# transformants/recipient cell/ng DNA $\times 10^{-10}$)[c], mean (SEM)		Ratio
	Recipient strain[d]		
	R+	R−	R−/R+
Δ*hpyAXIIM*/*R::catsacB*[a]	0.15 (0.08)[e]	320 (21)	2100
Δ*hpyAXIIM::cat*[b]	0.17 (0.17)[f]	1100 (350)	6500
ΔHP0368::*cat*[b]	310 (9.2)	180 (38)	0.58

[a]NSH57 wild-type genomic DNA was constructed by replacing the entire HpyAXII R-M system with the *catsacB* marker. 500 ng genomic DNA per transformation was used.
[b]A PCR product was engineered to disrupt genes *hpyAXIIM* or HP0368. The 5′ and 3′ homologous ends of the respective gene was fused to the antibiotic resistance cassette and transformed to generate a null allele by allelic replacement. A 150 ng DNA per transformation was used for Δ*hpyAXIIM::cat* and ΔHP0368::*cat*.
[c]Transformation efficiency was based on the average of three replicates.
[d]As in Table 2.
[e]A total of two clones was recovered.
[f]A total of five clones was recovered.

unmethylated exogenous homologous DNA into the *H. pylori* chromosome.

The HpyAXII R–M system follows the selfish hypothesis

Kobayashi and colleagues suggested that type II R–M systems act as selfish genetic elements and cannot be easily replaced with a homologous stretch of DNA (36). This phenomenon, more commonly known as postsegregational killing, has been attributed to the gradual dilution of the modification and restriction enzymes in the descendants of the cell that lost the R–M system. Since there is asymmetry between the roles of the MTase and REase, newly replicated chromosomes will be inadequately protected by methylation and rendered susceptible to cleavage by the residual REase, leading to cell death. To determine whether the type II HpyAXII R–M system follows the selfish hypothesis, we deleted this chromosomal locus by allelic replacement from *H. pylori* strain NSH57 engineered to harbor an active (R+) or inactive (R−) R.HpyAXII endonuclease. We found that R− but not R+ recipient bacteria were easily transformed with donor NSH57 genomic DNA in which *hpyAXIIM/R* was replaced with a *catsacB* marker (strain NSH57 Δ*hpyAXIIM/R::catsacB*) (Table 4, Δ*hpyAXIIM/R::catsacB*). This 2100-fold decrease in transformation frequency seen for the R + recipient strain is independent of the digestion of GTAC sites in the donor DNA because they are fully protected by DNA methylation. We verified that the recipient strains were not generally defective at transforming DNA by showing allelic replacement at a different chromosomal region (ΔHP0368). In summary, we showed that chromosomal type II HpyAXII R–M system resists its replacement with an empty site in the

presence of the active endonuclease R.HpyAXII, as described for other bacterial type II R–M systems (38,39).

An active R.HpyAXII REase can only exist in the presence of its active DNA MTase partner

We predicted that bacteria harboring a functional REase must also encode an active cognate DNA MTase to protect the genome from digestion. The R + M− phenotype should therefore be lethal, a condition especially relevant for the HpyAXII R-M system since many of the cognate sites are present within the rRNA genes (Table 1). Thus we deleted *hpyAXIIM* by allelic replacement using the Δ*hpyAXIIM::cat* PCR construct in bacteria harboring active (R+) or inactive (R−) endonuclease R.HpyAXII (Supplementary Figure S3). We observed a dramatic reduction in transformation efficiency (6500-fold) for the R + bacteria (Table 4). In comparison, allelic replacement at a different chromosomal region (ΔHP0368) was carried out at a similar efficiency in R + and R- bacteria (Table 4). Importantly, the observed decrease in transformation efficiency is independent of the cleavage of incoming Δ*hpyAXIIM::cat* donor DNA by the endonuclease R.HpyAXII since this fragment is devoid of GTAC sites. Rather, recombinants for the R + bacteria probably died from double-stranded breaks in their chromosomal DNA.

We were intrigued by the rare transformants obtained when gene *hpyAXIIM* was deleted in the R + bacteria because it suggests that the R + M− phenotype is occasionally viable. We investigated these clones further and found evidence of gene amplification and adaptive mutations at the M.HpyAXII/R.HpyAXII locus. From three independent transformations, we analyzed a total of 16 clones; seven displayed evidence of gene amplification and nine showed adaptive mutations. For gene amplification, *hpyAXIIM* was present in both a wild-type and
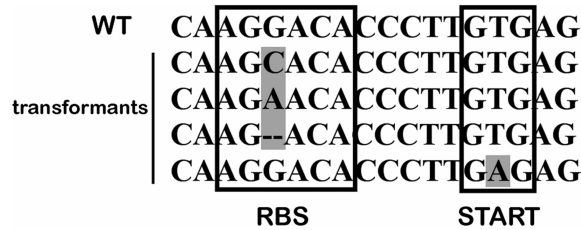
Figure 3. Deletion of *hpyAXIIM* from a *H. pylori* strain harboring an active endonuclease R.HpyAXII yielded rare transformants that display adaptive mutations in the promoter region of *hpyAXIIR*. Gene *hpyAXIIR* was sequenced in nine surviving recombinants obtained following *hpyAXIIM* deletion. When compared to wild-type, single base deletions and substitutions (highlighted in gray) were detected in the transformants in both the ribosomal binding site (RBS) and start codon (START) of *hpyAXIIR*.
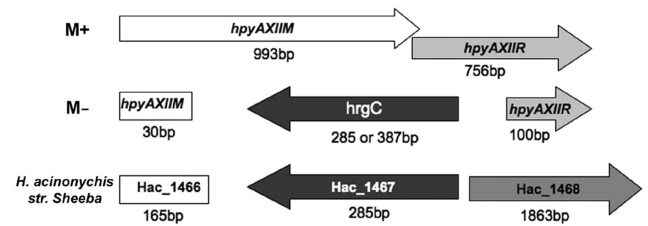

Figure 4. M.HpyAXII MTase activity is occasionally disrupted by insertion of the alternate gene *hrgC*. The HpyAXII R–M system was sequenced in the *H. pylori* clinical isolates where M.HpyAXII was found inactivated (M−). In all M− strains, most of the HpyAXII locus was replaced with alternate gene *hrgC* that shows strong homology to Hac_1467 (80% nucleotide identity) from *Helicobacter acinonychis sheeba*. Depending on the strain analyzed, gene *hrgC* is either 285- or 387-bp long.

a mutated form at the same locus (Supplementary Figure S3), overall resulting in the phenotype R+ M+. This diploidy for *hpyAXIIM* was unstable and the mutant form of the gene was generally lost after several passages *in vitro* (data not shown). In other clones, point mutations were found in the promoter region of gene *hpyAXIIR*, specifically in the ribosome binding sequence and in the start codon of the gene (Figure 3). In all cases, these mutations likely prevent effective translation of the endonuclease R.HpyAXII; the overall phenotype of these clones therefore became R−M−. We conclude that *H. pylori* bacteria only harbor an active R.HpyAXII endonuclease in the presence of an active M.HpyAXII MTase.

The adenine MTase M.HpyAXII is highly conserved among *H. pylori* strains

Helicobacter pylori R–M systems are remarkably heterogeneous among strains. Some are present in a fraction of strains, absent in others, or inactivated by deletion or frameshift mutations. It has been estimated that less than one-third of *H. pylori* R–M systems are fully active (6). We were interested in determining whether M.HpyAXII is present and active in a panel of *H. pylori* clinical isolates. We found that the HpyAXII R–M system is well conserved as we amplified this locus from all 50 pediatric clinical isolates isolated from 47 patients (S. Talarico, J. Fero, D.T. Thompson, J. Guarner, S. Czinn. B.D. Gold, N.R. Salama, personal communication) though its size was reduced in four (8%) strains (data not shown). To assess M.HpyAXII MTase activity for the clinical isolates, we digested each strain's genomic DNA with the REase RsaI, which only cleaves unmethylated GTAC sites. Of the 50 clinical isolates DNA, only four strains were digested with RsaI, all of which displayed a smaller amplified product for the HpyAXII R–M system (data not shown). M.HpyAXII activity is therefore remarkably conserved among *H. pylori* strains (92% active as tested). To determine how M.HpyAXII was inactivated in the four *H. pylori* strains lacking DNA MTase activity, we sequenced the HpyAXII R–M system. In all strains, most of *hpyAXIIM* and *hpyAXIIR* coding region was substituted by a gene in reverse orientation (Figure 4). This replacing gene, that we name *hrgC* (<u>H</u>pyAXII

replacing gene <u>C</u>), shows strong homology to *Helicobacter acinonychis sheeba* Hac_1467 (80% nucleotide identity). Hac_1467 is present at a different chromosomal region in this species (557 kb away from the HpyAXII locus) and encodes a 94 amino acid protein of unknown function. In the four clinical isolates where gene replacement occurred, *hrgC* encodes a protein of 94- or 128-amino acids. A search for orthologs of Hac_1467 at the *H. acinonychis* locus using a PCR approach was negative for all 50 clinical isolates and only positive for *H. pylori* strain NSH57. Surprisingly, in both *H. acinonychis sheeba* and *H. pylori* NSH57, Hac_1467 is preceded by a 165 bp DNA fragment (Hac_1466) that shows high homology to the 5′-end of *hpyAXIIM* (89% sequence identity, Figure 4), suggesting that some recombination event took place at this locus. In summary, we demonstrated that the activity of the DNA MTase M.HpyAXII is highly conserved among *H. pylori* strains but can be disrupted by insertion of the alternate gene *hrgC* at the same locus in several unrelated strains.

The REase R.HpyAXII is poorly conserved among *H. pylori* strains

The conservation of M.HpyAXII MTase activity in *H. pylori* isolates led us to evaluate R.HpyAXII activity in these same isolates. We developed a genetic assay to assess R.HpyAXII endonuclease activity based on the previous finding that the MTase M.HpyAXII cannot be deleted in the presence of an active REase. Since *H. pylori* strains vary in their ability to be transformed by exogenous DNA, we chose to engineer strain NSH57 to contain the HpyAXII R-M system from the different clinical isolates using allelic replacement and a knockin approach (Materials and methods section). We transformed the resulting clones with the Δ*hpyAXIIM*::*cat* donor DNA to delete gene *hpyAXIIM*. When a large number of transformants were recovered, we deduced that R.HpyAXII from that particular strain lost endonuclease activity. In contrast, when few or no transformants were obtained, we concluded that R.HpyAXII must be active because of the lethality of the R + M− phenotype. The DNA fragment ΔHP0368 was used as transformation control. We first evaluated R.HpyAXII activity in

the NSH57 clones engineered to contain the HpyAXII R–M system from *H. pylori* strains NSH57, J99, NSH79 and HPAG1. As expected from our genomic and biochemical results (Figures 1B and 2), only the clone harboring the HpyAXII R–M system from strain HPAG1 showed a dramatic reduction in transformation efficiency when gene *hpyAXIIM* was deleted (Supplementary Figure S4). In contrast, clones harboring HpyAXII from strain NSH57, NSH79 and J99 yielded a large number of transformants, confirming that the endonuclease R.HpyAXII is inactive in these strains. This result therefore validates the use of our genetic assay to investigate R.HpyAXII endonuclease activity. We constructed NSH57 clones harboring the HpyAXII R-M system from 37 clinical isolates and found that R.HpyAXII was only active in seven clones (18.9%) based on the number of transformants recovered. All strains could successfully be transformed with the ΔHP0368 PCR product, ruling out a general defect in transformation. We examined the nature of mutations that inactivated R.HpyAXII in the 30 *H. pylori* isolates by sequencing gene *hpyAXIIR*: nine strains (30%) contain a large deletion (39AA) similar to the one found in strain NSH57 (Figure 1B), seven strains (23%) harbor nonsense mutations that result in a truncated protein, two strains (7%) have both the deletion and nonsense mutations, and 12 strains (40%) contain single nucleotide polymorphisms (SNPs) in the coding region of gene *hpyAXIIR*. For four strains, we confirmed that the *hpyAXIIR* sequence was identical in the engineered strain and in the original isolate, ruling out the possibility that mutations were introduced during the genetic manipulations.

To analyze the inactivating SNPs in more details, we aligned the protein sequences from the 12 inactive and seven active R.HpyAXII enzymes with that of *H. pylori* strain HPAG1, shown to be active. We identified between two and five amino acid substitutions per strain that may inactivate R.HpyAXII function, resulting in a total of 31 distinct coding SNPs. The nature and frequency of these substitutions are described in Figure 5A and the SNPs for each strain are described in Supplementary Table S3. These coding SNPs were mapped on the primary protein sequence of R.HpyAXII (Figure 5B) as well as on the three-dimensional model based on R.PabI crystal structure (Figure 5C) (40). Out of these 31 SNPs, two (#5 and #18) overlap with residues previously shown to be required for R.PabI activity (40), and seven are found in multiple *H. pylori* strains (Figure 5A). Several structural motifs (β3, α3, β6, β8, β10 and α4 based on R.PabI) contain a high density of SNPs and may be crucial for R.HpyAXII endonuclease activity.

H. pylori M.HpyAXII MTase activity is not required for establishing infection in mice

The remarkable preservation of M.HpyAXII activity in *H. pylori* strains and the intriguing distribution of chromosomal GTAC sites led us to investigate whether this DNA MTase confers a selective advantage to *H. pylori* during infection. We were also interested in determining whether M.HpyAXII has a role during infection that is
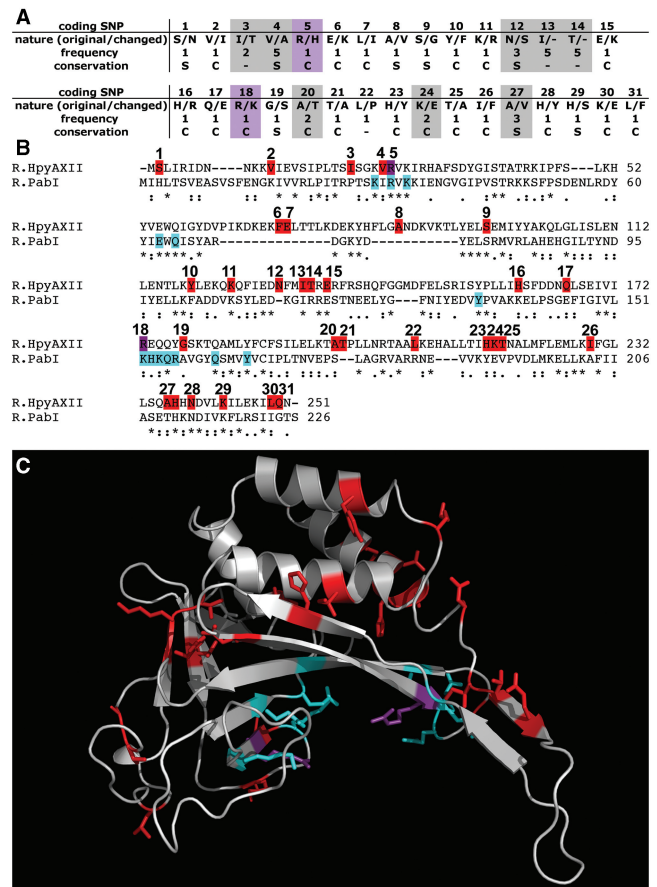

Figure 5. Protein sequence analysis of 12 inactive R.HpyAXII enzymes reveal 31 distinct coding SNPs. (A) Comparison of R.HpyAXII protein sequences in 12 *H. pylori* strains harboring inactive enzyme and in seven strains harboring active enzyme with that from strain HPAG1 identified 31 coding SNPs, enumerated #1 through #31 with the conservation and frequency of each substitution indicated. (C) - conserved, (S) - semiconserved, (-) - unrelated. Changes highlighted in gray were found in multiple strains and changes highlighted in purple were previously shown to be required for R.PabI activity (40). (B) HPAG1 R.HpyAXII and R.PabI protein sequences were aligned and the position of each SNP highlighted. Red, potentially inactivating SNPs numbered #1 through #31; cyan, previously shown to be required for R.PabI activity and target the predicted DNA-binding domain; purple, coding SNPs in R.HpyAXII also required for R.PabI activity. (*) - identical residues, (:) - conserved substitution, and (.) semi-conserved substitution. (C) Coding SNPs were mapped on the three-dimensional model generated for R.HpyAXII based on the R.PabI crystal structure using the same color code used in (B).

independent of its DNA MTase activity. For this reason, we infected mice with *H. pylori* NSH57 deletion mutant (Δ*hpyAXIIM*::*cat*) and with the catalytically inactive mutant (*hpyAXIIM*[N87A]), which MTase can presumably still bind DNA but not methylate it. Mouse stomachs were harvested after one week of infection and viable bacteria were recovered and enumerated to calculate the competitive index for each infection [colony forming units (CFU) mutant bacteria/CFU wild-type bacteria] (Materials and methods section). When mice were infected with NSH57Δ*hpyAXIIM*::*cat* mutant and wild-type bacteria, equal number of bacteria were recovered (*n* = 9, Figure 6 left) indicating that they could establish infection
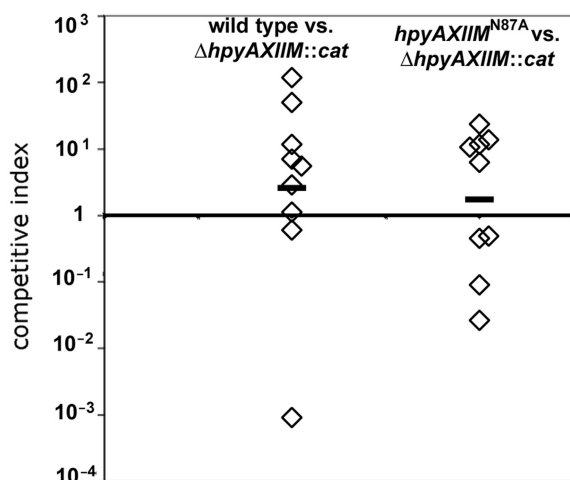
Figure 6. M.HpyAXII MTase activity is not required by *H. pylori* for establishing infection in a mouse infection model. Mice were infected in 1:1 competition with the indicated *H. pylori* strains and the competitive index (CI) determined after one week of infection. The CI was calculated by dividing colony forming units (CFU) recovered of null mutant bacteria over CFU wild-type or point mutant bacteria and corrected by dividing the actual input ratio enumerated from plating the inoculums. Left, mice (*n* = 9) were infected with *H. pylori* NSH57 wild-type and an isogenic null mutant *ΔhpyAXIIM*::*cat*. Both strains infected mice at a similar efficiency (geometric mean CI slightly above one). Right, mice (*n* = 9) were infected with the NSH57*ΔhpyAXIIM*::*cat* mutant and the catalytically inactive mutant (*hpyAXIIM*[N87A]) constructed in strain NSH57.

at a similar efficiency. M.HpyAXII protein expression is therefore not required for establishing infection in *H. pylori* strain NSH57. We also infected mice in competition with the catalytically inactive mutant *hpyAXII*[N87A] and the deletion mutant *ΔhpyAXIIM*::*cat*. Both strains infected mice similarly (*n* = 9, Figure 6 right) indicating that M.HpyAXII does not have a role during establishment of infection that is independent of its DNA MTase activity.

DISCUSSION

Helicobacter pylori naturally takes up exogenous DNA, a condition thought to promote genetic diversification by spreading new alleles in the population. However, genetic exchange is limited by the large number and heterogeneity of *H. pylori* R–M systems (6). The contribution of individual R–M systems in restricting incoming DNA remains for the most part undetermined since only a few of these systems have individually been tested in *H. pylori* (37,41). In addition, it is possible that R–M systems carry out other biological roles. In our study, we have biochemically characterized a novel *H. pylori* type II R–M system, M.HpyAXII/R.HpyAXII, and showed that it targets GTAC sites thus confirming prior predictions (35). The endonuclease R.HpyAXII effectively restricted both unmethylated plasmid (60-fold) and genomic (240-fold) DNA (Tables 2 and 3). Our results agree with previous reports documenting effective restriction of plasmid DNA by R.HpyAIII (encoded by HP0091, GATC) (37), and of

chromosomal DNA by R.HpyAII (encoded by HP1366, GAAGA) (41).

Since R–M systems are highly polymorphic in *H. pylori*, we analyzed the conservation of both M.HpyAXII MTase and R.HpyAXII endonuclease activities in a panel of clinical isolates. We found a striking asymmetry between the highly conserved MTase activity (92% active, *n* = 50) and that of the REase (18.9% active, *n* = 37). In fact, our data indicate that M.HpyAXII is the third most highly conserved adenine MTase in *H. pylori*, preceded only by M.HpyAI (CATG, 100% conserved) and M.HpyAIII (GATC, 97% conserved) (42,43). Other studies have suggested that MTase activities are generally selected for in *H. pylori* (6,29,44) but how DNA methylation benefits the organism is unknown. To begin to probe how M.HpyAXII affects *H. pylori* biology, we generated a null allele in gene *hpyAXIIM* in mouse-adapted strain NSH57 either by deletion or by targeted mutation of the catalytic site of the enzyme (*hpyAXIIM*[N87A]). Mouse infection with these strains indicated that M.HpyAXII expression is not required by *H. pylori* for the establishment of infection (Figure 6) in contrast to previous work (19). As this study did not complement the mutant phenotype, it is possible that second site mutations or polar effects accounted for the defect observed. Further experiments are needed to determine if M.HpyAXII activity is important during persistent infection and to elucidate how GTAC methylation provides a selective advantage in *H. pylori*. It was proposed that the maintenance of an active MTase allows reactivation of a given R–M system through recombination with the genomic DNA of a strain that encodes an active REase. However, the probability of this event must be relatively low given that R.HpyAXII is active in a small number of strains. It was also postulated that 'orphan' MTases are conserved because they act as a 'molecular vaccine' against the parasitism of R–M systems. This hypothesis was based on the finding that Dcm in *E. coli* can alleviate postsegregational killing induced by the EcoRII R–M complex that recognizes the same sequence as Dcm (45). This also seems unlikely because no R–M system that acts on the same site as HpyAXII has been identified in *H. pylori*. DNA methylation may alternatively be selected for in *H. pylori* because it confers immunity from the action of the cognate REase and therefore increases the chances that an entire gene is successfully propagated in a population. However, the finding that the average size of recombined fragment is unusually small (417 bp) in *H. pylori* as compared to other bacteria indicates that incoming DNA is generally susceptible to cleavage by REases (46).

Another important role of DNA methylation is to modulate the interaction between proteins and DNA as demonstrated for the stand-alone adenine MTase Dam in α- and γ-Proteobacteria (14,15,47). DNA MTases involved in R–M systems that are not coupled to an active REase may have evolved similar regulatory functions since they are no longer under the pressure to act on all chromosomal sites and may therefore preferentially methylate sequences required for transcription. The genome-wide analysis of conserved GTAC sites identified genes of various functional classes including several genes

of the cag pathogenicity island but it is unknown whether their expression is affected by M.HpyAXII methylation. Although GTAC sites are scarce in the *H. pylori* chromosome, they are present at high frequency in the 23S and 16S rRNA genes. Our preliminary experiments showed no effect of M.HpyAXII methylation on 23S rRNA transcripts levels or on *in vitro* growth of the bacteria (data not shown), arguing against a role for GTAC methylation in *H. pylori* ribosomal biogenesis. It is possible that ribosomal GTAC sites are preserved because they are crucial for the folding of the RNA into functional secondary structures in Helicobacter species and that methylation of these sites does not have any biological role. In agreement with this hypothesis, the analysis of 23S rRNA sequences in 55 Helicobacter strains representing 41 taxa (48) shows conservation of most GTAC sites even though many of these species probably lack a M.HpyAXII enzyme. Alternatively, ribosomal GTAC sites may not have been erased from these chromosomal regions because they evolve at a different rate than the rest of the genome; these elements are duplicated in *H. pylori* and are subjected to homogenization processes such as gene conversion (49). Overall, the significance of ribosomal GTAC sites in *H. pylori* remains unsolved but echoes a similar anomaly in site distribution with the tetramer CTAG that is very rare in *E. coli* and in other proteobacterial genomes and over-represented in rRNA genes (50,51).

The comparison of tetranucleotide combinations in 27 representative microbial genomes revealed that ACGT, GTAC and TCGA are strongly underrepresented in both coding and noncoding DNA in *H. pylori* (52). A significant correlation between the avoidance of palindromic words and type II R–M systems was suggested (53) presumably because these systems exert significant pressure on an organism when restriction is intact but methylation is incomplete. However, the highly conserved *H. pylori* R–M systems M.HpyAI/*iceA* (CATG) or M.HpyAIII/R.HpyAIII (GATC) target sites are not underrepresented in the *H. pylori* genome. Since cognate site avoidance is probably dependent on the timing of acquisition of a given R–M system during evolution, the HpyAXII R–M system must have been acquired a long time ago, consistent with the finding that this system is present in all *H. pylori* isolates analyzed. Interestingly, GTAC sites are also scarce in the *Helicobacter hepaticus* genome even though it does not encode an M.HpyAXII/R.HpyAXII ortholog. This can either be explained by the recent loss of this system or by the existence of a different R–M system that acts on GTAC sites in *H. hepaticus*.

Our study of the conservation of M.HpyAXII MTase activity uncovered the occasional replacement of the HpyAXII R-M system with an alternate gene that we re-annotate as *hrgC* (<u>H</u>pyAXII <u>r</u>eplacing <u>g</u>ene <u>C</u>). Other *H. pylori* R–M systems encoding the two most highly conserved MTases M.HpyAI and M.HpyAIII are similarly replaced with genes of unknown function. In the HpyAIII R–M system, gene *hrgA* inactivated the REase R.HpyAIII in 33% of the 208 strains analyzed but rarely inactivated M.HpyAIII function (54). *hrgA* encodes a 370 amino acid protein of unknown function and this gene was associated with gastric cancer in *H. pylori* isolates

from Asian patients but not among Western strains (55). In the HpyAI R–M system, the REase encoded by *iceA1* was replaced with *iceA2* in 44% of the strains analyzed (56). *iceA2* (renamed *hrgB*) encodes a polymorphic protein of unknown function ranging from 24 to 59 amino acid long depending on the strain studied, and epidemiological studies have correlated the presence of gene *iceA1* in *H. pylori* strains with the incidence of peptidic ulcers disease (57). In our study, we showed that similarly to *hrgA* and *hrgB*, *hrgC* encodes a protein of unknown function but its origin could be traced to a different locus in *H. acinonychis sheeba* and *H. pylori* strain NSH57 (G27). Unlike *hrgA* and *hrgB*, *hrgC* replacement always compromised both the MTase and REase activities in our analysis of 37 clinical isolates. Further work will define whether *hrgC* has any role in *H. pylori* pathogenicity.

As predicted for a type II R–M system, we showed that deletion of gene *hpyAXIIM* is lethal in the presence of an active R.HpyAXII endonuclease. The lethal phenotype R + M− can be rescued by adaptive mutations, in the form of gene amplification of *hpyAXIIM* and null mutations in the promoter region of *hpyAXIIR*. This result is reminiscent of extensive research describing stress-induced mutations in *E. coli* with the Lac assay system (58). Based on these findings, we suggest that these adaptive mutations preexisted in the *H. pylori* population and were selected for as a result of *hpyAXIIM* deletion. Tandem amplification was previously documented for R–M systems, as when the entire BamHI R–M system was deleted from *Bacillus subtilis* (38). In another study, large-scale genomic rearrangements were demonstrated when the PaeR71 R–M system was deleted from *E. coli* (39). These rearrangements involved recombination between copies of the transposon IS3 and were proposed to have resulted from multiple rounds of unequal crossing-over or rolling circle amplification. Our study of the deletion of the MTase gene in the HpyAXII R-M system is the first to demonstrate the occurrence of compensatory mutations in the promoter of the respective REase gene.

Orthologs of M.HpyAXII generally show poor sequence homology except for one gene of unknown function from *Campylobacter upsaliensis* RM3195 (CUP1644) that shows 43% identity with M.HpyAXII. Other adenine MTase that target GTAC sites such as M. PabI from *Pyroccocus abyssi* and M.CviQI from Chlorella virus only show 14.6% and 17.8% protein identity, respectively. M.RsaI (7.9% protein identity) from *Rhodopseudomonas sphaeroides* and M.MjaV from *Methanococcus janaschii* (12.7% protein identity) act on the same sequence but methylates cytosines (m4C). GTAC sites are strongly avoided from the genomes of *R. sphaeroides* and *M. jannashi* but not from *P. abyssi*, suggesting that the PabI R–M system invaded this organism recently as discussed previously (35,59). The crystal structure of the endonuclease R.PabI was recently solved and unveiled a novel protein fold termed 'half-pipe' (40). Although the two isochizomers R.PabI and R.HpyAXII show relatively weak sequence identity (25.4%), several conserved motifs are shared between the two proteins (Figure 5B). The R.HpyAXII polypeptide could be threaded with high confidence onto the tertiary structure of R.PabI

(PHYRE, *E*-value 3e-26, Figure 5C) suggesting that R.HpyAXII folds into a similar half-pipe structure. Among the thousands of confirmed and hypothetical REase enzymes identified, Orlowski and Bujnicki predicted that only eight present the half-pipe fold (60). R.HpyAXII therefore constitutes an excellent candidate for refined biochemical analyses that will enrich our understanding of this new class of enzymes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Kusters,J.G., van Vliet,A.H. and Kuipers,E.J. (2006) Pathogenesis of *Helicobacter pylori* infection. *Clin. Microbiol. Rev.*, **19**, 449–490.
2. Alm,R.A., Ling,L.S., Moir,D.T., King,B.L., Brown,E.D., Doig,P.C., Smith,D.R., Noonan,B., Guild,B.C., deJonge,B.L. *et al.* (1999) Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature*, **397**, 176–180.
3. Oh,J.D., Kling-Backhed,H., Giannakis,M., Xu,J., Fulton,R.S., Fulton,L.A., Cordum,H.S., Wang,C., Elliott,G., Edwards,J. *et al.* (2006) The complete genome sequence of a chronic atrophic gastritis *Helicobacter pylori* strain: evolution during disease progression. *Proc. Natl Acad. Sci. USA*, **103**, 9999–10004.
4. Roberts,R.J., Belfort,M., Bestor,T., Bhagwat,A.S., Bickle,T.A., Bitinaite,J., Blumenthal,R.M., Degtyarev,S., Dryden,D.T., Dybvig,K. *et al.* (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.*, **31**, 1805–1812.
5. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2007) REBASE—enzymes and genes for DNA restriction and modification. *Nucleic Acids Res.*, **35**, D269–D270.
6. Lin,L.F., Posfai,J., Roberts,R.J. and Kong,H. (2001) Comparative genomics of the restriction-modification systems in *Helicobacter pylori*. *Proc. Natl Acad. Sci. USA*, **98**, 2740–2745.
7. Nobusato,A., Uchiyama,I. and Kobayashi,I. (2000) Diversity of restriction-modification gene homologues in *Helicobacter pylori*. *Gene*, **259**, 89–98.
8. Tock,M.R. and Dryden,D.T. (2005) The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.*, **8**, 466–472.
9. Israel,D.A., Lou,A.S. and Blaser,M.J. (2000) Characteristics of *Helicobacter pylori* natural transformation. *FEMS Microbiol. Lett.*, **186**, 275–280.
10. Jeltsch,A. (2003) Maintenance of species identity and controlling speciation of bacteria: a new function for restriction/modification systems? *Gene*, **317**, 13–16.
11. Price,C. and Bickle,T.A. (1986) A possible role for DNA restriction in bacterial evolution. *Microbiol. Sci.*, **3**, 296–299.
12. Naito,T., Kusano,K. and Kobayashi,I. (1995) Selfish behavior of restriction-modification systems. *Science*, **267**, 897–899.
13. Mochizuki,A., Yahara,K., Kobayashi,I. and Iwasa,Y. (2006) Genetic addiction: selfish gene. *Genetics*, **172**, 1309–1323.
14. Low,D.A. and Casadesus,J. (2008) Clocks and switches: bacterial gene regulation by DNA adenine methylation. *Curr. Opin. Microbiol.*, **11**, 106–112.
15. Wion,D. and Casadesus,J. (2006) N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat. Rev. Microbiol.*, **4**, 183–192.
16. Gressmann,H., Linz,B., Ghai,R., Pleissner,K.P., Schlapbach,R., Yamaoka,Y., Kraft,C., Suerbaum,S., Meyer,T.F. and Achtman,M. (2005) Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. *PLoS Genet.*, **1**, e43.
17. van der Woude,M.W. (2006) Re-examining the role and random nature of phase variation. *FEMS Microbiol. Lett.*, **254**, 190–197.
18. de Vries,N., Duinsbergen,D., Kuipers,E.J., Pot,R.G., Wiesenekker,P., Penn,C.W., van Vliet,A.H., Vandenbroucke-Grauls,C.M. and Kusters,J.G. (2002) Transcriptional phase variation of a type III restriction-modification system in *Helicobacter pylori*. *J. Bacteriol.*, **184**, 6615–6623.
19. Baldwin,D.N., Shepherd,B., Kraemer,P., Hall,M.K., Sycuro,L.K., Pinto-Santini,D.M. and Salama,N.R. (2007) Identification of *Helicobacter pylori* genes that contribute to stomach colonization. *Infect. Immun.*, **75**, 1005–1016.
20. Peek,R.M. Jr., Thompson,S.A., Donahue,J.P., Tham,K.T., Atherton,J.C., Blaser,M.J. and Miller,G.G. (1998) Adherence to gastric epithelial cells induces expression of a *Helicobacter pylori* gene, *iceA*, that is associated with clinical outcome. *Proc. Assoc. Am. Physicians*, **110**, 531–544.
21. Lin,T.L., Shun,C.T., Chang,K.C. and Wang,J.T. (2004) Isolation and characterization of a HpyC1I restriction-modification system in *Helicobacter pylori*. *J. Biol. Chem.*, **279**, 11156–11162.
22. Skoglund,A., Bjorkholm,B., Nilsson,C., Andersson,A.F., Jernberg,C., Schirwitz,K., Enroth,C., Krabbe,M. and Engstrand,L. (2007) Functional analysis of the M.HpyAIV DNA methyltransferase of *Helicobacter pylori*. *J. Bacteriol.*, **189**, 8914–8921.
23. Ausubel, F., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1997) *Short Protocols in Molecular Biology*. John Wiley & Sons, New York.
24. Amundsen,S.K., Fero,J., Hansen,L.M., Cromie,G.A., Solnick,J.V., Smith,G.R. and Salama,N.R. (2008) *Helicobacter pylori* AddAB helicase-nuclease and RecA promote recombination-related DNA repair and survival during stomach colonization. *Mol. Microbiol.*, **69**, 994–1007.
25. Menard,R., Sansonetti,P.J. and Parsot,C. (1993) Nonpolar mutagenesis of the *ipa* genes defines IpaB, IpaC, and IpaD as effectors of *Shigella flexneri* entry into epithelial cells. *J. Bacteriol.*, **175**, 5899–5906.
26. Copass,M., Grandi,G. and Rappuoli,R. (1997) Introduction of unmarked mutations in the *Helicobacter pylori vacA* gene with a sucrose sensitivity marker. *Infect Immun.*, **65**, 1949–1952.
27. Wang,Y., Roos,K.P. and Taylor,D.E. (1993) Transformation of *Helicobacter pylori* by chromosomal metronidazole resistance and by a plasmid with a selectable chloramphenicol resistance marker. *J. General Microbiol.*, **139**, 2485–2493.
28. Woodcock,D.M., Crowther,P.J., Doherty,J., Jefferson,S., DeCruz,E., Noyer-Weidner,M., Smith,S.S., Michael,M.Z. and Graham,M.W. (1989) Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucleic Acids Res.*, **17**, 3469–3478.
29. Kong,H., Lin,L.F., Porter,N., Stickel,S., Byrd,D., Posfai,J. and Roberts,R.J. (2000) Functional analysis of putative restriction-modification system genes in the *Helicobacter pylori* J99 genome. *Nucleic Acids Res.*, **28**, 3216–3223.
30. Peterson,J.D., Umayam,L.A., Dickinson,T., Hickey,E.K. and White,O. (2001) The comprehensive microbial resource. *Nucleic Acids Res.*, **29**, 123–125.
31. Bennett-Lovsey,R.M., Herbert,A.D., Sternberg,M.J. and Kelley,L.A. (2008) Exploring the extremes of sequence/structure

space with ensemble fold recognition in the program Phyre. *Proteins*, **70**, 611–625.

32. Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., McWilliam,H., Valentin,F., Wallace,I.M., Wilm,A., Lopez,R. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.

33. DeLano, W.L. (2002) The PyMOL Molecular Graphics System. Available at http://www.pymol.org.

34. Malone,T., Blumenthal,R.M. and Cheng,X. (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.*, **253**, 618–632.

35. 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.

36. Kobayashi,I. (2001) Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res.*, **29**, 3742–3756.

37. Ando,T., Xu,Q., Torres,M., Kusugami,K., Israel,D.A. and Blaser,M.J. (2000) Restriction-modification system differences in *Helicobacter pylori* are a barrier to interstrain plasmid transfer. *Mol. Microbiol.*, **37**, 1052–1065.

38. Sadykov,M., Asami,Y., Niki,H., Handa,N., Itaya,M., Tanokura,M. and Kobayashi,I. (2003) Multiplication of a restriction-modification gene complex. *Mol. Microbiol.*, **48**, 417–427.

39. Handa,N., Nakayama,Y., Sadykov,M. and Kobayashi,I. (2001) Experimental genome evolution: large-scale genome rearrangements associated with resistance to replacement of a chromosomal restriction-modification gene complex. *Mol. Microbiol.*, **40**, 932–940.

40. 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.

41. Aras,R.A., Small,A.J., Ando,T. and Blaser,M.J. (2002) *Helicobacter pylori* interstrain restriction-modification diversity prevents genome subversion by chromosomal DNA from competing strains. *Nucleic Acids Res.*, **30**, 5391–5397.

42. Takata,T., Aras,R., Tavakoli,D., Ando,T., Olivares,A.Z. and Blaser,M.J. (2002) Phenotypic and genotypic variation in methylases involved in type II restriction-modification systems in *Helicobacter pylori*. *Nucleic Acids Res.*, **30**, 2444–2452.

43. Vale,F.F. and Vitor,J.M. (2007) Genomic methylation: a tool for typing Helicobacter pylori isolates. *Appl. Environ. Microbiol.*, **73**, 4243–4249.

44. Vitkute,J., Stankevicius,K., Tamulaitiene,G., Maneliene,Z., Timinskas,A., Berg,D.E. and Janulaitis,A. (2001) Specificities of eleven different DNA methyltransferases of *Helicobacter pylori* strain 26695. *J. Bacteriol.*, **183**, 443–450.

45. Takahashi,N., Naito,Y., Handa,N. and Kobayashi,I. (2002) A DNA methyltransferase can protect the genome from postdisturbance attack by a restriction-modification gene complex. *J. Bacteriol.*, **184**, 6100–6108.

46. Falush,D., Kraft,C., Taylor,N.S., Correa,P., Fox,J.G., Achtman,M. and Suerbaum,S. (2001) Recombination and mutation during

long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. *Proc. Natl Acad. Sci. USA*, **98**, 15056–15061.

47. Casadesus,J. and Low,D. (2006) Epigenetic gene regulation in the bacterial world. *Microbiol. Mol. Biol. Rev.*, **70**, 830–856.

48. Dewhirst,F.E., Shen,Z., Scimeca,M.S., Stokes,L.N., Boumenna,T., Chen,T., Paster,B.J. and Fox,J.G. (2005) Discordant 16S and 23S rRNA gene phylogenies for the genus *Helicobacter*: implications for phylogenetic inference and systematics. *J. Bacteriol.*, **187**, 6106–6118.

49. Liao,D. (2000) Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA genes in bacteria and archaea. *J. Mol. Evol.*, **51**, 305–317.

50. Burge,C., Campbell,A.M. and Karlin,S. (1992) Over- and under-representation of short oligonucleotides in DNA sequences. *Proc. Natl Acad. Sci. USA*, **89**, 1358–1362.

51. Vandenbogaert,M. and Makeev,V. (2003) Analysis of bacterial RM-systems through genome-scale analysis and related taxonomy issues. *In Silico Biol.*, **3**, 127–143.

52. Pride,D.T., Meinersmann,R.J., Wassenaar,T.M. and Blaser,M.J. (2003) Evolutionary implications of microbial genome tetranucleotide frequency biases. *Genome Res.*, **13**, 145–158.

53. Gelfand,M.S. and Koonin,E.V. (1997) Avoidance of palindromic words in bacterial and archaeal genomes: a close connection with restriction enzymes. *Nucleic Acids Res.*, **25**, 2430–2439.

54. Ando,T., Aras,R.A., Kusugami,K., Blaser,M.J. and Wassenaar,T.M. (2003) Evolutionary history of *hrgA*, which replaces the restriction gene *hpyIIIR* in the hpyIII locus of Helicobacter pylori. *J. Bacteriol.*, **185**, 295–301.

55. Ando,T., Wassenaar,T.M., Peek,R.M. Jr., Aras,R.A., Tschumi,A.I., van Doorn,L.J., Kusugami,K. and Blaser,M.J. (2002) A *Helicobacter pylori* restriction endonuclease-replacing gene, *hrgA*, is associated with gastric cancer in Asian strains. *Cancer Res.*, **62**, 2385–2389.

56. Figueiredo,C., Quint,W.G., Sanna,R., Sablon,E., Donahue,J.P., Xu,Q., Miller,G.G., Peek,R.M. Jr., Blaser,M.J. and van Doorn,L.J. (2000) Genetic organization and heterogeneity of the *iceA* locus of *Helicobacter pylori*. *Gene*, **246**, 59–68.

57. van Doorn,L.J., Figueiredo,C., Sanna,R., Plaisier,A., Schneeberger,P., de Boer,W. and Quint,W. (1998) Clinical relevance of the *cagA*, *vacA*, and *iceA* status of *Helicobacter pylori*. *Gastroenterology*, **115**, 58–66.

58. Hastings,P.J. (2007) Adaptive amplification. *Crit. Rev. Biochem. Mol. Biol.*, **42**, 271–283.

59. 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.

60. 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.

61. Tomb,J.F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G., Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S., Dougherty,B.A. *et al.* (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature*, **388**, 539–547.