Genomic evolution and epigenetic DNA modification in *Helicobacter pylori* phenotypic adaptation and transcriptional regulation

Dissertation zum Erwerb des Doctor of Philosophy (Ph.D.) an der Medizinischen Fakultät der Ludwig-Maximilians-Universität zu München

vorgelegt von

Iratxe Estibariz

aus

Vitoria-Gasteiz, Spain

am

2020

Principal supervisor: Prof. Dr. med. Sebastian Suerbaum

Second supervisors: Prof. Dr. rer. nat. Christine Josenhans
Prof. Dr. med. Dr. rer. nat. Burkhard Tümmler

Dean: Prof. Dr. med. dent. Reinhard Hickel

Oral defence: 18.02.2020

TABLE OF CONTENTS

LIST OF MANUSCRIPTS

Manuscript I

Genome and methylome variation in *Helicobacter pylori* with a *cag* pathogenicity island during early stages of human infection

Sandra Nell*, **Iratxe Estibariz***, Juliane Krebes*, Boyke Bunk, David Y. Graham, Jörg Overmann, Yi Song, Cathrin Spröer, Ines Yang, Thomas Wex, Jonas Korlach, Peter Malfertheiner, and Sebastian Suerbaum

**Authors declared shared first co-authorship*

Gastroenterology 2018; 154(3):612-623 (doi: 10.1053/j.gastro.2017.10.014)

Manuscript II

The core genome ^{m5}C methyltransferase JHP1050 (M.Hpy99III) plays an important role in orchestrating gene expression in *Helicobacter pylori*

Iratxe Estibariz, Annemarie Overmann, Florent Ailloud, Juliane Krebes, Christine Josenhans*, and Sebastian Suerbaum*

*Shared corresponding authors

Nucleic Acids Research 2019; Mar 18; 47(5):2336-2348 (doi: 10.1093/nar/gky1307)

1. SUMMARY

The human gastric pathogen *Helicobacter pylori* establishes a long-term infection leading to chronic inflammation of the stomach. Most of the infected individuals do not suffer symptoms or the manifestation occurs after many years. In some cases, the infection leads to gastric ulcers, gastric cancer or MALT lymphoma. Higher probabilities to develop clinical evidence of *H. pylori* infection have been associated with virulence factors like the *cag* pathogenicity island (*cag*PAI) and the more toxic alleles of the cytotoxin VacA.

H. pylori displays a high mutation rate and frequent recombination during mixed infections, which contribute to its great allelic diversity and genome plasticity. The high level of genetic variation of *H. pylori* has been proposed to contribute to its adaptation to different human hosts.

Interestingly, despite the small genome size of *H. pylori*, the genome contains an exceptional number of strain-specific genes encoding Restriction-Modification (R-M) systems. They have been proposed to act as "primitive immune systems" in bacteria. However, other roles have been assigned to the methylation catalyzed by the MTases, including control of gene expression.

So far, it is not well understood how *H. pylori* takes advantage of its genome variability to adapt to novel stomach niches and the role of the extraordinary number of R-M systems. Moreover, there is not much known about how methylation influences the *H. pylori* transcriptome and how it modifies the phenotype.

In previous studies, *H. pylori* isolates were obtained at different intervals of time in chronically infected patients. Genome analysis of sequential isolates allowed studying the distribution of genetic changes, the calculation of mutation rates and recombination frequencies. Imports clustered in a non-arbitrary distribution. Thus, it was shown that there was a selection for the diversification of genes encoding outer membrane proteins (OMP), which have an impact on the interaction of the bacteria with the gastric epithelium.

At the onset of this thesis, only few studies had attempted to understand *H. pylori* genome evolution during acute infection. The diversification of genes during early colonization cannot be investigated using isolates from chronically infected patients since initial genetic changes would be hard to distinguish from long-term adaptation or might have disappeared following a population bottleneck. We have compared the genomes of isolates from human volunteers who were experimentally infected with a fully virulent *H. pylori* strain. The volunteers were given either placebo or a prophylactic vaccine candidate and subsequently challenged with *H. pylori*. Isolates were recovered 12 weeks (62 weeks in one case) post infection, and their genomes were sequenced. The data

showed that OMP-related genes were the ones most prone to vary between isolates and the mutation rates were in agreement with the ones calculated during chronic *H pylori* infections. Additionally, the activities of multiple virulence factors were affected during short-term infection. Interestingly, the gene encoding for the vacuolating cytotoxin VacA was inactivated in three isolates, suggesting that a vaccine-induced selection pressure modulated the activity of this virulence factor. Therefore, *H. pylori* exhibits a rapid genome evolution already within a few weeks of infection that may be crucial for the adaptation to novel stomach niches.

The R-M system portfolio varies between strains leading to variable methylomes. R-M systems are typically composed of a restriction enzyme (REase) and a methyltransferase (MTase). The REase cleaves foreign unmethylated DNA at a specific target site, while the MTase adds methyl groups at the same motif. Thus, R-M systems act as a barrier against invading DNA. Nevertheless, other functions have been attributed to methylation apart from self-DNA recognition, such as modulation of gene expression, control of the cell cycle and DNA replication. We compared the methylomes of isolates from human volunteers infected with *H. pylori* and found that variation in the methylomes was caused by a switch in the expression of phase-variable MTase genes, which might play a role in colonization by regulating gene expression.

So far, only few studies had shown that methylation impacts gene transcription in *H. pylori*. In this work, I focused on an extraordinarily conserved ^{m5}C-MTase gene (JHP1050) shared by every *H. pylori* strain. Analysis of 459 *H. pylori* genomes showed that the MTase gene was always present and putatively active. In contrast, the corresponding REase gene was found in 61 genomes only, and predicted to be functional in 15 of these. A phylogenetic analysis of the MTase and REase genes showed a tree structure similar to the overall population structure of the strains (as computed from seven gene multilocus haplotypes), suggesting that this particular R-M system was acquired early in the history of *H. pylori.*

We used RNA sequencing to analyze the transcriptome of two *H. pylori* wild type strains (J99 and BCM-300) and their corresponding isogenic MTase mutants. Transcriptome comparison of J99 and J99 MTase mutant showed 225 differentially expressed genes. In contrast, the transcriptomes of BCM-300 and the mutated strain exhibited 29 genes with different expression. Of those, 10 genes were differentially expressed in both, J99 and BCM-300. Moreover, changes in gene expression affected several phenotypic attributes such as adherence to host cells, bacterial competence for DNA uptake, copper susceptibility and cell morphology.

Using site-directed mutagenesis, we modified different GCGC motifs to GAGC sequences, which were not susceptible to methylation. The motifs were selected due to their location: one motif was located within an antitoxin gene and two motifs were placed within the 500 bp upstream region of the gene.

One of the upstream motifs overlapped with the -10 box of the predicted promoter. Quantification of the gene expression disclosed that the modification of the GCGC motif overlapping the promoter had a direct impact on gene expression. Similar result was observed when the whole R-M system was inactivated.

The work of the current thesis provided novel insights into *in vivo* genome and methylome modifications arising in the very first stages of *H. pylori* infection. Additionally, we contributed to the understanding of how the same MTase is able to modulate the expression of several genes and the phenotype of *H. pylori* in a strain-specific manner. Furthermore, we showed that motifs within regulatory regions have a direct impact upon transcription.

2. ABBREVIATIONS

^{m6}A	*N6*-methyladenine
^{m4}C	*N4*-methylcytosine
^{m5}C	5-methylcytosine
ANOVA	analysis of variance
ATP	adenosine triphosphate
BabA	blood group antigen-binding adhesin A
cagA	cytotoxin associated gene A
***cag*PAI**	*cag* pathogenicity island
CAT	chloramphenicol acetyl transferase
CDS	coding sequences
CEACAMS	carcinoembryonic antigen-related cell adhesion molecules
CH_3	methyl group
DEGs	differentially expressed genes
dsDNA	double-stranded DNA
ssDNA	single-stranded DNA
FC	fold change
FCS	fetal calf serum
gDNA	genomic DNA
HGT	horizontal gene transfer
IARC	International Agency for Research on Cancer
IL	interleukin
Km	kanamycin
L/D	live or dead
Le^b	Lewis b
LPS	lipopolysaccharides

MAMPs	microbe-associated molecular patterns
MLST	multi-locus sequence typing
MMR	mismatch repair
MTase	methyltransferase
MuGent	multiplex genome editing
NAP	neutrophil-activating protein
NGS	next generation sequencing
O/E	observed / expected
OMPs	outer membrane proteins
PPIs	proton pump inhibitors
PRRs	pattern recognition receptors
qPCR	quantitative PCR
REase	restriction endonuclease
R-M	restriction-modification
rRNA	ribosomal RNA
RUT	rapid urease test
S subunit	specificity subunit
SabA	sialic acid binding adhesin A
SAM	S-adenyl methionine
sLe$^{x/a}$	sialyl-Lewis x/a antigens
SMRT	Single Molecule, Real-Time
SNPs	single nucleotide polymorphisms
TA	toxin-antitoxin
TSS	transcription start site
T4SS	type IV secretion system
TLR's	toll-like receptors
VacA	vacuolating cytotoxin A
WHO	World Health Organization

3. INTRODUCTION

3.1. *Helicobacter pylori*: general microbiology, clinical aspects and therapies

3.1.1 Discovery of *H. pylori* and general microbiology

The gastric pathogen *Helicobacter pylori* was discovered in 1982 by the Australian scientists Barry J. Marshall and J. Robin Warren. They observed that a spiral bacterium was present in almost all gastric biopsies obtained from individuals with active chronic gastritis and with duodenal ulcers, and the majority of those with gastric ulcers (Warren and Marshall 1983). A pioneer study using gastric biopsies from patients allowed these two scientists to relate the presence of this previously uncharacterized pathogen with gastritis and peptic ulcers (Marshall and Warren 1984). To confirm the results, Barry J. Marshall infected himself with an *H. pylori* culture and subsequently developed severe active gastritis (Marshall, Armstrong et al. 1985). In 2005, Marshall and Warren were awarded with the Nobel Prize in Physiology or Medicine for the discovery of a human pathogen causing gastric diseases.

H. pylori is a spiral-shaped, Gram-negative bacterium that belongs to the phylum *Proteobacteria* and to the class *Epsilonproteobacteria*. The rotation of 4-6 flagella at one of the bacterial poles provides motility to *H. pylori*. The growth of *H. pylori* depends on microaerobic conditions (lower levels of oxygen, 5%), elevated concentrations of CO_2 (5-10%), and a temperature of 37°C (Bury-Moné, Kaakoush et al. 2006). *H. pylori* is catalase, oxidase and urease positive; thus, tests based on the activity of these enzymes are classically used in the identification of the infection (Kusters, van Vliet et al. 2006).

3.1.2. Clinical aspects and therapies

About half of the world's population is infected by *H. pylori*, but the infection rates differ among countries, with a higher prevalence in places where the access to appropriate health and sanitary conditions is restricted (Eusebi, Zagari et al. 2014, Peleteiro, Bastos et al. 2014). A recent systematic review and meta-analysis about the prevalence of *H. pylori* indicated that African countries have the highest rates of infection (70.1%) while the lowest prevalence was found in Oceania (24.4%). In Germany, the infection rate was estimated to be 35.5% (Hooi, Lai et al. 2017).

H. pylori is likely to establish a life-long infection when not treated. While the majority of the patients do not develop clinical symptoms, chronic gastritis can progress to several diseases like gastric or duodenal ulcers (10%) or even gastric cancer and lymphoma of the mucosa-associated lymphoid tissue (MALT) (1%) (Suerbaum and Michetti 2002). Stomach cancer is, based on data provided by *The International Agency for Research on Cancer* (IARC), the fifth most common cancer in the world and

the third malignancy in terms of mortality (Bray, Ferlay et al. 2018). *H. pylori* is associated with 90% of all new stomach cancers; thus, it has been classified since 1994 by the *World Health Organization* (WHO) as a class I carcinogenic agent (IARC 1994).

Eradication of *H. pylori* is a difficult task. The mucus layer of the stomach provides protection to the pathogen, making the bacteria difficult to approach by antimicrobials (Shimizu, Akamatsu et al. 1996). Established treatment regimens consist of a combination of antibiotics and proton pump inhibitors (PPIs) during 7-14 days, in order to ensure the eradication of the infection (Malfertheiner, Megraud et al. 2017, Zagari, Rabitti et al. 2017). However, a rapid increase in antimicrobial resistance has led to the inclusion of *H. pylori* in the list of high-risk pathogens by the WHO in 2017, calling for the development of novel treatments against this bacterium (WHO 2017).

Substantial efforts have been made to develop a vaccine against *H. pylori*. Vaccines could prevent the acquisition of the bacteria, avoid future infections and reduce antimicrobial resistance. However, the majority of the vaccine candidates under development were abandoned after preclinical or phase I trials (Michetti, Kreiss et al. 1999, Banerjee, Medina-Fatimi et al. 2002, Czinn and Blanchard 2011, Moss, Moise et al. 2011, Sutton and Boag 2018). Two types of vaccines have been mostly developed and applied: a prophylactic vaccine to prevent the effects of future infections via the stimulation of the immune responses or a therapeutic vaccine given after the infection to alter the disease outcome.

Prior to clinical studies to test vaccine candidates, there was a need to develop a safe human model to study *H. pylori* infection and the immune response. This model was established by Graham and colleagues. Human volunteers were infected with a *cag* Pathogenicity Island (*cag*PAI) negative *H. pylori* strain (BCS 100) and treated with antibiotics to eradicate the infection after 4 or 12 weeks. With this study, the authors demonstrated that the infection and eradication of *H. pylori* were possible in human volunteers and that the subjects developed typical symptoms associated with *H. pylori* infection (Graham, Opekun et al. 2004). Afterwards, clinical studies used human volunteers to test prophylactic vaccine candidates. In one study, individuals where challenged with the *cag*PAI negative *H. pylori* strain BCS 100 and subsequently given a *Salmonella* Ty21a vaccine expressing the *H. pylori* urease A and B subunits (Aebischer, Bumann et al. 2008). Later, Malfertheiner and colleagues evaluated a prophylactic vaccine candidate (expressing three recombinant *H. pylori* proteins: CagA, VacA and NAP) in healthy human volunteers who were subsequently challenged with a fully pathogenic *H. pylori* strain (BCM-300) (Malfertheiner, Selgrad et al. 2018). In both cases, even though well tolerated and able to induce an immune response, the vaccines were ineffective in protecting the human volunteers against a new *H. pylori* infection (Aebischer, Bumann et al. 2008, Malfertheiner, Selgrad et al. 2018).

Thus, so far, there is no vaccine available against *H. pylori*. Some of the reasons leading to the difficulties in the improvement of functionally active vaccines are the high genetic variability of *H. pylori* and its ability to escape from the immune system (Aebischer, Schmitt et al. 2005). Only one vaccine candidate reached clinical phase III. The recombinant vaccine was provided to children, and after 1 and 3 years follow up, the authors observed a reduction in *H. pylori* infection rate (Zeng, Mao et al. 2015). However, there is no more information available if the vaccine study was continued.

3.2. Colonization and virulence factors

H. pylori is restricted to the extreme environment provided by the stomach niche. The bacterium is able to adhere to the gastric epithelium or swim in the mucus layer. In order to ensure a successful colonization, *H. pylori* requires a battery of resources. Once *H. pylori* has been acquired, the bacteria must face the acidic pH in the gastric fluid (pH of 1-2 between meals) by activating cytoplasmic urease (Kusters, van Vliet et al. 2006). Next, *H. pylori* is able to move along the gastric epithelium due to flagella-based chemotactic motility (Schreiber, Konradt et al. 2004, Lertsethtakarn, Ottemann et al. 2011, Behrens, Schweinitzer et al. 2013). *H. pylori* swims to the mucus layer and multiplies when an optimal pH 5-7 is reached. Part of the bacterial population attaches to the gastric epithelial cells using various adhesins. Finally, the bacterium secretes several virulence factors, like the cytotoxin-associated gene product A (CagA) and the vacuolating cytotoxin A (VacA) (Dunne, Dolan et al. 2014, Kao, Sheu et al. 2016) (Figure 1).

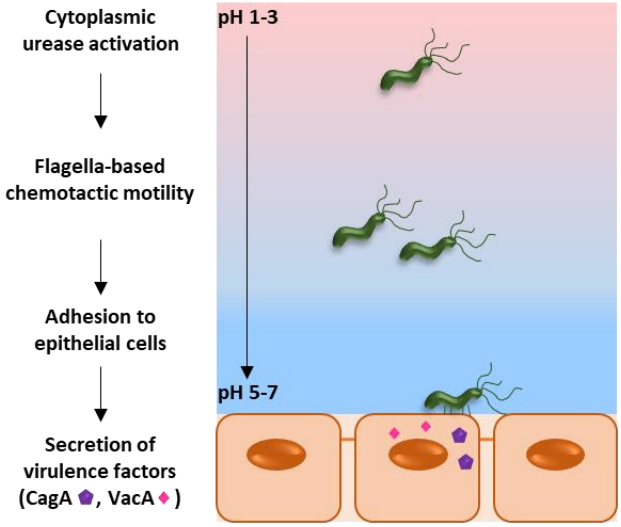


Figure 1. Process of *H. pylori* infection from the gastric fluid (acid pH) to the epithelium (neutral pH). *H. pylori* confronts the acid pH by activating cytoplasmic urease. Urea is hydrolyzed, leading to the buffering of the periplasm and adjusting the membrane potential. Chemotactic motility allows the bacterium to move through the gastric epithelium and reach neutral pH. Attachment to epithelial cells takes place due to a battery of adhesins. Finally, several virulence factors are secreted.

3.2.1. Inflammation and immune evasion

In order to establish a life-long infection, *H. pylori* must evade the immune system. When the bacterium is transmitted to a new host, two scenarios are possible: the infection is cleared by the immune response or, in contrast, the gastric pathogen is able to escape from the immune system and establish a persistent infection leading to gastric inflammation (Baldari, Lanzavecchia et al. 2005). To successfully establish a chronic infection, *H. pylori* is able to disrupt and modulate the immune system and ultimately escape immunity-mediated clearance.

Every individual infected with *H. pylori* develops active gastritis. *H. pylori* uses several mechanisms like virulence factors and outer membrane proteins (OMPs) to recruit immune cells to the site of infection and induce inflammation in the gastric epithelial cells (Suerbaum and Michetti 2002, Lamb and Chen 2013). The activation of several host transcription factors induces the up-regulation of inflammatory molecules like cytokines and chemokines. Inflammatory cytokines IL-1, IL-2 IL-6, IL-8 and TNF are overexpressed in *H. pylori* infected patients (Suerbaum and Michetti 2002).

Lipopolysaccharides (LPS) and flagellin are common examples of *H. pylori* immune evasion. The best-studied pattern recognition receptors (PRRs) are the toll-like receptors (TLRs). LPS act as microbe-associated molecular patterns (MAMPs) and they can be recognized mostly by TLR4 in Gram-negative bacteria (Takeuchi, Hoshino et al. 1999, Mogensen 2009). Modifications within the lipid core A domain of *H. pylori*'s LPS lead to lower activation of TLR4 (Salama, Hartung et al. 2013, Varga and Peek Jr. 2017). TLR5 senses bacterial flagellins. However, *H. pylori* flagellins (FlaA and FlaB) avoid TLR5 detection due to modifications in the TLR5 recognition site (Lee, Stack et al. 2003, Gewirtz, Yu et al. 2004, Andersen-Nissen, Smith et al. 2005).

Virulence factors can also modulate the immune system. VacA can suppress adaptive immune activity by targeting lymphocytes and disrupt, for example, the phagocytic killing of *H. pylori*. In addition, VacA interferes with T-cell proliferation, production of cytokines and dendritic cells, modulating their normal function (Djekic and Müller 2016).

3.2.2. Urease

In order to cope with the low pH present in the stomach lumen, *H. pylori* produces urease, a cytoplasmic enzyme, as a first line of defense. Urease is one of the most predominant enzymes in *H. pylori*, representing 10-15% of the total protein content (Bauerfeind, Garner et al. 1997). Regulation of urease levels depends on the availability of the cofactor nickel (van Vliet, Kuipers et al. 2001, de Reuse, Vinella et al. 2013). The urease gene cluster consists of two operons. The first operon contains the two structural subunits *ureAB* and the second operon, located downstream, harbours the accessory *ureIEFGH* genes (Mobley 2001). At an external pH < 6.0, urea goes through the outer and inner membranes thanks to the urea channel UreI and meets activated cytoplasmic urease (Krulwich,

Sachs et al. 2011). Hydrolysis of urea leads to the buffering of the periplasm and adjusting the transmembrane potential to adequate levels that allow protein synthesis and growth in acid environments (Scott, Weeks et al. 1998). Direct measurements showed that the periplasmic pH is increased to and maintained at pH 6.2 when the medium is acidic (Wen, Scott et al. 2018). Thus, urease is considered a key virulence factor that enables *H. pylori* survival in the acidic stomach (Eaton, Brooks et al. 1991).

3.2.3. Motility and chemotaxis

Motility based on flagella and chemotaxis is an essential factor for colonization and establishing the infection, since it allows the pathogen to migrate and move through the mucus layer (O'Toole, Lane et al. 2000, Josenhans and Suerbaum 2002, Lertsethtakarn, Ottemann et al. 2011). *H. pylori* possesses 4-6 unipolar flagella, which provide motility (Geis, Leying et al. 1989, Josenhans, Eaton et al. 2000). *H. pylori* harbors core chemotactic proteins (CheA, CheW and CheY) and four chemoreceptors (TlpA, TlpB, TlpC and TlpD) (Lertsethtakarn, Ottemann et al. 2011). TlpA, TlpB and TlpC are integral membrane proteins while TlpD is a cytoplasmic protein. Several compounds influence *H. pylori* chemotactic activity like urea, lactate, mucins, and others (Spohn 2001, Croxen, Sisson et al. 2006, Behrens, Schweinitzer et al. 2013).

3.2.4. Outer membrane proteins and adhesins

H. pylori attaches to the epithelial cells using a set of OMPs that can act as adhesins. Attachment to the gastric epithelial cells is possible due to the interaction between the bacterial adhesins and host oligosaccharides, glycans, and host surface-proteins (Testerman, McGee et al. 2001). It has been described that 4% of *H. pylori*'s genome encode OMPs (Dossumbekova, Prinz et al. 2006). The Hop family of proteins is the biggest family of OMPs, followed by the Hor, Hof and Hom families (Tomb, White et al. 1997, Alm, Bina et al. 2000). Many of the known *H. pylori* adhesins belong to the Hop family of OMPs. Among others, the major adhesins for which the host receptor has been identified so far are BabA, SabA and HopQ.

BabA: the blood group antigen-binding adhesin A binds to fucosylated Lewis b (Le^b) antigen and related ABO blood group antigens present on epithelial cells and gastric mucins (Borén, Falk et al. 1993, Aspholm-Hurtig, Dailide et al. 2004). BabA binding to Le^b is reversible and acid-sensitive, with higher affinities when the pH increases (Bugaytsova, Bjornham et al. 2017). Gene conversion with the paralogous gene *babB* and phase variation led to the loss of *babA* expression during chronic infection of Rhesus monkeys and in some clinical isolates (Solnick, Hansen et al. 2004). The loss of *babA* expression was not dependent on Le^b binding, suggesting additional functions of this important adhesin (Solnick, Hansen et al. 2004, Nell, Kennemann et al. 2014, Hansen, Gideonsson et al. 2017). Modifications in the C-terminus of BabA can affect the expression and binding abilities of the protein,

since the C-terminus has been reported to be relevant for the correct protein folding (Nell, Kennemann et al. 2014).

SabA: the sialic acid binding adhesin recognizes the sialyl-Lewis x/a antigens (sLe^x and sLe^a) that are usually replacing the normal Le antigens during chronic gastric inflammation and gastric carcinoma (Mahdavi, Sonden et al. 2002). *H. pylori* binds to neutrophils via SabA, inducing an oxidative burst (Mahdavi, Sonden et al. 2002, Unemo, Aspholm-Hurtig et al. 2005).

HopQ: It has been recently discovered that HopQ binds to several human and murine carcinoembryonic antigen-related cell adhesion molecules (CEACAMS) (Javaheri, Kruse et al. 2016, Königer, Holsten et al. 2016). This particular adhesin seems to be important in the pathogenesis of *H. pylori* since its interaction is essential for the translocation of the oncoprotein CagA into the host cells (Javaheri, Kruse et al. 2016, Königer, Holsten et al. 2016).

Although some adhesins and their roles in infection are described in *H. pylori*, the great numbers of OMPs and still uncharacterized genes suggest there might be other undescribed adhesins contributing to colonization.

3.2.5. The *cag*PAI and the oncoprotein CagA

The *cag*PAI is one of the major virulence factors of *H. pylori* and it has been associated with a higher risk of developing gastric diseases (Blaser, Perez-Perez et al. 1995, Parsonnet, Friedman et al. 1997). The *cag*PAI consists of a 40 kb chromosomal region containing more than 30 predicted genes coding for a Type IV Secretion System (T4SS) (Censini, Lange et al. 1996, Olbermann, Josenhans et al. 2010). The T4SS of *H. pylori* includes genes that are considered orthologs of components of the Vir T4SS system of *Agrobacterium tumefaciens*, but also additional genes. The products of some genes were shown to be required for the functionality of the *cag*PAI (Fischer, Puls et al. 2001).

The presence of the *cag*PAI in *H. pylori* strains varies among geographical regions, with a prevalence of approximately 60% in Western countries and 100% in Asia (Yamaoka, Kodama et al. 1999, Gressmann, Linz et al. 2005, Olbermann, Josenhans et al. 2010). The strains harboring the island, *cag*PAI+ strains, are usually associated with a higher risk of developing gastric malignancies due to a combination of host and strain factors (Figueiredo, Machado et al. 2002, Noto and Peek Jr. 2012, Cover 2016). One important bacterial element associated with gastric cancer is the oncoprotein CagA. This protein contains multiple tyrosine phosphorylated (EPIYA) motifs in the carboxy-terminal region. The EPIYA motifs are essential for the association of CagA with the membrane (Higashi, Yokoyama et al. 2005). The type and number of EPIYA motifs has been associated with the risk of developing gastric cancer and ulcers, although several studies displayed controversial results (Jones, Joo et al. 2009, Li, Liu et al. 2017).

The oncoprotein CagA is injected into the gastric epithelial cells and subsequently phosphorylated by host cellular kinases, triggering multiple effects on cellular signal transduction cascades and morphological alterations in the host cells, like the "hummingbird phenotype", characterized by cell elongation. CagA binds and deregulates the SHP-2 Tyrosine Phosphatase that is an oncoprotein prone to mutate in several human cancers (Hatakeyama 2004) (Figure 2).

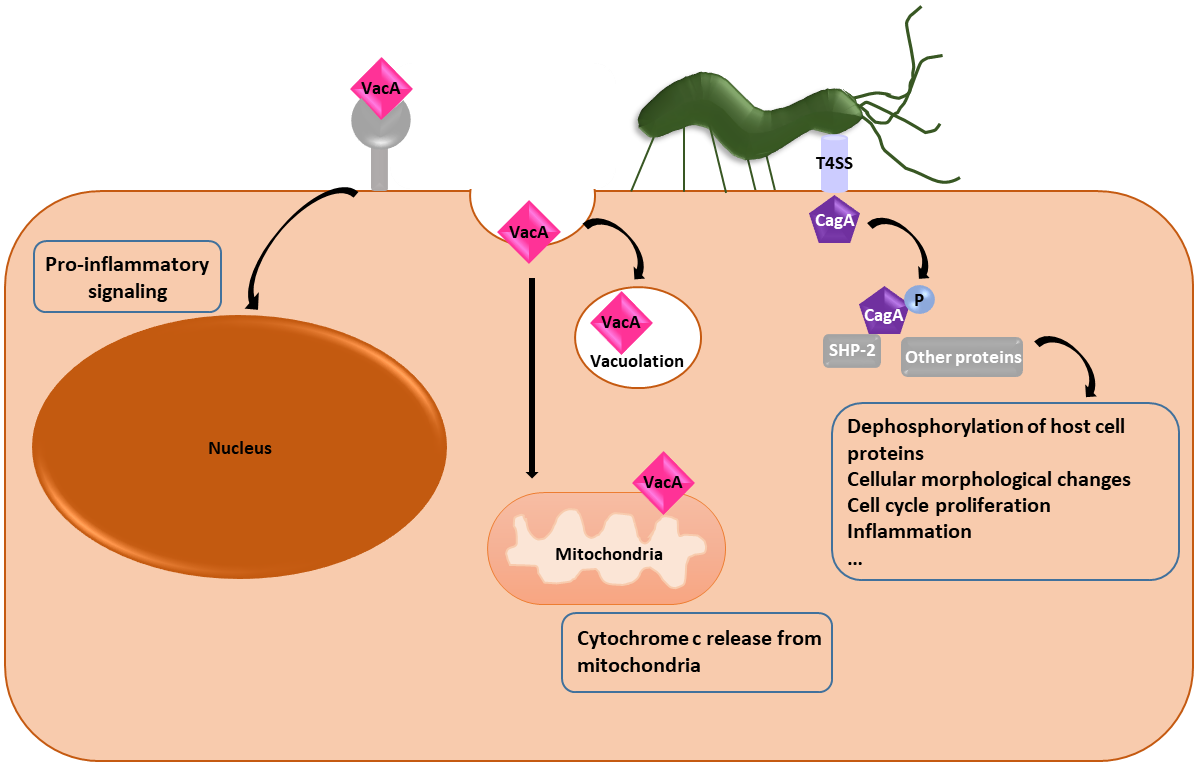


Figure 2. Representation of two major virulence factors, CagA and VacA. The oncoprotein CagA is injected in the epithelial cell and phosphorylated. CagA binds to several proteins triggering multiple effects on the host cell. VacA produces several effects on the host cell. Some examples are the formation of vacuoles, the localization in the mitochondria where produces cytochrome c release and possibly apoptosis, and binding to cell-membrane receptors activating pro-inflammatory signaling.

3.2.6. Vacuolating cytotoxin A (VacA)

VacA is considered a multifunctional toxin in *H. pylori*, causing a wide range of effects on the host cells (Cover and Blanke 2005). VacA molecules are secreted to the extracellular space or they can stay on the surface of *H. pylori* (Foegeding, Caston et al. 2016). Several alleles are present among strains and were shown to possess different abilities to cause cell damage. The *s1m1* type is the most pathogenic allelic variant. Strains carrying the *s1m1* allele have been linked to a higher risk of developing peptic ulcers (Atherton, Cao et al. 1995).

VacA intoxicates several types of human cells, including immune cells, leading to distinct alterations. Internal vacuolization was the first direct effect caused by VacA that was discovered (de Bernard,

Arico et al. 1997). VacA is able to modify mitochondrial permeability via reduction of the mitochondrial transmembrane potential and cytochrome c release (Willhite and Blanke 2004). The toxin also acts on several immune cells driving to localized immunosuppression and production of pro-inflammatory cytokines (Gebert, Fischer et al. 2003, Montecucco and de Bernard 2003) (Figure 2).

3.3. Mechanisms driving genetic diversity in *H. pylori*

Every infected individual is believed to carry a unique *H. pylori* strain, since the bacterium displays an extraordinarily high genetic diversity that is greater than in most other bacteria (Achtman, Azuma et al. 1999). Whole genome analysis of the first two sequenced *H. pylori* strains (J99 and 26695) revealed that 6% of the gene content differed between strains (Tomb, White et al. 1997, Alm, Ling et al. 1999). *H. pylori*'s genetic variability is thought to play an important role in adaptation to different human hosts and to the variable conditions of the gastric environment. Several mechanisms are responsible for the extraordinary sequence variability of the gastric pathogen.

3.3.1. Mutation

H. pylori mutation rates (10^{-5} - 10^{-7}) are several orders of magnitude higher than the ones found for the majority of other bacterial species (Björkholm, Sjolund et al. 2001). This phenomenon is related to the lack of a classic DNA mismatch repair (MMR) system and the hypermutator role of its DNA polymerase I (Garcia-Ortiz, Marsin et al. 2011). Moreover, homopolymeric and dinucleotide repeats are prone to phase-vary by slipped-strand mispairing, switching the activity of genes when they are located in promoters or within gene sequences (Josenhans, Eaton et al. 2000, Salaun, Linz et al. 2004, Kraft and Suerbaum 2005, Baltrus, Blaser et al. 2009). Thus, spontaneous mutations within coding sequences or regulatory regions can alter the amino acid sequence of the translated protein, leading to modified functions of the protein (gain or loss) and therefore, modify the fitness of the bacteria.

3.3.2. Recombination

Despite the high mutation rate, recombination is the process introducing most of the allelic changes in the genome during mixed infections with two or more *H. pylori* strains (Suerbaum, Maynard Smith et al. 1998, Kang and Blaser 2006, Didelot, Nell et al. 2013). The natural competence of *H. pylori* allows the uptake of exogenous double-stranded DNA (dsDNA) by the ComB system (Hofreuter, Odenbreit et al. 2001, Stingl, Muller et al. 2010, Dorer, Sessler et al. 2011). Following uptake, dsDNA is transformed to single-stranded DNA (ssDNA), transferred to the cytoplasm and able to recombine with the recipient's genome after interaction with RecA (Fischer and Haas 2004, Dorer, Sessler et al. 2011). Acquisition of imports by *H. pylori* generates a bimodal distribution of import lengths, with short (less than 50 bp) and long (peak length 1,645 bp) patches of imported sequences (Bubendorfer,

Krebes et al. 2016). Hence, recombination is a key mechanism driving to genomic evolution and plasticity in *H. pylori*.

3.3.3. *H. pylori* populations reflect human migrations

Multi Locus Sequence Analysis (MLSA) based on the sequencing of 7 housekeeping genes in a collection of 370 strains from 27 geographical and ethnic human regions led to the assignment of *H. pylori* to several modern populations and permitted the reconstruction of inferred ancestral populations (Falush, Wirth et al. 2003) (Table 1). The high genetic diversity of *H. pylori* and the coevolution and migrations with its human host, which began at least 80,000 years ago, resulted in the separation into phylogeographic populations that reflect human migrations (Linz, Balloux et al. 2007, Moodley, Linz et al. 2009, Breurec, Guillard et al. 2011, Moodley, Linz et al. 2012).

Table 1. Phylogeographic modern populations, subpopulations and ancestral population of *H. pylori*.

Modern population	Modern subpopulation	Ancestral population
hpAfrica2		Ancestral Africa 2
hpAfrica1	hspSAfrica	Ancestral Africa 1
	hspWAfrica	
	hspCAfrica	
hpNEAfrica		Ancestral Europe 2
hpEurope		Ancestral Europe 1 and 2
hpSahul		Ancestral Sahul
hpEastAsia	hspEAsia	Ancestral East Asia
	hspAmerind	
	hspMaori	
hpAsia2		Ancestral Europe 1

3.4. The *H. pylori* methylome: diversity of the Restriction-Modification (R-M) systems

DNA methylation is the process that occurs by the addition of methyl groups (CH_3) from the donor S-adenyl methionine (SAM) to DNA sequences. In eukaryotes, methylation typically occurs at the fifth position of the pyrimidine ring of cytosines leading to 5-methylcytosine (^{m5}C). In mammals, most of the ^{m5}C-methylations take place at cytosine residues as part of dinucleotide cytosine-guanine sequences, called CpG islands. In prokaryotes, there are another two types of methylation where CH_3

is added to the N6 position of adenines resulting in N^6-methyladenine (^{m6}A), or to the N4 position of cytosines generating N^4-methylcytosine (^{m4}C) (Figure 3).

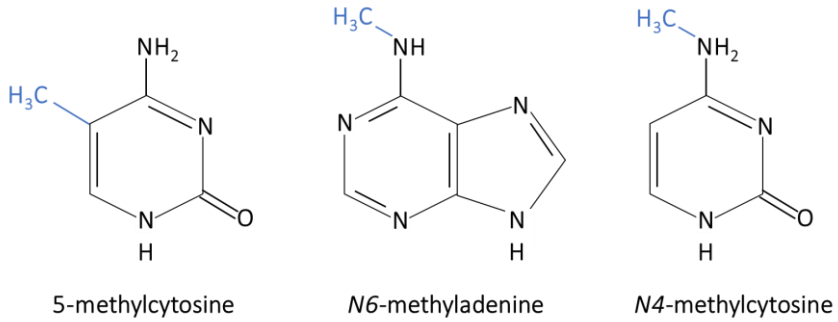


Figure 3. Chemistry of methylated nucleotides. From left to right: 5-methylcytosine (^{m5}C), *N6*-methyladenine (^{m6}A) and *N4*-methylcitosine (^{m4}C).

Methylation is catalyzed by methyltransferase enzymes (MTases), which add methyl groups to adenines or cytosines of a target motif. MTases in prokaryotes are often part of Restriction-Modification (R-M) systems. Such systems are usually composed of one MTase that methylates host DNA, and the restriction enzyme (REase) that cleaves foreign unmethylated DNA at the same target sequence. The R-M systems can be divided into four main groups based on the mechanism recognizing the target motif.

3.4.1. The discovery and functions of the R-M systems

In the early 50s, it was observed that one bacteriophage was able to grow in a particular bacterial host but was restricted in growth in other strains. This singularity was called by the investigators "host-induced variation" (Luria and Human 1952, Bertani and Weigle 1953). The observation inspired many researchers to understand the mechanism responsible for this phenomenon. In 1978, W. Arber, D. Nathans and H. Smith shared the Nobel Prize in Physiology and Medicine for the discovery of DNA-specific sequences, the enzymes in charge of cleaving DNA and their application in molecular genetics (Arber and Dussoix 1962, Dussoix and Arber 1962, Smith and Wilcox 1970, Danna and Nathans 1971). Since their discovery, REases have been used in molecular biology as genetic engineering tools.

R-M systems are widely distributed among bacterial species, as well as in archaea and eubacteria (Vasu and Nagaraja 2013). They have been classified as "primitive immune systems" (Bickle 2004), due to their protective role against exogenous DNA, contributing to bacterial genomic evolution and limiting horizontal gene transfer (HGT) (Gogarten, Doolittle et al. 2002). Although it has been shown that R-M systems are successful entities cutting heterologous DNA, their effectiveness diminishes with homeologous incoming DNA (Bubendorfer, Krebes et al. 2016). Moreover, R-M systems have

been suggested to act as selfish-mobile genetic elements causing adverse effects on the host cell (Kobayashi 2001). However, some MTases are not associated to any R-M system. They are known as "orphan" MTases.

In addition to self-DNA protection, methylation plays additional roles in bacteria. In *E. coli*, the Dam-MTase, methylating G^{m6}ATC sites, is a key element in DNA replication. SeqA binds to hemimethylated GATC sites in the *oriC* and in the promoter of *dnaA*, sequestering the origin of replication. The sequestration avoids re-initiation of the DNA replication because the *oriC* and the *dnaA* promoter must be fully methylated (Messer, Bellekes et al. 1985, Russell and Zinder 1987, Bogan and Helmstetter 1997). Another well-studied example is the regulation of the cell cycle in *Caulobacter crescentus* by the CcrM-MTase methylating G^{m6}ANTC motifs. Methylation within the promoter of the regulatory protein CtrA during a certain period adjusts the replication of the chromosome to only one per cycle (Berdis, Lee et al. 1998, Kozdon, Melfi et al. 2013, Gonzalez, Kozdon et al. 2014).

Furthermore, many MTases are prone to phase-variation and can coordinate the switch of the expression of several genes, which is called "phasevarion" (Srikhanta, Maguire et al. 2005, Srikhanta, Fox et al. 2010). Phase variable MTases and associated phasevarions have been found in many bacterial pathogens, including *Haemophilus influenzae*, *Neisseria meningitidis* and *H. pylori* (Fox, Dowideit et al. 2007, Srikhanta, Dowideit et al. 2009, Srikhanta, Gorrell et al. 2011).

3.4.2. Classification of R-M systems

The R-M systems are classified in four main groups. Type I, II and III are represented in Figure 4:

Type I R-M systems: These are the most complex type of R-M systems since they consist of multi-subunit proteins functioning as one module (Dryden, Murray et al. 2001). Furthermore, Type I R-M systems can be sub-divided into four categories (A, B, C, D) (Roberts, Belfort et al. 2003). They comprise three genes: the REase or R subunit (*hsdR*), the MTase or M subunit (*hsdM*) and the specificity subunit (*hsdS*). The *hsdS* carry two target recognition domains (TRDs) defining the sequence that will be recognized by the R-M system. Type I R-M system genes form a $R_2M_2S_1$ complex that requires adenosine triphosphate (ATP) hydrolysis. The entity M_2S_1 methylates the target motif in the absence of the REase (Murray 2000, Roberts, Belfort et al. 2003, Kennaway, Obarska-Kosinska et al. 2009). So far, most of the Type I R-M systems methylate adenines within asymmetric motifs, although few systems methylating cytosines have been recently identified (Morgan, Luyten et al. 2016). The target sequences typically comprise two segments of 3-4 bp separated by a spacer of 6-8 bp (Murray 2000).

Type II R-M systems: Type II REases and MTases usually act as monomers independently of each other. Type II MTases transfer the methyl-group from the donor SAM to adenines or cytosines, generating ^{m6}A, ^{m4}C and ^{m5}C types of methylation (Roberts, Belfort et al. 2003). Typically, Type II R-M

systems recognize palindromic sequences, but multiple exceptions exist, leading to several sub-groups. There are numerous criteria to classify the Type II R-M systems in the different sub-categories; therefore, many Type II R-M systems can be allocated within more than one group. Richard J. Roberts and colleagues summarized all the Type II sub-categories in (Roberts, Belfort et al. 2003). In *H. pylori*, Type II R-M systems are the most predominant ones as it has been shown in various methylome studies (Kong, Lin et al. 2000, Krebes, Morgan et al. 2014, Lee, Anton et al. 2015).

Type III R-M systems: The *mod* gene coding for the MTase or Mod and the *res* gene encoding the REase or Res, compose the Type III R-M systems. While two Mod subunits can achieve the enzymatic activity without the Res, the M_2R_2 complex is required for ATP-dependent cleavage (Dryden, Murray et al. 2001, Rao, Dryden et al. 2014). Many Type III R-M systems have been described to be phase-variable due to simple repetitive DNA sequences that are prone to length changes via slipped strand mispairing. The reversible ON/OFF switch of the Mod activity enables modifications in the methylome, driving to phasevarion (Srikhanta, Maguire et al. 2005).

Type IV R-M systems: These type of systems differ from the other three since they cleave modified DNA target sequences, including methylation, hydroxymethylation or glucosyl-hydroxymethylation (Vasu and Nagaraja 2013). This class of R-M systems are formed by one or two genes and their activity is not ATP-dependent.

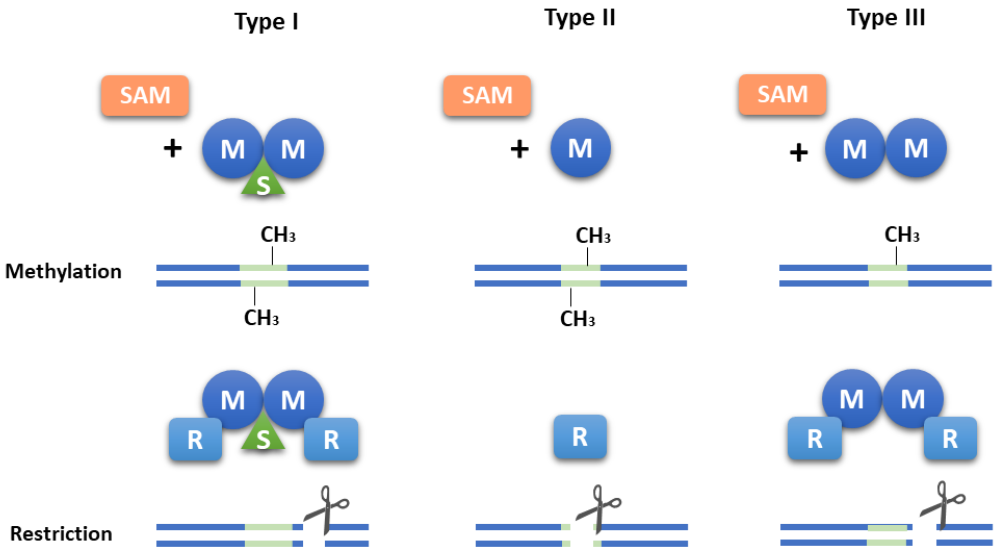


Figure 4. Classification of R-M systems. R-M systems add methyl groups to the target motif using SAM as donor. When the motif is not methylated, REases act and cleave the DNA. Cleavage of DNA by Type I R-M systems occur at a variable location away from the motif. Type II R-M systems usually cut within the palindromic motif. Type III R-M systems cleave at a fixed position (25-27 bp) from the target sequence (Srikhanta, Fox et al. 2010). M refers to MTase, S to specificity subunit, R to REase and SAM means S-adenyl methionine. The blue lines are the dsDNA and the motif is colored in green.

3.4.3. The R-M systems in the gastric pathogen *H. pylori*

Despite its small genome, *H. pylori* encodes an extraordinarily high number of R-M systems, where every strain carries a unique set of R-M system genes leading to variable methylomes (Nobusato, Uchiyama et al. 2000, Roberts, Vincze et al. 2015). More than half of the strain-specific genes of *H. pylori* code for R-M systems (Tomb, White et al. 1997).

The first two *H. pylori* methylomes from the strains 26695 and J99-R3 were studied using Single Molecule Real-Time (SMRT®) sequencing technology (Krebes, Morgan et al. 2014). The DNA polymerase used by the SMRT® sequencing technology catalyzes the incorporation of fluorescently labeled nucleotides. The kinetics of the polymerase are altered when there are modifications on the DNA sequence, such as methylation. Thus, every DNA modification produces different kinetics, allowing the identification of methylated nucleotides on the DNA (Schadt, Banerjee et al. 2013, Ardui, Ameur et al. 2018).

In the past years, many other *H. pylori* methylomes have been published, confirming methylomic inter-strain diversity (Krebes, Morgan et al. 2014, Lee, Anton et al. 2015, Lamichhane, Chua et al. 2019) and, indeed, few shared methylated motifs have been found between *H. pylori* strains (Vale, Megraud et al. 2009). Furthermore, using SMRT® sequencing technology, many novel recognition sites and their respective associated R-M system genes have been identified.

Although the role of some MTases in epigenetic regulation in *H. pylori* has been elucidated (Donahue, Israel et al. 2002, Srikhanta, Gorrell et al. 2017, Kumar, Karmakar et al. 2018), the reason why *H. pylori* possesses so many R-M systems and their function is still not well understood. Functionally active R-M systems might promote homeologous recombination between *H. pylori* strains generating allelic diversity in a given population. Nevertheless, there are many orphan MTases that are not part of an active R-M system since the REase gene is truncated or absent, whose functions are still undescribed.

4. AIMS

The comparison of the first two complete genome sequences of *H. pylori* strains identified great genetic diversity, which was higher than of most other bacteria. Analysis of sequential isolates from chronically infected individuals showed that *H. pylori* displays high mutation and recombination rates. It was observed that OMP genes were found to be prone to modifications. Despite the contribution of these studies to the knowledge of *H. pylori* genetic diversity, the study of genomes from chronic isolates does not allow the investigation of genetic modifications arising in the first stages of the infection. Moreover, *H. pylori* diversity also comes from strain-specific genes encoding R-M systems, leading to variable methylomes between strains. DNA modifications that produce changes in gene expression but do not alter the gene sequence are known as epigenetic modifications. Methylation protects bacterial genomes from invading DNA, but also plays key roles in their physiology. Up to now, the majority of the bacterial methylome studies focused on methylation within adenine residues, since it is the most common type of modification in bacteria.

However, the molecular mechanisms contributing to genetic and epigenetic evolution and the influence of methylation in transcription are so far not well understood. During this thesis, I tried to address questions to improve our understanding of genetic and methylomic evolution during early adaptation to novel stomach niches and the influence of methylation on gene transcription and *H. pylori* physiology.

At the time of this writing, there is no licensed vaccine capable of preventing *H. pylori* infections. The potential to escape from the immune system by the high genetic variability of this bacterium may make the development of a successful vaccine difficult. However, this hypothesis had not been formally proven yet. We obtained *H. pylori* isolates from human volunteers who participated in an experimental vaccination study, and who were challenged with a fully virulent *H. pylori* strain. Bacterial isolates were collected 12 weeks post infection (62 weeks in one particular case). We aimed to investigate the *in vivo* genome and methylome evolution of *H. pylori* during the initial phase of chronic infection. With that purpose, we planned to use a combination of NGS techniques, like SMRT® sequencing, and advanced bioinformatics, together with genetic and biochemical experimental approaches. Thus, we anticipated that the study of evolved genomes from a known *H. pylori* strain would allow us to study *de novo* modifications taking place during the adaptation process to a novel niche. Moreover, because the isolates belonged either to placebo or vaccination groups, we tried to understand specifically the effect caused by the vaccine in the genetic modifications and phenotypic changes during early infection.

Next, I intended investigate the role of methylation in gene transcription. Because every *H. pylori* strain carries a unique set of R-M systems leading to variable methylomes, there are very few methylated motifs shared between *H. pylori* strains. However, we observed that one motif ($G^{m5}CGC$) was present in all *H. pylori* strains whose methylomes are publically available. Some studies have attempted to understand how methylation within specific target motifs influences transcription in *H. pylori*. Nonetheless, the MTases studied were not universally present and active in all *H. pylori* strains, indicating that regulation by those enzymes was strain-specific.

Therefore, during this thesis I aimed to dissect the function of a highly conserved MTase (JHP1050), present in all *H. pylori* strains analyzed so far, that methylates the sequence GCGC resulting in a $G^{m5}CGC$ motif. The objective was to identify whether methylation of GCGC motifs had an impact on gene regulation. Further, I intended to understand whether the effects were comparable between *H. pylori* strains. To do so, we were planning to apply a combination of RNA sequencing and advanced bioinformatic tools to two different *H. pylori* strains and their mutants carrying an interrupted MTase gene. Deciphering the transcriptomes, we were aiming to observe differentially expressed genes between both, the wild type and mutant strain and between the two *H. pylori* strains analyzed. Differential gene expression caused by the absence of methylation can be direct and indirect. Hence, we wanted to address if there was a direct impact of methylated GCGC motifs on gene expression. For that specific aim, we created a set of mutants in which selected GCGC motifs located within or upstream of a specific gene. The GCGC sequences were modified to GAGC motifs not susceptible to methylation. Then, we planned to quantify the expression of the target gene.

Moreover, I aimed to characterize whether the absence of methylation had an effect on phenotypic traits and relate the phenotypic alterations to the differences observed in gene expression. With that purpose, we planned to characterize four *H. pylori* wild type and mutant strains with several phenotypic assays and to analyze whether adherence to host cells, natural competence for DNA uptake, bacterial cell shape, or susceptibility to copper were affected by the absence of methylation.

5. RESULTS

Contributions of the authors to the manuscripts

Manuscript I: S.N., I.E. and J.K. are joint first authors and contributed equally. The study was initiated by S.S., experiments and analyses were planned by S.N., I.E, J.K. and S.S.

S.N., J.K. and S.S. wrote the first version of the manuscript, while the rewriting of the revised version was jointly done by I.E. and S.S., with input from all coauthors. B.B., C.S., J.O., Y.S., performed PacBio SMRT® sequencing. S.N, J.K. and I.E. analyzed the genome and methylome data and performed the experiments.

D.Y.G. provided strain BCM-300. T.W. and P.M. performed the clinical vaccine trial from which the bacterial isolates were obtained.

Manuscript II: I.E. is the sole first author of this study. C.J. and S.S. shared senior authorship.

S.S. initiated the study. Experiments were planned by I.E., C.J. and S.S., and performed by I.E. and A.O. Data were analyzed by I.E., F.A., C.J. and S.S.

I.E., S.S. and C.J. wrote and revised the manuscript with input from all coauthors.

5.1. Manuscript I

Genome and methylome variation in *Helicobacter pylori* with a cag Pathogenicity island during early stages of human infection

Sandra Nell*, **Iratxe Estibariz***, Juliane Krebes*, Boyke Bunk, David Y. Graham, Jörg Overmann, Yi Song, Cathrin Spröer, Ines Yang, Thomas Wex, Jonas Korlach, Peter Malfertheiner, and Sebastian Suerbaum

**Authors declared shared first co-authorship*

Published in:

Gastroenterology 2018; 154(3):612-623 (doi: 10.1053/j.gastro.2017.10.014)

Short summary

In this article, we studied genome and methylome evolution of 12 *H. pylori* isolates obtained 12 weeks (62 weeks in one case) after the challenge of human volunteers with a fully virulent *H. pylori* strain during a vaccination trial. Whole genomes comparisons showed sequence modifications between the isolates, many of them affecting virulence factors and adhesins. Differences in the methylomes were due to changes in the activity of phase-variable MTases. The study provides evidence of rapid mutational and epigenetic adaptation of *H. pylori* during the first weeks of human infection.

Genome and Methylome Variation in *Helicobacter pylori* With a *cag* Pathogenicity Island During Early Stages of Human Infection

Sandra Nell,[1,2,*] **Iratxe Estibariz**,[1,2,3,*] **Juliane Krebes**,[1,2,*] Boyke Bunk,[2,4] David Y. Graham,[5] Jörg Overmann,[2,4] Yi Song,[6] Cathrin Spröer,[2,4] Ines Yang,[1,2] Thomas Wex,[7] Jonas Korlach,[6] Peter Malfertheiner,[7] and Sebastian Suerbaum[1,2,3,8]

[1]*Institute of Medical Microbiology and Hospital Epidemiology, Hannover Medical School, Hannover, Germany;* [2]*German Center for Infection Research (DZIF), Hannover-Braunschweig Site, Hannover, Germany;* [3]*Medical Microbiology and Hospital Epidemiology, Max von Pettenkofer Institute, Faculty of Medicine, LMU Munich, München, Germany;* [4]*Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany;* [5]*Baylor College of Medicine, Michael E. DeBakey VAMC, Houston, Texas;* [6]*Pacific Biosciences, Menlo Park, California;* [7]*Department of Gastroenterology, Hepatology, and Infectious Diseases, Otto-von-Guericke University, Magdeburg, Germany; and* [8]*National Reference Center for Helicobacter pylori, München, Germany*

BACKGROUND & AIMS: *Helicobacter pylori* is remarkable for its genetic variation; yet, little is known about its genetic changes during early stages of human infection, as the bacteria adapt to their new environment. We analyzed genome and methylome variations in a fully virulent strain of *H pylori* during experimental infection. **METHODS:** We performed a randomized Phase I/II, observer-blind, placebo-controlled study of 12 healthy, *H pylori*–negative adults in Germany from October 2008 through March 2010. The volunteers were given a prophylactic vaccine candidate (n = 7) or placebo (n = 5) and then challenged with *H pylori* strain BCM-300. Biopsy samples were collected and *H pylori* were isolated. Genomes of the challenge strain and 12 reisolates, obtained 12 weeks after (or in 1 case, 62 weeks after) infection were sequenced by single-molecule, real-time technology, which, in parallel, permitted determination of genome-wide methylation patterns for all strains. Functional effects of genetic changes observed in *H pylori* strains during human infection were assessed by measuring release of interleukin 8 from AGS cells (to detect *cag* pathogenicity island function), neutral red uptake (to detect vacuolating cytotoxin activity), and adhesion assays. **RESULTS:** The observed mutation rate was in agreement with rates previously determined from patients with chronic *H pylori* infections, without evidence of a mutation burst. A loss of *cag* pathogenicity island function was observed in 3 reisolates. In addition, 3 reisolates from the vaccine group acquired mutations in the vacuolating cytotoxin gene *vacA*, resulting in loss of vacuolization activity. We observed interstrain variation in methylomes due to phase variation in genes encoding methyltransferases. **CONCLUSIONS:** We analyzed adaptation of a fully virulent strain of *H pylori* to 12 different volunteers to obtain a robust estimate of the frequency of genetic and epigenetic changes in the absence of interstrain recombination. Our findings indicate that the large amount of genetic variation in *H pylori* poses a challenge to vaccine development. ClinicalTrials.gov no: NCT00736476.

Keywords: Microbe; Stomach; Pathogen; Cancer.

Helicobacter pylori is a highly prevalent bacterial pathogen that infects the human stomach. If untreated, *H pylori* can establish a lifelong infection that can either remain asymptomatic, or lead to severe sequelae, including peptic ulcer disease and gastric cancer.[1] The species *H pylori* is notable for its exceptionally high genetic diversity and variability. Elevated rates of spontaneous mutations are attributed to the lack of a number of classic DNA repair genes,[2–4] in combination with specific mutagenic properties of its DNA polymerase I.[5] In addition, recombination during mixed infections with multiple *H pylori* strains within one stomach was shown to be the dominant driving force of genetic variability.[6–8] The genetic variability of *H pylori* is thought to be important for its adaptation to different individual hosts, and to the constantly changing conditions of the gastric niche.[2] To date, in vivo genome evolution of *H pylori* has been mainly studied in isolates obtained from chronically infected individuals.[6,9–12] These studies showed that outer membrane protein (OMP) encoding genes were more frequently affected by genomic changes than other genes, pointing to a strong selection for the diversification of proteins that interact with the host during chronic infection.[10] Immune evasion is suggested to be one potential driving force for the diversification of *H pylori* in vivo.

H pylori therapy aims at the eradication of infection, yet the increasing spread of antibiotic resistance necessitates the development of alternative approaches for the control of *H pylori* infection. Therefore, since the early 1990s, multiple attempts have been made to develop a vaccine, but to date, no effective therapeutic or prophylactic vaccine is commercially available,[13–16] and the high genetic diversity and variability of *H pylori* may have contributed to this situation.

***Authors share co-first authorship.**

Abbreviations used in this paper: **ANOVA, analysis of variance; *cag*PAI, *cag* pathogenicity island; IL, interleukin; MTases, methyltransferases; OMP, outer membrane protein; PCR, polymerase chain reaction; R-M, Restriction-Modification; SMRT sequencing, single-molecule, real-time sequencing; SNPs, single-nucleotide polymorphisms; T4SS, type IV secretion system; UBT, urea breath test.**

Most current article

0016-5085/$36.00
https://doi.org/10.1053/j.gastro.2017.10.014

EDITOR'S NOTES

BACKGROUND AND CONTEXT

Helicobacter pylori bacteria have been shown to evolve and diversify in chronically infected patients. Little has been known about the genetic adaptation of *H pylori* in the early phase infection following the initial acquisition.

NEW FINDINGS

The researchers determined the genome sequences of *H pylori* before challenge and after several months of infection, permitting to precisely determine rates of mutations in early-stage infections, and to identify the bacterial genes affected by the mutations.

LIMITATIONS

Only one bacterial re-isolate clone was available for each volunteer. Thus, an assessment of the variability within the individual stomach was not possible. The number of volunteers in the vaccinated and non-vaccinated subgroups were relatively small.

IMPACT

The findings indicate that the large amount of genetic variation in *H pylori* poses a challenge to vaccine development.

The analysis of strains from chronically infected individuals, as performed previously,[6,8–10,12] does not permit investigation of the in vivo diversification of OMPs and other important virulence genes during early colonization, because infections might have been established for many years and initial changes are most likely masked by subsequent changes and purifying selection. Experimental infections of *H pylori*–negative individuals with a defined strain are better suited to specifically investigate early adaptation to the individual host. So far, only 2 studies have analyzed genomic changes of *H pylori* isolates obtained during the initial phase of an experimental human infection in 1 and 2 volunteers, respectively.[10,17] Although Kennemann and coworkers[10] found only marginal genomic changes and not a single recombination event, Linz and coworkers[17] reported rapid genome evolution through a mutation burst and numerous recombination events. The reason for the discrepancy between the 2 studies was unknown.

Methylation of DNA is an important form of epigenetic modification catalyzed by methyltransferases (MTases). *H pylori* is characterized by a striking abundance and substantial interstrain diversity of MTases and restriction-modification systems.[18–22] Recent studies have taken advantage of the Single-Molecule, Real-Time (SMRT) sequencing technology to characterize the genome-wide DNA methylation in multiple *H pylori* strains.[19,20,23] However, to date, no study has investigated methylation in the context of functional adaptation during human infection in vivo.

In this study, we applied SMRT sequencing to analyze both genome and methylome variation in 12 *H pylori* isolates obtained after experimental infection of human volunteers during a vaccine trial with *H pylori* challenge strain BCM-300. This *babA*-positive strain has the *vacA* s1m1 genotype, carries an intact *cag* pathogenicity island (*cag*PAI) and expresses the effector protein CagA. We identify individual sequence differences in all reisolates, many of which affect virulence and host interaction factors, such as *cagA* and *vacA*. Variations in the methylome were likewise detected, resulting from phase-variable expression of 2 MTase genes.

Materials and Methods

H pylori *Strains and Ethics Statement*

The experimental human infection study was a randomized Phase I/II, observer-blind, placebo-controlled, single-center study performed in healthy *H pylori*–negative adults from October 2008 to March 2010 (ClinicalTrials.gov: NCT00736476). The study was performed at the Clinic of Gastroenterology, Hepatology and Infectious Diseases at the Otto-von-Guericke University of Magdeburg, Germany; it followed all good clinical practice criteria and International Conference on Harmonization guidelines, and received the approval from the local ethical committee and written informed consent from all subjects.

The challenge strain *H pylori* BCM-300 that was used for experimental human infection[24] was originally isolated from an asymptomatic volunteer with mild superficial gastritis (ATCC BAA-1606). Gastric biopsies obtained from study participants were subjected to culture with single colony purification for isolation of *H pylori*. The histological typing and grading of gastritis was performed according to the recommendations of the updated Sydney classification.[25] Inflammation and all other parameters were semiquantitatively scored as either 0 (absent), 1 (mild), 2 (moderate), or 3 (severe).

Microbiological and Molecular Biology Techniques

Details of the culture conditions for *H pylori* isolates and mutants as well as *Escherichia coli* strains, DNA preparation, Sanger sequencing, quantitative polymerase chain reaction (PCR), neutral red uptake assay, IL8 induction, BabA expression, Le(b) binding, insertion mutagenesis in *H pylori*, overexpression of Hpy300XI in *E coli,* and restriction analyses are described in Supplementary Materials and Methods.

SMRT Sequencing and Base Modification Analysis

Total genomic DNA was extracted using QIAGEN Genomic-tip 100/G columns (QIAGEN, Hilden, Germany). BCM-300 and 10 reisolates were genome-sequenced at Pacific Biosciences (Menlo Park, CA). SMRTbell template library construction of 15-kb shotgun libraries was performed as previously described.[26] Genomes were sequenced on the Pacific Biosciences RSII instrument using 1 SMRT Cell per strain applying P4/C2 chemistry. Strains HE134/09 and HE178/09, *H pylori* mutant strains and *E coli* expression strains were sequenced at DSMZ (Braunschweig, Germany) as follows: SMRTbell template libraries were prepared according to the instructions from Pacific Biosciences following the Procedure & Checklist for 10 kb Template Preparation and Sequencing. Sequencing of 2 SMRT Cells per strain was performed using the Pacific Biosciences RSII instrument and P6/C4 chemistry. De novo genome assembly was carried out with HGAP2 (Pacific Biosciences) and HGAP3 (DSMZ)[27] and genome consensus using

Quiver.[27] Genome-wide detection of base modification and motif analysis was performed using the standard settings in the "RS_Modification_and_Motif_Analysis.1" protocol.

Genome Analyses

The genome of BCM-300 was annotated using Prokka version 1.7.[28] Whole-genome comparison was performed using Kodon (Applied Math, Austin, TX) with BCM-300 as reference. Sequence differences, including single-nucleotide polymorphisms (SNPs), indels, and intrachromosomal rearrangements, were identified (Supplementary Data 1). SNPs defined as base substitutions flanked by identical sequence of at least 200 bp[10] were validated by targeted Sanger sequencing (Table 1). The mutation rate of each reisolate was calculated by the following formula: no. of SNPs/genome length of BCM-300 [nt]/time of infection [d] * 365 d.

Identification and Assignment of MTase Genes

The annotation of the genomes revealed numerous putative Type I, II, and III restriction-modification (R-M) genes. In addition, the genome sequence of BCM-300 was scanned for the presence of R-M genes as previously described.[29] The specificities of most identified putative MTase genes were predicted by homology search using the REBASE BLAST tool.[21] These predicted specificities were subsequently matched with the motifs identified through base modification analyses from SMRT sequencing data. In addition, several putative MTase genes showed strong homology to not yet functionally characterized enzymes, only very weak homology to characterized enzymes, or no homology at all. These MTases represented strong candidates for the assignment of the novel recognition sites detected through SMRT sequencing.

Detection of Antibody Responses Against Vaccine Antigens

IgG antibody responses against the 3 vaccine antigens were determined using an enzyme-linked immunosorbent assay described previously.[15] The data presented in Supplementary Figure 1 are a subset of the full dataset reported in the full description of the clinical trial (Malfertheiner et al., submitted).

Statistical Analyses

GraphPad Prism 6 (La Jolla, CA) was used for statistical analyses and generation of graphs.

Data Availability

Genome sequence data of all strains have been deposited in the European Nucleotide Archive with the study number PRJEB17945. The authors declare that all other data supporting the findings of this study are available within the article and its Supplementary Information, or from the corresponding author on request.

Results

Prophylactic Vaccination, Challenge of Human Volunteers, and Recovery of H pylori Reisolates

In a clinical trial that will be published with full details in a separate article (Malfertheiner et al., submitted), 12 human volunteers who had been tested negative for *H pylori* infection by both urea breath test (UBT) and bacterial culture of gastric tissue biopsies were either administered a prophylactic tricomponent vaccine candidate (n = 7) that contained 3 recombinant *H pylori* proteins, the translocated effector CagA, the vaculating cytotoxin VacA, and the neutrophil-activating protein NAP, or placebo (n = 5), respectively. The volunteers were subsequently challenged with the *cag*PAI-positive *H pylori* strain BCM-300.[24] Twelve weeks, or, in 1 case, 62 weeks post challenge, gastric biopsies were taken from all study participants during gastroendoscopy, and *H pylori* cultures were performed (Figure 1). To enable follow-up investigations on reisolates, 1 single *H pylori* strain was isolated and archived for each study participant. Gastric biopsies obtained during initial screening and after challenge with *H pylori* were analyzed by histopathology (Figure 1). Although none of the participants showed any sign of inflammation before infection, various levels of gastric

Table 1. In Vivo Mutation Rates of *H pylori* Challenge Strain BCM-300 in 12 Infected Human Volunteers

Group	Strain	Gastroendoscopy, *wpc*	Chromosome Size, *bp*	No. of SNPs	Mutation rate, mutations site^{-1} year^{-1}
Challenge strain	BCM-300	-	1,667,883	-	
Placebo	HE136/09	12	1,667,978	2	5.2×10^{-6}
	HE141/09	12	1,670,384	2	5.2×10^{-6}
	HE143/09	12	1,667,804	2	5.2×10^{-6}
	HE147/09	12	1,667,712	4	1.0×10^{-5}
	HE170/09	12	1,670,321	1	2.6×10^{-6}
Vaccine	HE93/10	62	1,682,800[a]	4	2.0×10^{-6}
	HE101/09	12	1,667,821	1	2.6×10^{-6}
	HE132/09	12	1,668,174	3	7.8×10^{-6}
	HE134/09	12	1,667,858	2	5.2×10^{-6}
	HE142/09	12	1,667,936	3	7.8×10^{-6}
	HE171/09	12	1,667,625	1	2.6×10^{-6}
	HE178/09	12	1,667,894	2	5.2×10^{-6}

wpc, weeks post challenge.
[a]This chromosome is not closed due to a long repeat element at the end of the genome sequence.

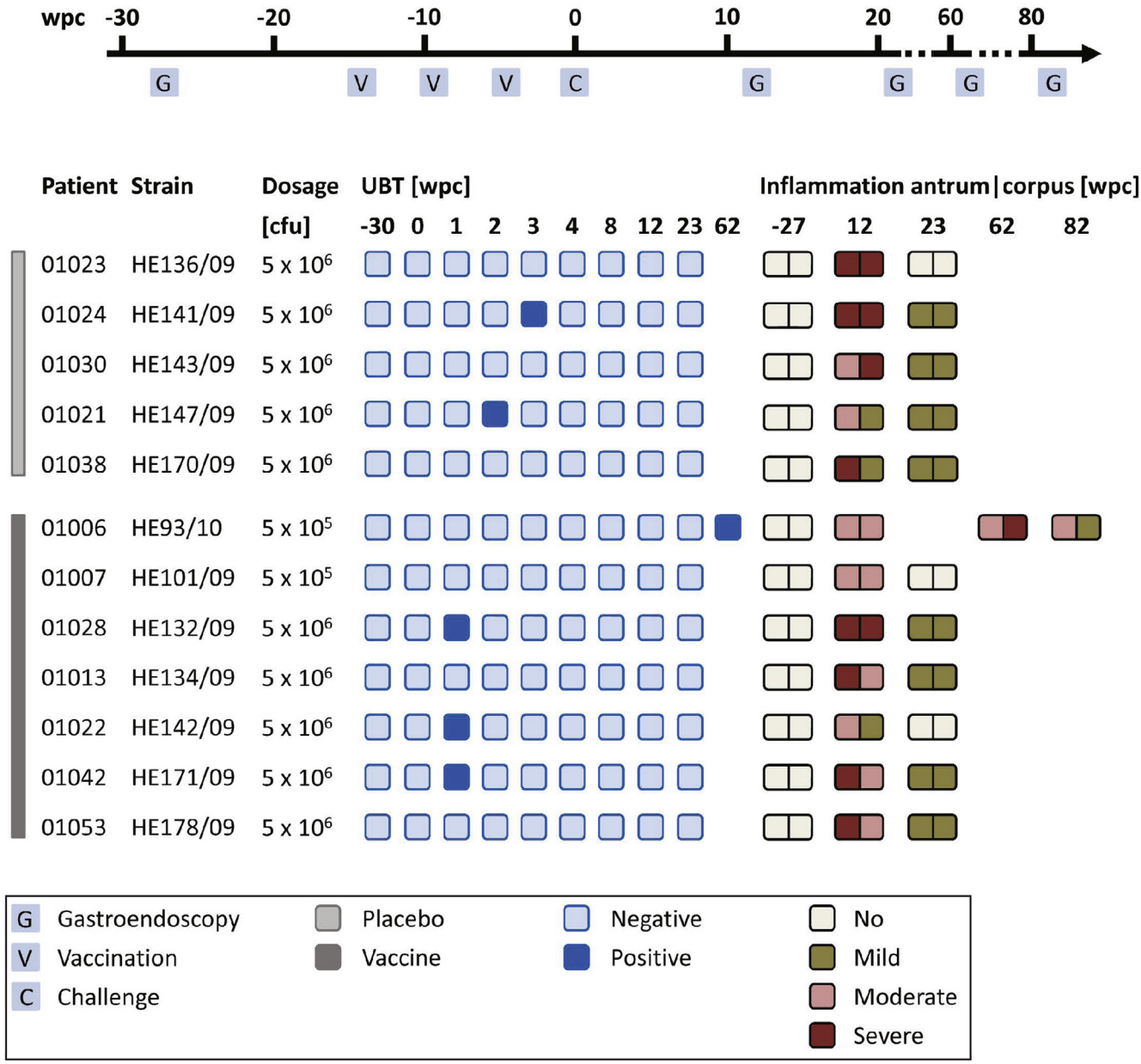


Figure 1. Schematic representation of study design and parameters. After a gastroendoscopy demonstrating *H pylori* negativity, study participants received 3 doses of vaccine or placebo and were subsequently challenged with *H pylori* strain BCM-300. Participants were monitored with regular UBTs and underwent gastroendoscopy for cultivation of *H pylori* (12 or 62 wpc) and histology. C, challenge; cfu, colony-forming units; G, gastroendoscopy; V, vaccination; wpc, weeks post challenge.

inflammation ranging from mild to severe were detected 12 weeks post challenge, without any significant difference between the vaccine and placebo groups. UBTs were performed regularly during the study to test for *H pylori* infection (Figure 1). All UBTs performed before challenge were negative, as expected. Most UBTs remained negative after the challenge, in some cases even at the time point when positive *H pylori* culture was achieved from biopsies.

Genome Evolution of BCM-300 in Different Human Volunteers

The genomes of challenge strain *H pylori* BCM-300 and 1 reisolate from each of the 12 volunteers were sequenced by SMRT technology[27,30] (Table 1). All but 1 reisolate genome could be assembled into a closed single chromosome. Annotation with Prokka[28] identified 1576 coding sequences in BCM-300. Genome analysis of the challenge strain showed the presence of a complete *cag*PAI, 5 copies of the insertion element IS*Hp609,*[31] the major adhesin *babA,* and the *vacA* s1m1 genotype. Although the genome sizes of most reisolates varied only slightly, 3 reisolates differed considerably. Two isolates (HE141/09, HE170/09) harbored an additional copy of IS*Hp609* (2398 bp) each located in the *cag*PAI. Strain HE93/10 harbored a long repeat element at both remaining contig ends, accounting for the difference in genome length of 14.9 kb. As

a result, we were not able to close this genome during the assembly process. The repeat element was located at the 3' end of the *cag*PAI and consisted of 4 repeats of approximately 4.2 kb containing the *cagA* gene (Figure 2*A*). Three *cagA* copies were identical to BCM-300, whereas 1 copy differed by six 1-bp deletions. The presence of multiple consecutive *cagA* copies in strain HE93/10 was confirmed by PCR amplification (Figure 2*B* and *C*). To determine the number of *cagA* gene copies in the genome sequence, the copy number of *cagA* was quantified by quantitative PCR (Figure 2*D*). The data indicated the presence of 7 *cagA* copies in the HE93/10 genome.

Whole-genome comparison of all reisolates with BCM-300 identified genomic changes including SNPs, insertions or deletions (indels), and intrachromosomal rearrangements (Supplementary Data 1). The reisolates harbored between 1 and 4 SNPs (Table 2), yielding an average mutation rate of 5.2 x 10^{-6} per year per site (range 0.2 x 10^{-6} to 1.0 x 10^{-5}) (Table 1). All SNPs were confirmed by Sanger sequencing. Twenty-three SNPs were located in open reading frames, resulting in 6 synonymous and 17 nonsynonymous mutations, whereas 4 SNPs were found in intergenic regions. Interestingly, some genes including the OMP-encoding gene *hopF* and the important virulence genes *vacA* and *cagA* were affected in multiple reisolates. Most predicted indels were located in homopolymeric or dinucleotide repeats. Selected insertions/deletions in homopolymeric sequences were confirmed by Sanger sequencing (eg, in R-M systems or virulence genes, see later in this article); however, the length of homopolymeric repeats is not reliably determined by SMRT sequencing, and we chose not to systematically resequence all predicted indels. Resequencing was performed for 51 randomly selected indels, confirming 31 length changes that had occurred during infection, and 20 indels could not be confirmed. No recombination event originating from an unrelated *H pylori* strain was identified, indicating the absence of a mixed infection in the volunteers.

Changes in cagPAI *Functionality After Short-Term Human Infection*

BCM-300 harbors the *cag*PAI that encodes a type IV secretion system (T4SS) associated with increased virulence.[32] To check for the functionality of the *cag*PAI in BCM-300 and the reisolates, we determined the ability of the strains to induce IL8 secretion in human gastric AGS cells. Compared with the mock-infected control, BCM-300 and most of the reisolates were able to induce IL8 secretion (Figure 3*A*); however, 3 reisolates (HE101/09, HE142/09, HE170/09) had completely lost the ability to induce IL8 secretion. These strains harbored differences in the *cag*PAI, which might explain the impaired IL8 induction phenotype (Supplementary Data 1). In strain HE170/09, the insertion of an additional copy of IS*Hp609* resulted in the disruption of *cagE*, whereas HE101/09 and HE142/09 had frameshift mutations in *cagY*. Two further reisolates (HE132/09, HE178/10) showed a small but significant reduction of their ability to induce IL8 secretion. Both strains had acquired nonsynonymous mutations in *cag*-PAI genes *cagA* or *cagW* that may account for these phenotypes (Table 2). The mean IL8 induction was not significantly different between strains isolated from vaccinated volunteers and those isolated from placebo-treated individuals (nested analysis of variance [ANOVA] test, $P > .1$).

Modulation of H pylori *Cytotoxic Activity During Infection*

Three of the 7 reisolates of the vaccine group (HE93/10, HE101/09, HE178/09) contained premature stop codons in the *vacA* gene, which were predicted to code for nonfunctional truncated proteins (Table 2). To test for a loss of VacA function, we analyzed the capacity of the *H pylori* strains to induce

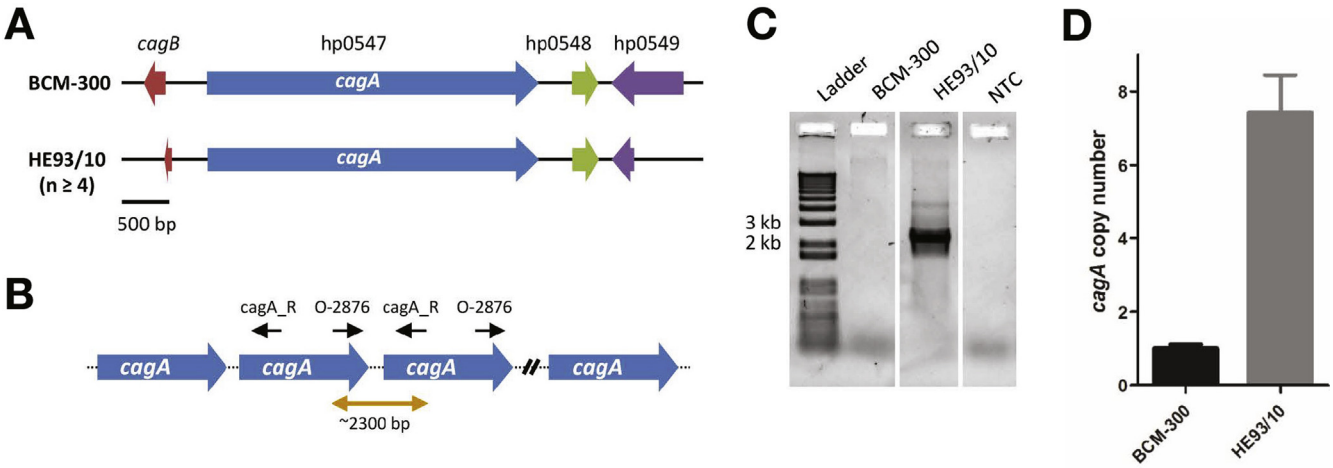


Figure 2. Multiplication of *cagA* in reisolate HE93/10. (*A*) Schematic representation of repeat region of HE93/10 compared with the BCM-300 locus. (*B*) Representation of multiple binding sites of primers cagA_R and O-2876 within the repeat region of HE93/10. (*C*) Confirmation of the presence of at least 2 consecutive *cagA* copies in HE93/10 by PCR using the primers cag_R and O-2876. NTC, no template control. (*D*) The genomic copy number of *cagA* was estimated by quantitative PCR and normalized to reference gene *efp*. The *cagA*/*efp* ratio of BCM-300 was set to 1. Data are presented as mean ± standard deviation of triplicate measurements.

Table 2. Mutations Identified in *H pylori* Reisolates From Volunteers Challenged With Strain BCM-300

Group	Strain ID	#	Position in BCM-300	Nt	SNP type	AA	Annotation	BCM-300 locus tag
Placebo	HE136/09	1	924452	G > A	Synonymous		Pantothenate kinase	BCM300_00904
		2	1029483	C > A	Nonsynonymous	K > N	Hypothetical protein	BCM300_00997
	HE141/09	1	1029281	G > A	Nonsynonymous	H > Y	Hypothetical protein	BCM300_00997
		2	1471570	C > T	Synonymous		ATP synthase subunit B	BCM300_01431
	HE143/09	1	259853	G > A	Nonsynonymous	G > R	OMP (*hopF*)	BCM300_00256
		2	867350	C > T	Nonsynonymous	V > M	Molybdenum cofactor biosynthesis protein C (*moaC*)	BCM300_00842
	HE147/09	1	242241	G > T	Nonsynonymous	A > E	Beta-lactamase HcpE	BCM300_00238
		2	581568	C > T	Nonsynonymous	S > L	cag pathogenicity island protein (*cagA*)	BCM300_00568
		3	934310	C > T	Synonymous		Hypothetical protein	BCM300_00916
		4	1113391	A > G	Intergenic			
	HE170/09	1	1340881	G > A	Intergenic			
Vaccine	HE93/10	1	258783	G > A	Nonsynonymous	G > D	OMP (*hopF*)	BCM300_00256
		2	301606	C > T	Synonymous		Toxin-like outer membrane protein	BCM300_00292
		3	857371	C > T	Nonsynonymous	D > N	Hypothetical protein	BCM300_00828
		4	949223	C > T	Nonsynonymous	Q > *	Vacuolating cytotoxin (*vacA*)	BCM300_00927
	HE101/09	1	948014	C > T	Nonsynonymous	Q > *	Vacuolating cytotoxin (*vacA*)	BCM300_00927
	HE132/09	1	581663	C > T	Nonsynonymous	R > C	cag pathogenicity island protein (*cagA*)	BCM300_00568
		2	697107	A > G	Synonymous		Fucosyltransferase	BCM300_00672
		3	875914	G > A	Intergenic			
	HE134/09	1	260916	T > C	Nonsynonymous	L > P	OMP (*hopG*)	BCM300_00257
		2	401200	G > A	Synonymous		Membrane protein	BCM300_00395
	HE142/09	1	762531	A > G	Intergenic			
		2	1488893	C > T	Nonsynonymous	Q > *	Sel1 repeat-containing protein	BCM300_01451
		3	1540693	C > T	Nonsynonymous	M > I	Zinc-metallo protease	BCM300_01503
	HE171/09	1	259244	G > A	Nonsynonymous	G > S	OMP (*hopF*)	BCM300_00256
	HE178/09	1	563321	C > T	Nonsynonymous	M > I	cag pathogenicity island protein (*cagW*)	BCM300_00551
		2	949904	C > T	Nonsynonymous	Q > *	Vacuolating cytotoxin (*vacA*)	BCM300_00927

[1]The genome annotation was automatically generated by Kodon based on annotation of *H pylori* strain 26695 (NC_000915.1), and then manually curated if necessary.
[2]The genome annotation of BCM-300 was generated by Prokka (v1.7).

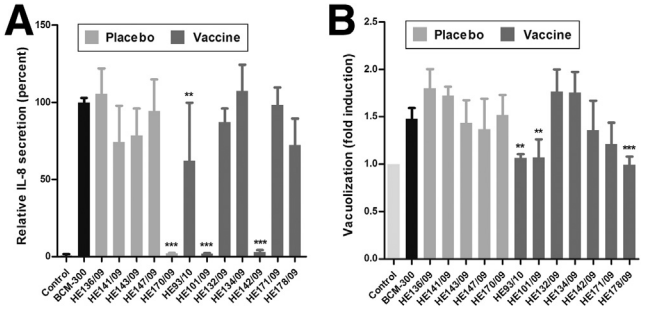


Figure 3. Induction of IL8 secretion and vacuolization by *H pylori* BCM-300 and reisolates. (*A*) AGS cells were infected with *H pylori* at MOI 50 for 4 hours. IL8 secretion into cell supernatants was measured by enzyme-linked immunosorbent assay. Data are presented as mean ± standard deviation of at least 4 independent experiments. BCM-300 was set to 100%. (*B*) MKN-28 cells were infected with *H pylori* at MOI 50 for 24 hours. Cell vacuolization was determined by uptake of neutral red. Vacuolization is shown relative to mock-infected control cells set to 1. Data are presented as mean ± standard deviation of at least 6 independent experiments. Statistical significance was analyzed by 1-way ANOVA ($P < .001$) followed by Bonferroni's multiple comparison test comparing BCM-300 with each of the reisolates (**$P < .01$; ***$P < .001$).

vacuoles in host cells by measuring the uptake of neutral red. BCM-300 induced vacuolization in MKN-28 cells by approximately 1.5-fold compared with mock-infected control cells (Figure 3*B*). In contrast, cells infected with the 3 reisolates harboring a putatively nonfunctional *vacA* gene behaved like the control. All other reisolates retained their ability to induce vacuolization. The mean vacuolating activity was slightly higher for reisolates from placebo-treated volunteers than for vaccinated volunteers (nested ANOVA test, $P = .0662$, which is suggestive, but not conclusive). We note that VacA activity was lost in *H pylori* isolates from 3 of the vaccinated vs none of the placebo-treated individuals. Because VacA was one component of the prophylactic vaccine, this might suggest that bacteria with inactivated VacA might have been selected for to evade a vaccine-induced immune response. However, the number of isolates was rather small, such that this difference in VacA inactivation between the 2 groups did not reach statistical significance.

Loss of Virulence Factors During Challenge Infection and Immune Responses Against Vaccine Components

We analyzed serum antibody responses against the 3 vaccine antigens, VacA, CagA, and NAP, at baseline and 12

weeks after challenge (ie, at the time point when the biopsies for the reisolate cultures were taken). For all 3 vaccine antigens, mean serum antibody titers were higher in the vaccinated subjects than in placebo-treated subjects, and the differences were statistically highly significant for all proteins, although 1 volunteer in the placebo group had a high antibody titer against CagA at baseline (Supplementary Figure 1).

We next compared antibody titers in individuals from whom the strains that had lost *cag*PAI or VacA activity had been isolated, vs the remaining individuals in whom the functions of VacA and/or *cag*PAI had stayed intact. Although the mean antibody titers were higher for those 3 individuals who lost VacA or *cag*PAI activity, respectively, vs the individuals whose strains kept the activity, the differences did not reach statistical significance (Supplementary Figure 1).

Variation and Partial Inactivation of OMP-Encoding Genes

Previous studies of genome evolution of *H pylori* during human infection showed that genes belonging to the *hop/hof/hor* family of paralogous *Helicobacter* OMP-encoding genes are frequently affected by genomic changes suggesting diversifying selection.[9,10,17,33] Four of the 12 reisolates contained SNPs in OMP-encoding genes, *hopF* or *hopG* (Table 2). Additional sequence differences were identified in other OMP genes (Supplementary Data 1), including differences in the length of the CT dinucleotide repeats of *sabA* and *sabB* in 3 reisolates each that led to gene inactivation. Furthermore, 2 strains had undergone intrachromosomal recombination affecting the putative adhesin domain of SabA (HE171/09) and SabB (HE147/09), respectively.

In contrast, none of the isolates displayed sequence differences in *babA*, the gene encoding the well-characterized adhesin of *H pylori* that mediates binding to Le(b).[34,35] BabA was expressed in BCM-300 and all reisolates (Supplementary Figure 2*A*), and we demonstrated binding to Le(b) for the challenge strain (Supplementary Figure 2*B*).

Methylome Variation During Early H pylori Infection

SMRT sequencing allowed us to analyze the genome-wide methylation patterns of challenge strain BCM-300 and all reisolates from human volunteers.[36] Between 1.8% and 2.3% of the genomic positions were detected as methylated, predominantly at adenosine nucleotides (87.5% ± 4.9%). Altogether, we identified 15 distinct motifs characteristic for Type II and Type III R-M systems (Supplementary Figure 3, Table 3). Most of the target sequences were almost fully methylated (> 90%); in 1 case of ^{m5}C methylation, which could not be reliably detected by SMRT sequencing,[37] the methylation of the corresponding motif was confirmed by restriction digestion analysis (Supplementary Figure 4).

Eleven of the 15 motifs could be assigned to MTases based on sequence homology searches using the REBASE database.[21] The specificity of 8 of these MTases in *H pylori* had already been experimentally verified in previous studies,[18,20,22,38] the MTases targeting 3 additional motifs could be predicted by homology and were functionally confirmed in this study (Supplementary Text 1, Supplementary Figures 5 and 6).

The MTases methylating the 4 remaining motifs could not be reliably assigned by homology with MTases of known specificity, indicating that BCM-300 contained yet uncharacterized *H pylori* MTases. Inactivation of selected candidate

Table 3. Methylated Sequence Motifs Detected by SMRT Sequencing in *H pylori* BCM-300 and Reisolates

No.	MTase specificity[a]	Assignment	BCM-300 locus tag	Type/subtype of R-M system	Presence of motifs
1	5′-C^{m6}**A**TG-3′	M.Hpy300I	BCM300_01150	II, alpha	All strains
2	5′-TGC ^{m6}**A**-3′	M.Hpy300II	BCM300_01310	II, gamma	All strains
3	5′-GA ^{m6}**A**TTC-3′	M.Hpy300III	BCM300_01060	II	All strains
4	5′-TCNNG ^{m6}**A**-3′	M.Hpy300IV	BCM300_01490	II, gamma	All strains
5	5′- ^{m4}**C**CGG-3′	M.Hpy300V	BCM300_00266	II, beta	All strains
6	5′-CC ^{m6}**A**TC-3′	M1.Hpy300VI	BCM300_01363	II, alpha	All strains
7	5′-G ^{m6}**A**TGG-3′	M2.Hpy300VI	BCM300_01364	II, alpha	All strains
8	5′-G ^{m6}**A**GG-3′	M1.Hpy300VII	BCM300_00054	II, beta	All strains
9	5′-AGG ^{m6}**A**G-3′[b]	Hpy300VIII	BCM300_01346	IIG	All strains
10	5′-G ^{m6}**A**TC-3′	M.Hpy300IX	BCM300_00102	II, beta	All strains
11	5′-G ^{m6}**A**CY-3′[b]	M.Hpy300X	BCM300_01342	III	Not present in HE93/10, HE147/09, HE171/09
12	5′-CCTYN ^{m6}**A**-3′[b]	Hpy300XI	BCM300_01297	IIG	Only present in HE143/09
13	5′-G ^{m5}**C**GC-3′[c]	M.Hpy300XII	BCM300_01446	II	All strains
14	5′-GTS ^{m6}**A**C-3′	M.Hpy300XIII	BCM300_00056	II, beta	All strains
15	5′-CAC ^{m6}**A**T-3′[b]	—	—	—	All strains

[a]The methylated position within the motif is highlighted in bold. Underlining indicates the modified base in the complementary strand of palindromic sites.
[b]Motifs are methylated by previously uncharacterized MTases.
[c]This motif was not reliably detected by SMRT sequencing, but methylation was experimentally confirmed by restriction digestion analysis.

genes by insertion mutagenesis and subsequent analysis of the methylation profiles of the isogenic mutants by SMRT sequencing permitted us to assign 3 novel MTases to 3 of the remaining motifs (Table 3, Supplementary Text 1).

We next focused on differences of methylation patterns between the BCM-300 challenge strain and reisolates from volunteers (Supplementary Data 2). Thirteen of the 15 methylated sequence motifs were detected as methylated in all strains. In contrast, the motif 5′-G^{m6}**A**CY-3′ lacked methylation in 3 of the reisolates (HE93/10, HE147/09, and HE171/09), and the motif 5′-CCTYN^{m6}**A**-3′ was unmethylated in BCM-300 and 11 reisolates, and only became methylated in reisolate HE143/09 (Table 3). Changes in *H pylori* methylation patterns have been shown to be frequently due to frameshifts in MTase genes. Two putative MTases, M.Hpy300X and Hpy300XI, displayed phase variation between the challenge strain and reisolates, suggesting that these might be responsible for the observed changes in methylation patterns. Sanger sequencing of the repeat sequences permitted to confirm this hypothesis in both cases (Supplementary Figure 7, Table 3). Hpy300XI shows substantial homology to the recently characterized systems HpyAXVI (*H pylori* 26695) and Hpy99XIV (*H pylori* J99-R3), respectively, which both were recently shown to undergo a unique frameshift-mediated switch of sequence specificity.[20] We were able to confirm that length variation of the second repeat in Hpy300XI (which was, however, not observed among the reisolates) had a similar impact on sequence specificity (Supplementary Text 1, Supplementary Figure 7*B*).

Discussion

Genome evolution of *H pylori* during the initial phase of human infection has so far only been analyzed by 2 recent studies that differed considerably in the number of observed genomic changes.[10,17] In the present study, we used SMRT sequencing to comprehensively characterize both genome and methylome evolution of the *cag*PAI-positive *H pylori* challenge strain BCM-300 in the early phase of experimental human infection. Additionally, the administration of a prophylactic vaccine to some of the volunteers enabled us to study genomic changes under vaccine-induced selective pressure.

The average mutation rate of 5.2 $\times$ 10^{-6} mutations per site per year calculated from this dataset is in good agreement with previous estimates observed during chronic human infection[6,9,10,12] that exceeded the mutation rates of many other bacterial species.[39] Studies on the genome evolution of *H pylori* during the initial phase of human infection are scarce, because natural acute *H pylori* infections are almost never diagnosed and they therefore require experimental infection of human volunteers with a defined *H pylori* strain. The only 2 previous studies of this kind were based on a very small number of isolates, and came to substantially different conclusions.[10,17] In the first study that was based on 1 reisolate obtained 12 weeks post volunteer infection with *cag*PAI-negative challenge strain BCS 100, we calculated a mutation rate of 2.6 x 10^{-6} mutations per site per year.[10] This is in good agreement with the results of the present study. These results are in marked contrast with the second study that analyzed 2 reisolates originating from 2 volunteers, who had been reinfected with their own strains after eradication of their natural *H pylori* infection.[17] The average mutation rate of 7.3 x 10^{-4} mutations per site per year was approximately 140-fold higher than our previous estimate. In addition to this presumed mutation burst, Linz et al[17] also observed a high recombination rate, which is strong evidence for the presence of a mixed infection. Although acquisition of an unrelated *H pylori* strain after reinfection and subsequent recombinational exchange between strains cannot be excluded, the substantial amount of recombination may also point to an ongoing low-level preestablished infection at the time of reinfection. Although eradication of the natural *H pylori* infection had been followed up by 2 consecutive negative UBTs, this observation does not unequivocally exclude the presence of a low-level *H pylori* infection. Note that in our study, *H pylori* was successfully cultured from gastric tissue biopsies of all volunteers, despite negative UBTs. Sampling of a strain from the preexisting natural infection instead of a descendant of the reinfecting strain would be a plausible alternative explanation for the high number of mutations and import events. Our earlier study[10] and the present study agree in that mutations occur at a similar rate during early and chronic infection. A similar low sensitivity of UBTs to detect experimental infections was previously reported in rhesus macaques.[40]

We observed changes in 2 major virulence factors of *H pylori*, the *cag*PAI and *vacA*, which were accompanied by functional consequences. Both virulence factors play a central role in the pathogenesis of *H pylori*.[41,42] The *cag*PAI-encoded T4SS mediates translocation of the oncoprotein CagA into gastric epithelial cells leading to a multitude of downstream effects on host cellular physiology[43–46] and is required for the induction of the proinflammatory cytokine IL8.[47] Three of 12 reisolates lost the function of this important virulence module during early-phase infection of volunteers. There was no significant difference between placebo and vaccine groups in their rates of *cag*PAI function loss. Loss of *cag*PAI function was due to frameshift mutations in *cagY* and insertion of a mobile element in *cagE*, respectively. Both genes were previously shown to be essential for IL8 induction in vitro.[48,49] A link between recombination in *cagY* and loss of T4SS function had recently been demonstrated in both rhesus macaques and mice.[50,51] Loss of cagPAI activity through inactivation of *cagY* had also been observed in 1 of 14 pairs of sequential isolates obtained from chronically infected human individuals.[51] Our data thus confirm and extend previous evidence that *cagY* plays an important role in the adaptation of *cag*PAI function to the individual host, probably facilitated by its length and high content of repeats. In addition, our previous analysis of the genetic diversity of individual *cag*PAI genes indicated that *cagY* is under diversifying selection.[52] Our observations indicate a selection against *cag*PAI function during acute human infection, at least in the genetic context of strain BCM-300, and in the absence of mixed infection. In addition to the complete loss of *cag*PAI

function, we observed also a slight but significant reduction in IL8 induction in 2 isolates that contained nonsynonymous SNPs in the *cag*PAI genes *cag*A and *cag*W, respectively. The functional relevance of these amino acid exchanges remains to be addressed. The same is true for the multiplication of *cagA* in strain HE93/10, which is in agreement with 2 recent studies that also reported dynamic changes of *cagA* copy numbers during infection.[53,54]

A further remarkable finding of this study was the inactivation of VacA that occurred exclusively in 3 reisolates cultured from volunteers who had received the candidate vaccine (which contained VacA). Although this difference between vaccine and placebo groups, as well as the difference of antibody titers between those individuals whose strains had lost VacA vs those who had not were both not statistically significant due to the small number of individuals in both groups, this may point to a specific selection in response to vaccine-induced selective pressure (ie, possibly to evade host immunity).

Infection with BCM-300 led to mild to severe levels of gastric inflammation in all individuals. The fact that we did not detect significant differences in inflammation levels for individuals harboring a strain with a nonfunctional *cag*PAI or VacA (and in 1 case, both) could be attributable to the presence of further proinflammatory bacterial traits, such as strength of bacterial adhesion and various bacterial MAMPs activating Toll-like receptors or NOD-like receptors/inflammasomes.[10,55–57] Moreover, because only 1 reisolate per individual was archived and available for genome characterization, we cannot exclude the possibility that bacteria with functional VacA or with an intact *cag*PAI were still present in the respective volunteers. Loss of *cag*PAI and VacA activity may have occurred after the challenge had induced strong inflammation that did not have sufficient time to resolve. Finally, vaccination itself may have an impact on the inflammatory response to the challenge in some individuals, potentially obscuring the effect of the deletion of virulence factors. Such inflammation-enhancing effects of vaccines have been described for animal models in the literature,[58] although a previous study using a challenge in humans did not observe an exacerbation of gastritis in vaccinated vs nonvaccinated subjects.[14]

Previous studies about the in vivo evolution of *H pylori* showed that OMP-encoding genes, in particular members of the *hop* family, were significantly affected by genomic changes.[10,11,17,59] It was proposed that surface-exposed proteins with a role in pathogen-host interactions undergo diversifying selection to facilitate host adaptation.[10] In this study, we observed a number of genomic changes in *hop* family genes, including SNPs, repeat length changes, and intrachromosomal rearrangements. The adhesin genes *sabA* and *sabB* were subject to both inactivating frameshift mutations and intrachromosomal rearrangements, whereas we detected no changes in *babA*, the gene encoding the Le(b) binding adhesin BabA.

Considering the overall low degree of genome variation detected in our study, the striking incidence of sequence differences in the *cag*PAI, *vac*A, and OMP-encoding genes strongly points to specific adaptation processes of the isolates to their individual host. Genetic diversification of these targets likely contributes to natural and vaccine-induced immune evasion and the modulation of adherence properties of *H pylori* to the changing conditions in the gastric environment.

To our knowledge, this is the first study to analyze in vivo methylome variation of *H pylori* during human infection. Investigation of the BCM-300 methylome identified dense genome-wide methylation at 15 distinct motifs, including 4 novel MTase recognition sequences. This is in agreement with several recent methylome analyses of other *H pylori* strains that also reported dense methylation, a similar number of methylated motifs, and substantial strain-specific differences.[19,20,23] The biological functions of the strain-specific large R-M system portfolios remain unknown. We recently showed that methylation-dependent restriction of heterologous DNA sequences limits the import of nonhomologous sequences into the *H pylori* genome.[60] *H pylori* MTases were also found to participate in various processes, such as cell cycle control, and DNA replication and repair.[61] Although genome comparisons did not reveal differences in the overall R-M gene content between BCM-300 and the reisolates, we noticed variations in the genome-wide methylation patterns. Although methylation of 13 motifs was shared between all strains, phase-variable expression of 2 MTase genes (M.Hpy300X, Hpy300XI) accounted for the observed interstrain variability. Phase variation is a common mechanism of host-adapted bacterial pathogens to generate genetic heterogeneity by high-frequency, reversible ON/OFF switching of gene expression via simple sequence repeats.[62] Phenotypic variation in *H pylori* was suggested to promote the generation of a heterogeneous population capable of rapid host adaptation.[2] Remarkably, random switching of the genome-wide methylation status through phase variation of MTase genes was recently shown to coordinate the expression of multiple genes in *H pylori* and other bacteria, thereby modulating their virulence.[63] These so-called phasevarions (phase-variable regulon) were proposed to affect host colonization and adaptation in vivo, for example by regulating gene expression via differential methylation of promoter sequences.[63] In our study, 3 reisolates lost expression of the Type III MTase M.Hpy300X and strain HE143/09 gained function of the Type IIG system Hpy300XI. Because of the random nature of phase variation, we cannot exclude that ON/OFF switching of MTase gene expression also occurred during isolation and cultivation of the reisolates. Nevertheless, this finding indicates that switching of the genome-wide methylation status during the initial human infection might facilitate host colonization, and should be further studied by functional approaches.

The biological significance of variable DNA methylation in *H pylori*, in particular its impact on functional adaptation in vivo, will be subject to further study.

Conclusion

The striking genetic variability of *H pylori* has been known for decades. Although it has been widely assumed

that it would pose a challenge to vaccine development, the current study is the first to clearly demonstrate the potential of *H pylori* to use this genetic variation to adapt to challenges during the adaptation to new hosts, and to vaccine-induced selection pressure by inactivating nonessential functions, including major virulence modules, and potentially also by modulation of its methylome.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at https://doi.org/10.1053/j.gastro.2017.10.014.

References

1. Suerbaum S, Michetti P. *Helicobacter pylori* infection. N Engl J Med 2002;347:1175–1186.
2. Suerbaum S, Josenhans C. *Helicobacter pylori* evolution and phenotypic diversification in a changing host. Nat Rev Microbiol 2007;5:441–452.
3. Dorer MS, Sessler TH, Salama NR. Recombination and DNA repair in *Helicobacter pylori*. Annu Rev Microbiol 2011;65:329–348.
4. Blaser MJ, Kirschner D. The equilibria that allow bacterial persistence in human hosts. Nature 2007;449:843–849.
5. **Garcia-Ortiz MV, Marsin S**, Arana ME, et al. Unexpected role for *Helicobacter pylori* DNA polymerase I as a source of genetic variability. PLoS Genet 2011;7:e1002152.
6. Didelot X, Nell S, Yang I, et al. Genomic evolution and transmission of *Helicobacter pylori* in two South African families. Proc Natl Acad Sci U S A 2013;110:13880–13885.
7. Suerbaum S, Maynard Smith J, Bapumia K, et al. Free recombination within *Helicobacter pylori*. Proc Natl Acad Sci U S A 1998;95:12619–12624.
8. Falush D, Kraft C, Taylor NS, et al. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. Proc Natl Acad Sci U S A 2001;98:15056–15061.
9. Furuta Y, Konno M, Osaki T, et al. Microevolution of virulence-related genes in *Helicobacter pylori* familial infection. PLoS One 2015;10:e0127197.
10. Kennemann L, Didelot X, Aebischer T, et al. *Helicobacter pylori* genome evolution during human infection. Proc Natl Acad Sci U S A 2011;108:5033–5038.
11. Krebes J, Didelot X, Kennemann L, et al. Bidirectional genomic exchange between *Helicobacter pylori* strains from a family in Coventry, United Kingdom. Int J Med Microbiol 2014;304:1135–1146.
12. Morelli G, Didelot X, Kusecek B, et al. Microevolution of *Helicobacter pylori* during prolonged infection of single hosts and within families. PLoS Genet 2010;6:e1001036.
13. Banerjee S, Medina-Fatimi A, Nichols R, et al. Safety and efficacy of low dose *Escherichia coli* enterotoxin adjuvant for urease based oral immunisation against *Helicobacter pylori* in healthy volunteers. Gut 2002;51:634–640.
14. **Aebischer T, Bumann D, Epple HJ**, et al. Correlation of T cell response and bacterial clearance in human volunteers challenged with *Helicobacter pylori* revealed by randomised controlled vaccination with Ty21a-based Salmonella vaccines. Gut 2008;57:1065–1072.
15. Malfertheiner P, Schultze V, Rosenkranz B, et al. Safety and immunogenicity of an intramuscular *Helicobacter pylori* vaccine in noninfected volunteers: a phase I study. Gastroenterology 2008;135:787–795.
16. **Zeng M, Mao XH**, Li JX, et al. Efficacy, safety, and immunogenicity of an oral recombinant *Helicobacter pylori* vaccine in children in China: a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 2015;386:1457–1464.
17. Linz B, Windsor HM, McGraw JJ, et al. A mutation burst during the acute phase of *Helicobacter pylori* infection in humans and rhesus macaques. Nat Commun 2014;5:4165.
18. Lin LF, Posfai J, Roberts RJ, et al. Comparative genomics of the restriction-modification systems in *Helicobacter pylori*. Proc Natl Acad Sci U S A 2001;98:2740–2745.
19. Furuta Y, Namba-Fukuyo H, Shibata TF, et al. Methylome diversification through changes in DNA methyltransferase sequence specificity. PLoS Genet 2014;10:e1004272.
20. Krebes J, Morgan RD, Bunk B, et al. The complex methylome of the human gastric pathogen *Helicobacter pylori*. Nucleic Acids Res 2014;42:2415–2432.
21. Roberts RJ, Vincze T, Posfai J, et al. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 2015;43:D298–D299.
22. Xu Q, Morgan RD, Roberts RJ, et al. Identification of Type II restriction and modification systems in *Helicobacter pylori* reveals their substantial diversity among strains. Proc Natl Acad Sci U S A 2000;97:9671–9676.
23. Lee WC, Anton BP, Wang S, et al. The complete methylome of *Helicobacter pylori* UM032. BMC Genomics 2015;16:424.
24. Malfertheiner P, Selgrad M, Wex T, et al. Efficacy of an investigational recombinant antigen based vaccine against a CagA *H pylori* infectious challenge in healthy volunteers. Gastroenterology 2012;142(Supplement 1):S-184.
25. Dixon MF, Genta RM, Yardley JH, et al. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. Am J Surg Pathol 1996;20:1161–1181.
26. **Travers KJ, Chin CS**, Rank DR, et al. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res 2010;38:e159.
27. Chin CS, Alexander DH, Marks P, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 2013;10:563–569.
28. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014;30:2068–2069.
29. Murray IA, Clark TA, Morgan RD, et al. The methylomes of six bacteria. Nucleic Acids Res 2012;40:11450–11462.

30. **Eid J, Fehr A, Gray J**, et al. Real-time DNA sequencing from single polymerase molecules. Science 2009; 323:133–138.
31. Kersulyte D, Kalia A, Zhang M, et al. Sequence organization and insertion specificity of the novel chimeric IS*Hp609* transposable element of *Helicobacter pylori*. J Bacteriol 2004;186:7521–7528.
32. Censini S, Lange C, Xiang Z, et al. *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. Proc Natl Acad Sci U S A 1996;93:14648–14653.
33. Kennemann L, Brenneke B, Andres S, et al. *In vivo* sequence variation in HopZ, a phase-variable outer membrane protein of *Helicobacter pylori*. Infect Immun 2012;80:4364–4373.
34. Borén T, Falk P, Roth KA, et al. Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens. Science 1993; 262:1892–1895.
35. **Ilver D, Arnqvist A**, Ogren J, et al. *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging. Science 1998;279: 373–377.
36. Flusberg BA, Webster DR, Lee JH, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods 2010;7:461–465.
37. Clark TA, Lu X, Luong K, et al. Enhanced 5-methylcytosine detection in Single-Molecule, Real-Time sequencing via Tet1 oxidation. BMC Biol 2013;11:4.
38. Kong H, Lin LF, Porter N, et al. Functional analysis of putative restriction-modification system genes in the *Helicobacter pylori* J99 genome. Nucleic Acids Res 2000; 28:3216–3223.
39. **Didelot X, Bowden R, Wilson DJ**, et al. Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet 2012;13:601–612.
40. Solnick JV, Hansen LM, Canfield DR. [14C]Urea breath test is not sensitive for detection of acute Helicobacter pylori infection in rhesus monkeys (*Macaca mulatta*). Dig Dis Sci 2002;47:298–303.
41. Atherton JC, Peek RM Jr, Tham KT, et al. Clinical and pathological importance of heterogeneity in *vacA*, the vacuolating cytotoxin gene of *Helicobacter pylori*. Gastroenterology 1997;112:92–99.
42. Figueiredo C, Machado JC, Pharoah P, et al. *Helicobacter pylori* and interleukin 1 genotyping: an opportunity to identify high-risk individuals for gastric carcinoma. J Natl Cancer Inst 2002;94: 1680–1687.
43. Odenbreit S, Puls J, Sedlmaier B, et al. Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. Science 2000;287:1497–1500.
44. Buti L, Spooner E, Van der Veen AG, et al. *Helicobacter pylori* cytotoxin-associated gene A (CagA) subverts the apoptosis-stimulating protein of p53 (ASPP2) tumor suppressor pathway of the host. Proc Natl Acad Sci U S A 2011;108:9238–9243.
45. Hatakeyama M. *Helicobacter pylori* CagA and gastric cancer: a paradigm for hit-and-run carcinogenesis. Cell Host Microbe 2014;15:306–316.
46. **Amieva MR, Vogelmann R**, Covacci A, et al. Disruption of the epithelial apical-junctional complex by *Helicobacter pylori* CagA. Science 2003;300:1430–1434.
47. Segal ED, Lange C, Covacci A, et al. Induction of host signal transduction pathways by *Helicobacter pylori*. Proc Natl Acad Sci U S A 1997;94:7595–7599.
48. Fischer W, Puls J, Buhrdorf R, et al. Systematic mutagenesis of the *Helicobacter pylori* cag pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. Mol Microbiol 2001; 42:1337–1348.
49. Selbach M, Moese S, Meyer TF, et al. Functional analysis of the *Helicobacter pylori cag* pathogenicity island reveals both VirD4-CagA-dependent and VirD4-CagA-independent mechanisms. Infect Immun 2002;70: 665–671.
50. Barrozo RM, Cooke CL, Hansen LM, et al. Functional plasticity in the type IV secretion system of *Helicobacter pylori*. PLoS Pathog 2013;9:e1003189.
51. Barrozo RM, Hansen LM, Lam AM, et al. CagY is an immune-sensitive regulator of the *Helicobacter pylori* type IV secretion system. Gastroenterology 2016; 151:1164–1175.
52. **Olbermann P, Josenhans C**, Moodley Y, et al. A global overview of the genetic and functional diversity in the *Helicobacter pylori cag* pathogenicity island. PLoS Genet 2010;6:e1001069.
53. Draper JL, Hansen LM, Bernick DL, et al. Fallacy of the unique genome: sequence diversity within single *Helicobacter pylori* strains. MBio 2017;8.
54. Jang S, Su H, Blum FC, et al. Dynamic expansion and contraction of *cagA* copy number in *Helicobacter pylori* impact development of gastric disease. MBio 2017;8(1).
55. Viala J, Chaput C, Boneca IG, et al. Nod1 responds to peptidoglycan delivered by the *Helicobacter pylori cag* pathogenicity island. Nat Immunol 2004;5:1166–1174.
56. Rad R, Ballhorn W, Voland P, et al. Extracellular and intracellular pattern recognition receptors cooperate in the recognition of *Helicobacter pylori*. Gastroenterology 2009;136:2247–2257.
57. Stein SC, Faber E, Bats SH, et al. *Helicobacter pylori* modulates host cell responses by CagT4SS-dependent translocation of an intermediate metabolite of LPS inner core heptose biosynthesis. PLoS Pathog 2017; 13:e1006514.
58. DeLyria ES, Nedrud JG, Ernst PB, et al. Vaccine-induced immunity against *Helicobacter pylori* in the absence of IL-17A. Helicobacter 2011;16:169–178.
59. Linz B, Windsor HM, Gajewski JP, et al. *Helicobacter pylori* genomic microevolution during naturally occurring transmission between adults. PLoS One 2013; 8:e82187.
60. **Bubendorfer S, Krebes J**, Yang I, et al. Genome-wide analysis of chromosomal import patterns after natural transformation of *Helicobacter pylori*. Nat Commun 2016; 7:11995.
61. Vasu K, Nagaraja V. Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol Mol Biol Rev 2013;77:53–72.

62. Bayliss CD. Determinants of phase variation rate and the fitness implications of differing rates for bacterial pathogens and commensals. FEMS Microbiol Rev 2009;33:504–520.
63. Srikhanta YN, Fox KL, Jennings MP. The phasevarion: phase variation of Type III DNA methyltransferases controls coordinated switching in multiple genes. Nat Rev Microbiol 2010;8:196–206.

Author names in bold designate shared co-first authorship.

Received April 13, 2017. Accepted October 2, 2017.

Reprint requests
Address requests for reprints to: Sebastian Suerbaum, Chair of Medical Microbiology and Hospital Epidemiology, Max von Pettenkofer Institute, Faculty of Medicine, LMU Munich, Pettenkoferstraße 9a, 80336 Munich, Germany. e-mail: suerbaum@mvp.uni-muenchen.de; fax: +49 89218072802.

Acknowledgments
We thank Christine Josenhans for helpful discussions, many valuable suggestions, and critical reading of the manuscript. We also thank Yu-Chih Tsai for help with re-running some analyses of SMRT sequencing data in preparation of the final manuscript, and Xavier Didelot for advice on nested ANOVA testing.

Ines Yang's present address is: Lower Saxony Centre for Biomedical Engineering, Implant Research and Development, Department of Prosthetic Dentistry and Biomedical Materials Science, Hannover Medical School, Stadtfelddamm 34, 30625 Hannover, Germany.

Conflicts of interest
These authors disclose the following: Yi Song and Jonas Korlach are full-time employees at Pacific Biosciences, a company commercializing SMRT sequencing technologies. The remaining authors disclose no conflicts.

Funding
This work was supported by grants DFG SFB 900/A1 from the German Research Foundation and grant HELDIVPAT in the framework of the ERA-NET PathoGenoMics to Sebastian Suerbaum (0315905A) and Peter Malfertheiner (0315905D) from the German Ministry of Education and Research.

Supplementary Materials and Methods

Culture Conditions

Strains were cultured from frozen stocks on blood agar plates with antibiotic supplements as previously described.[1] Mutant strains were grown on blood agar plates additionally supplemented with kanamycin (20 μg/mL), chloramphenicol (20 μg/mL), or both antibiotics as required. Liquid cultures were performed in brain heart infusion (BD Difco, Heidelberg, Germany) with yeast extract (2.5 g/L), 10% heat-inactivated horse serum, and a mix of antibiotics (vancomycin (10 mg/L), polymyxin B (3.2 mg/L), amphotericin B (4 mg/L), and trimethoprim (5 mg/L). Cultivation was performed at 175 rpm and 37°C in microaerobic atmosphere using air-tight jars (Oxoid, Wesel, Germany) and Anaerocult C gas-generating bags (Merck, Darmstadt, Germany).

Escherichia coli strains MC1061 and ER2683 (Supplementary Table 1) were grown under aerobic conditions at 37°C on LB agar plates or in LB broth (Lennox L Broth; Thermo Fisher Scientific, Darmstadt, Germany) supplemented with ampicillin (200 μg/mL), kanamycin (20 μg/mL), and/or chloramphenicol (20 μg/mL), as required.

DNA Preparation

Bacterial chromosomal DNA was extracted with the QIAamp DNA Minikit (QIAGEN, Hilden, Germany). Preparation of plasmid DNA was performed using QIAprep Spin Miniprep Kit (QIAGEN).

Sanger Sequencing

PCRs were performed according to standard protocols. Primers were designed using Primer3.[2] Amplification products were purified using the QIAquick PCR purification kit (QIAGEN). PCR amplicons were sequenced bidirectionally using the BigDye terminator v1.1 cycle sequencing kit and the 3130xl genetic analyzer (Thermo Fisher Scientific). Sequence data were analyzed using BioNumerics v6.01 (Applied Maths, Sint-Martens-Latem, Belgium).

Quantitative PCR

The genomic copy number of *cagA* and *efp* was determined by quantitative PCR using the standard curve method. Quantitative PCR was performed in the CFX96 Real-Time PCR Detection System (Bio-Rad, Munich, Germany) with the QuantiTect SYBR Green PCR Kit (QIAGEN) according to manufacturer's information. Primer sequences are shown in Supplementary Table 2.

Neutral Red Uptake Assay

Induction of vacuolization in MKN-28 cells by *H pylori* was analyzed by neutral red uptake assay.[3] Briefly, 1 x 10^4 cells/well were seeded into 96-well plates in RPMI 1640 with 10% fetal calf serum (FCS) (Biochrom, Berlin, Germany). After 24 hours, cells were infected with *H pylori* strains at MOI 50 in RPMI 1640 with 1% FCS and incubated for 24 hours. After removal of medium, cells were incubated with 100 μL/well neutral red medium (40 μg/mL) for at least 2 hours at 37°C. After washing, the incorporated dye was solubilized by adding 150 μL/well neutral red destain solution (50% ethanol [96%], 49% deionized water, 1% glacial acetic acid). Absorbance at 540/690 nm was measured using a microplate reader.

IL8 Induction

Briefly, 1.5 x 10^5 AGS cells/well were seeded into 24-well plates in RPMI 1640 with 10% FCS (Biochrom). After 24 hours, cells were infected with *H pylori* strains at MOI 50 in RPMI 1640 with 10% FCS and incubated for 4 hours. The IL8 concentration in the cell culture supernatants was determined with the BD OptEIA Human IL8 Enzyme-Linked Immunosorbent Assay Set (BD Biosciences, San Jose, CA).

BabA Expression

H pylori was harvested from 24-hour-old blood agar plates in phosphate-buffered saline (PBS) and centrifuged (5000*g*, 4°C, 10 minutes). Pellets were homogenized in Tris buffer pH 7.4 by sonication. Equal amounts of protein were separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (10%), and analyzed by Western blotting. BabA antibody was kindly provided by Thomas Borén. Peroxidase-labeled AffiniPure goat anti-rabbit IgG (H+L) (Jackson ImmunoResearch Laboratories, West Grove, PA) was used as secondary antibody. Western blots were developed with SuperSignal West Pico chemiluminescent substrate (Thermo Fisher Scientific, Waltham, MA) and chemiluminescence was detected using the LAS-3000 imaging system (Fujifilm Life Science, Tokyo, Japan).

Le(b) Binding

Le(b) binding of *H pylori* was determined by an enzyme-linked immunosorbent assay, as previously described.[4] Briefly, *H pylori* was harvested from 24-hour-old blood agar plates in PBS by centrifugation (2795*g*, 5 minutes, 4°C); 2 x 10^8 bacteria were biotinylated with 125 μg NHS-LC-biotin for 1 hour. A 96-well Universal Covalent microtiter plate (Corning Costar, Corning, NY) was coated with 250 ng bovine serum albumin (BSA), or 250 ng Le(b)-BSA. Afterward, the plate was exposed to UV light for 30 seconds in a Stratalinker (Stratagene, Penzberg, Germany). After blocking of the plate with 5% BSA in PBS, 1 x 10^7 biotinylated bacteria/well were coincubated for 1 hour. Adherent bacteria were fixed with 100 μL paraformaldehyde (2% in 100 mM potassium phosphate, pH 7). After washing with 0.05% Tween 20 in PBS, blocking with 10% FCS in PBS and repeated washing, the plate was incubated with neutravidin-horseradish peroxidase-conjugate in PBS plus 10% FCS. After final washing, the plate was incubated with 100 μL/well 3,3',5,5'-tetramethylbenzidine (BD Biosciences). The reaction was stopped by addition of 1 M H_3PO_4 (50

μL/well), and the extinction at 450 nm was measured using a microplate reader.

Insertion Mutagenesis in H pylori

Mutants in putative R-M genes were generated by natural transformation-mediated allelic exchange[5] of the corresponding target genes. Oligonucleotides used for mutagenesis are provided in Supplementary Table 2. Mutant alleles were constructed by overlap PCR. Briefly, 2 fragments of approximately 500 bp omitting a region within the central part of each of the target genes and the *aphA-3* and the CAT cassette were amplified via PCR. The resulting overlap fragments were ligated into the plasmid pUC19 via PstI and BamHI restriction sites and propagated in *E coli* MC1061. The constructed plasmids (Supplementary Table 3) were used for natural transformation of *H pylori* isolates BCM-300 and HE143/09. PCR amplification of the relevant loci confirmed the successful chromosomal replacement of the target gene with the respective mutant allele.

Overexpression of Hpy300XI in E coli

The Type IIG R-M system Hpy300XI (BCM300_01297) of BCM-300 was amplified via PCR and ligated to pRRS (a pUC19 derivative). Oligonucleotides are listed in Supplementary Table 2. Frameshift correction and stabilization by site-specific mutagenesis was performed as described previously.[6] Briefly, in both allele variants, the first repeat region was modified to a nonrepeat sequence of 11 nucleotides in length (CCT CCA CCG CC) to allow stable expression of Hpy300XI in ON status. Additionally, the second repeat in variant 1 was modified to stably comprise 14 nucleotides (CCG CCA CCT CCA CC), which allowed stable expression of the full-length protein containing the naturally occurring C-terminal additional target recognition domain. In variant 2, the second repeat region was modified to a nonrepeat sequence of 12 nucleotides (CCG CCA CCT CCA), resulting in a premature stop codon to prevent full-length translation. All sequence alterations were designed to preserve amino acid sequence. Sequence accuracy was confirmed by Sanger sequencing. The resulting constructs (Supplementary Table 3) were expressed in *E coli* ER2683 (dam^+ dcm^+) under the regulation of the P_{lac} promoter present on pRRS. Genomic DNA of the *E coli* host ER2683 expressing the modified Hpy300XI alleles was subjected to SMRT sequencing to analyze the resulting methylation profile.

Restriction Analysis

MTase activity for BCM-300 mutant strains disrupted in BCM300_01060 (Hpy300III, M.EcoRI homolog) and BCM300_01362-01364 (Hpy300VI, BccI homolog) was assayed by incubation of 300 ng genomic DNA with EcoRI and BccI, respectively, in 1x CutSmart Buffer (New England Biolabs, Ipswich, MA) for 1 hour at 37°C. *H pylori* 26695, which is naturally deficient for both MTase activities, and wild-type BCM-300 were included as controls. All DNA samples were also incubated without the addition of the respective REase as a negative control. Reactions were analyzed by agarose gel electrophoresis (1% wt/vol).

References

1. Moccia C, Krebes J, Kulick S, et al. The nucleotide excision repair (NER) system of *Helicobacter pylori*: role in mutation prevention and chromosomal import patterns after natural transformation. BMC Microbiol 2012; 12:67.
2. Untergasser A, Cutcutache I, Koressaar T, et al. Primer3—new capabilities and interfaces. Nucleic Acids Res 2012;40:e115.
3. Schneider S, Carra G, Sahin U, et al. Complex cellular responses of *Helicobacter pylori*-colonized gastric adenocarcinoma cells. Infect Immun 2011;79:2362–2371.
4. Nell S, Kennemann L, Schwarz S, et al. Dynamics of Lewis b binding and sequence variation of the *babA* adhesin gene during chronic *Helicobacter pylori* infection in humans. MBio 2014;5:e02281–14.
5. Haas R, Meyer TF, van Putten JP. Aflagellated mutants of *Helicobacter pylori* generated by genetic transformation of naturally competent strains using transposon shuttle mutagenesis. Mol Microbiol 1993;8:753–760.
6. Krebes J, Morgan RD, Bunk B, et al. The complex methylome of the human gastric pathogen *Helicobacter pylori*. Nucleic Acids Res 2014;42:2415–2432.

Supplementary Text 1: Assignment of Known MTase Specificities and Identification of Functionally Uncharacterized MTases

Based on sequence homology search using the REBASE database,[21] 11 of the 15 methylated sequence motifs were assigned to Type II MTases. Although the specificity of 8 of these MTases was already experimentally verified in previous studies,[18,20,22,38] the MTases targeting 5′-GA^{m6}**A**TTC-3′ and 5′-CC^{m6}**A**TC-3′ (methylated base in bold face, underlined base indicates methylated base on complementary DNA strand) had not been functionally characterized in *H pylori* so far (Table 3).

The EcoRI MTase homolog M.Hpy300III (BCM300_01060) was predicted to methylate 5′-GA^{m6}**A**TTC-3′. Methylation of this motif was recently detected in *H pylori* strains P12 and F30, and assigned by homology.[19] We inactivated BCM300_01060 by insertion mutagenesis and subsequently showed that the mutant DNA became susceptible to cleavage by EcoRI, confirming loss of 5′-GA^{m6}**A**TTC-3′ methylation (Supplementary Figure 5).

The R-M system Hpy300VI is homologous to the BccI system and was suggested to methylate 5′-CC^{m6}**A**TC-3′. Hpy300VI consists of a REase (BCM300_01362) and 2 MTases (BCM300_01363, BCM300_01364) each expected to methylate 1 of these motifs (Supplementary Figure 6*A*). To analyze the functionality and specificity of this system, we constructed a BCM-300 triple mutant disrupting all 3 loci as well as a REase/MTase 1 double mutant (BCM300_01362-01363). Restriction analysis showed that the genomic DNA of the triple mutant was completely digested after BccI treatment, confirming a lack of methylation of the BccI restriction site (Supplementary Figure 6*B*). In contrast, the DNA of the double mutant was not susceptible to cleavage, demonstrating that the second MTase encoded by BCM300_01364 is active. To elucidate which MTase is responsible for methylation of each of the 2 motifs, we performed SMRT sequencing of the REase/MTase 1 double mutant. A lack of 5′-CC^{m6}**A**TC-3′ methylation showed that M1.Hpy300VI (BCM300_01363) is responsible for methylation of this motif. Therefore, we conclude that M2.Hpy300VI catalyzes methylation of 5′-G^{m6}**A**TGG-3′.

Four methylated motifs could not be assigned to the responsible MTase by homology (indicated by superscript b in Table 3). Analysis of the BCM-300 genome identified a number of putative MTases. To investigate which MTases are responsible for methylation of these motifs, we inactivated selected candidates by insertion mutagenesis and analyzed the methylation profiles of the isogenic mutants by SMRT sequencing.

(**i**) 5′-AGG^{m6}**A**G-3′: Hpy300VIII (BCM300_01346). Inactivation of BCM300_01346 resulted in a complete loss of 5′-AGG^{m6}**A**G-3′ methylation. This ORF encodes a Type IIG R-M system, combining REase and MTase activity with a shared target recognition domain. Methylation of this motif was also detected for *H pylori* F30,[19] but the respective MTase was not identified. Yet, sequence comparisons showed that Hpy300VIII shares 79% amino acid identity with Hpy30ORF1390P (HPF30_RS07180) of strain F30. Most differences were located in the N-terminal REase domain while the C-terminus was highly conserved. This strongly indicates that Hpy30ORF1390P is responsible for 5′-AGG^{m6}**A**G-3′ methylation in *H pylori* F30.

(**ii**) 5′-G^{m6}**A**CY-3′: M.Hpy300X (BCM300_01342). Functional inactivation of BCM300_01342 led to abrogation of methylation of 5′-G^{m6}**A**CY-3′. Methylation of this motif has not been described before. Methylation of 5′-G^{m6}**A**CY-3′ was not found in 3 of the reisolates, namely HE93/10 and HE171/09 from the vaccine group and HE147/09 from the placebo group. BCM300_01342 contains 1 homopolymeric nucleotide repeat that is prone to phase variation by slipped strand mispairing. Analysis of the SMRT sequence data revealed differences in repeat length; however, these did not correspond to the observed phenotypes. Resequencing by Sanger technology revealed that BCM-300 and all reisolates showing 5′-G^{m6}**A**CY-3′ methylation had a repeat length of 13 G nucleotides allowing full-length translation of the MTase. In contrast, the 3 reisolates lacking the methylated motif had a repeat length of 12 and 14 G nucleotides, respectively, causing frameshift mutations that rendered the MTase nonfunctional (Supplementary Figure 7*A*). Of note, methylation of 5′-G^{m6}**A**CC-3′, which is 1 of the 2 motifs covered by the degenerate 5′-G^{m6}**A**CY-3′ site (Y = C/T), was also detected for *H pylori* P12,[19] but the corresponding MTase was not identified. Analysis of the P12 genome indicated sequence homology to the putative *mod* gene HPP12_1497 (M.Hpy1497P). This gene also contains an in-frame G-repeat, and shows 76% amino acid identity with M.Hpy300X. Most differences are located in the region containing the putative target recognition domain, which might either account for the slight difference in the recognition sequence or point to a completely different specificity.

(**iii**) 5′-CCTYN^{m6}**A**-3′: Hpy300XI (BCM300_01297). Methylation of 5′-CCTYN^{m6}**A**-3′ was solely detected for isolate HE143/09. As the subset of R-M systems did not differ between BCM-300 and its reisolates, we assumed the presence of a further active phase-variable MTase. BCM300_01297, which encodes a putative Type IIG R-M system, contains 2 homopolymeric nucleotide repeats (Supplementary Figure 7*B*). Analysis of the genome sequences did not reveal any correlation between repeat lengths and methylation status. Thus, resequencing of both repeat regions was performed and showed that only HE143/09 had both repeats in-frame allowing expression of the full-length protein. In all other isolates, this R-M system was inactivated by frameshift mutations in the first repeat region resulting in a premature stop codon. In line with these results, functional inactivation of BCM300_01297 in HE143/09 resulted in a lack of 5′-CCTYN^{m6}**A**-3′ methylation, confirming the specificity of this novel system (Table 3).

Interestingly, this system shows 83% and 81% aa identity to the recently characterized systems HpyAXVI

(*H pylori* 26695) and Hpy99XIV (*H pylori* J99-R3), respectively, for which a unique frameshift-mediated sequence specificity switch was discovered.[20] Both systems each contain 2 homopolymeric repeat regions, and although the first region was shown to mediate a reversible ON/OFF switching of enzyme activity, the length of the second region determines sequence specificity. Thus, we were interested to investigate whether length variation of the second repeat in Hpy300XI also has an impact on sequence specificity. For this, 2 different variants of the wild-type allele of BCM-300 were constructed for recombinant expression in *Escherichia coli* (Figure 7*B*). In both variants, the first repeat was changed to be in-frame to putatively activate the enzyme. The second repeat, which is in-frame in the wild-type allele, was additionally mutated in variant 2 to prevent full-length translation of the protein. SMRT sequencing of the *E coli* host DNAs showed that both alleles encode active MTases, yet revealed slightly different recognition sequences. Expression of variant 1 resulted in methylation of 5′-CCTTN^{m6}**A**-3′ (87% detected), whereas variant 2 showed methylation of 5′-CCTNC^{m6}**A**-3′ (98%) and 5′-CCTTT^{m6}**A**-3′ (74%) (Figure 7*B*). Methylation of 5′-CCTTN^{m6}**A**-3′ by the full-length enzyme expressed in *E coli* differed somewhat from the 5′-CCTYN^{m6}**A**-3′ methylation (Y = C/T) detected for *H pylori* strain HE143/09. This finding might be due to the overexpression in *E coli*, as a similar phenomenon had also been observed for the homologous system Hpy99XIV of *H pylori* J99-R3.[20] Nevertheless, as reported for HpyAXVI and Hpy99XIV, alteration of protein length mediated by the second repeat region affects the specificity of the positions −1 and −2 relative to the modified A. Thus, Hpy300XI represents a further example of a remarkable Type IIG R-M system showing a frameshift-mediated sequence specificity switch.

(iv) 5′-CAC^{m6}**A**T-3′: Inactivation of Type II MTase candidate genes encoded by BCM300_00688-00689, BCM300_01392, and BCM300_01574 did not reveal a change in the methylation profile of the mutant strains, indicating that all enzymes are not active in BCM-300. Thus, methylation of the novel motif 5′-CAC^{m6}**A**T-3′ could not be assigned to one of the analyzed putative MTase genes.

Supplementary Table 1. Bacterial Strains

Strain	Genotype	Source
Escherichia coli		
ER2683	*fhuA2 glnV44 e14- rfbD1? relA1? endA1 spoT1? thi-1 Δ (mcrC-mrr) 114::IS10 Δ (lacI-lacA) 200/F′proAB lacI^q ΔlacZM15 zzf::miniTn10 (KanR)*	Reference 1
MC1061	*ara*D139, Δ(*ara*, *leu*)7697, Δ*lac*X74, *galU*$^-$, *galK*$^-$, *hsr*$^-$, *hsm*$^-$, *strA*	Reference 2
Helicobacter pylori		
BCM-300	*H pylori* challenge strain (ATCC® BAA-1606™), CagA protein, *babA2*, and *cag* pathogenicity island positive, *vacA* s1-m1 type, *oipA* functional	D. Y. Graham
BCM-300 *BCM300_01060*	BCM300_01060 (predicted M.EcoRI homolog) from BCM-300 inactivated with *aphA-3*	This study
BCM-300 *BCM300_01362-01363*	BCM300_01363 (predicted M1.BccI homolog) and BCM300_01362 (predicted R.BccI homolog) from BCM-300 jointly inactivated with *aphA-3*	This study
BCM-300 *BCM300_01362-01364*	BCM-300 *BCM300_01362-01363* derivative containing a BCM300_01362::CAT disruption (predicted M2-BccI homolog)	This study
BCM-300 *BCM300_01574*	BCM300_01574 (predicted Type II R-M modification protein) from BCM-300 inactivated with aphA-3	This study
HE143/09 *BCM300_01297*	BCM300_01297 (predicted Type IIG R-M modification system) from HE143/09 inactivated with aphA-3	This study
BCM-300 *BCM300_01342*	BCM300_01342 (predicted Type III R-M modification protein) from BCM-300 inactivated with *aphA-3*	This study
BCM-300 *BCM300_01346*	BCM300_01346 (predicted Type IIG R-M modification system) from BCM-300 inactivated with *aphA-3*	This study
BCM-300 *BCM300_00688-00689*	BCM300_00688 and BCM300_00689 (predicted Type IIG R-M modification system) from BCM-300 inactivated with *aphA-3*	This study
BCM-300 *BCM300_01392*	BCM300_01392 (predicted Type IIG R-M specificity protein) from BCM-300 inactivated with *aphA-3*	This study

Supplementary Table 2. Oligonucleotide Primers

Primer	Target gene	5' → 3' sequence	RS	Application
cagA analysis				
cagA_R	*cagA*	GGGTTGTATGATATTTTCCATAA	-	Detection of multiple *cagA* copies
O-2876		GTGCCTRCTAGTTTGTCAGCGA	-	
cagA_F2	*cagA*	GCGATCAAAAATCCTACCAAA	-	Quantification of *cagA* by qPCR
cagA_R		GGGTTGTATGATATTTTCCATAA	-	
HPefpF01	*efp*	GGCAATTGGGATGAGCGAGCTC	-	Quantification of *efp* by qPCR
HPefp_R2		CGCACTTATCCCCCGCATGGA	-	
Methylome analysis				
Km8_OL1_for	*aphA3*	gcactgtccgatccgAGCGAACCATTTGAGGTG	-	Amplification of *aphA3* cassette with random overhangs
Km9_OL2_rev		gcaaggctaccgagcATCATCGATAAGCTTTTTAGAC	-	
pCAT-uniOL-fwd	*cat*	gcactgtccgatccgAGAGTCAACCGTGATATAGATTGAAA		Amplification of *cat* cassette with random overhangs
pCAT-uniOL-rev		gcaaggctaccgagcGACAGAGAGTATAGAAGTGCGC		
BCM300_gene1_PstI_for	BCM300_01060	atactgcagTTGGTAAAAACAATGTCAAGAAAAG	PstI	Insertion mutagenesis
BCM300_gene1_OL1_rev		cggatcggacagtgcCTTTTAGCAATTGAATGCTCTCC	-	
BCM300_gene1_OL2_for		gctcggtagccttgcCGCAATTGATCGAACACAGC	-	
BCM300_gene1_BamHI_rev		ataggatccCTAAAATAGGGGCATTTACC	BamHI	
BCM300_gene2_PstI_for	BCM300_01362_01363	atactgcagTGAAAAAAAGTATTCGTTCTCC	PstI	Insertion mutagenesis
BCM300_gene2_OL1_rev		cggatcggacagtgcAAATCTACATTACCCACAGG	-	
BCM300_gene2_REase_OL2_for		gctcggtagccttgcCCATAAGGCTTTTAAAGGGG	-	
BCM300_gene2_REase_BamHI_rev		ataggatccCATGGATATTTGGAGCCAC	BamHI	
BCM300_gene2_MT2_PstI_for	BCM300_01364	atactgcagGATTTAATCTAAAAAACCGCC	PstI	Insertion mutagenesis
BCM300_gene2_MT2_OL1_rev		cggatcggacagtgcACCTTAAAATCAGCGTCTTG	-	
BCM300_gene2_MT2_OL2_for		cggatcggacagtgcGAGAGAAAAGACGCTAACGA	-	
BCM300_gene2_MT2_BamHI_rev		ataggatccCTATGAGCCTTTTCTTTAACGC	BamHI	
BCM300_gene3_PstI_for	BCM300_01574	atactgcagGTGTTTTATCATAGCAGCGC	PstI	Insertion mutagenesis
BCM300_gene3_OL1_rev		cggatcggacagtgcCACGCCGATAAATTCTATAG	-	
BCM300_gene3_OL2_for		gctcggtagccttgcGTTTGCGTGAATATAGACGG	-	
BCM300_gene3_BamHI_rev		ataggatccCAATGCCTAGAGTCATTTAC	BamHI	
BCM300_gene4_PstI_for	BCM300_01297	atactgcagATGCTAAAAGAATATTTAGAAGGC	PstI	Insertion mutagenesis
BCM300_gene4_OL1_rev		cggatcggacagtgcCTTAATGCGTTTGCGAAATC	-	
BCM300_gene4_OL2_for		gctcggtagccttgcCACCATAGGCGAATCTTAC	-	
BCM300_gene4_BamHI_rev		ataggatccTTAGGTAATACAGAATAAAGGGAAG	BamHI	
BCM300_gene4_pRRS_PstI_for	BCM300_01297	tgcctgcagttaaggtttaacatATGCTAAAAGAATATTTAGAAGGC	PstI	Expression and frameshift correction
BCM300_gene4_fix_fs1_for		CCGCCAACAAACCCTAAAACACC	-	
BCM300_gene4_fix_fs1_rev		TGGAGGTAACATGTGCTTGAATAC	-	
BCM300_gene4_fix_fs2_for		CCTCCACCATTTAATACCAATATCGC	-	
BCM300_g4_fix_fs2_for2		CCTCCAATTTAATACCAATATCGCC	-	
BCM300_gene4_fix_fs2_rev		TGGCGGTGGATTTTGCAAGATTTC	-	
BCM300_gene5_PstI_for	BCM300_01342	atactgcagGCAAAATAAAGAAATGGATC	PstI	Insertion mutagenesis
BCM300_gene5_OL1_rev		cggatcggacagtgcCAATCTGGGATACATGAAAC	-	
BCM300_gene5_OL2_for		gctcggtagccttgcGCACGCATGATTTTTGTTTG	-	
BCM300_gene5_BamHI_rev		ataggatccCTACCCCCTAATCTTTAAATC	BamHI	

Supplementary Table 2. Continued

Primer	Target gene	5' → 3' sequence	RS	Application
BCM300_gene6_PstI_for	BCM300_01346	atactgcagCCCAACACAAACTACCCAAG	PstI	Insertion mutagenesis
BCM300_gene6_OL1_rev		cggatcggacagtgcCCCTTTCTATCGTAGCAATC	-	
BCM300_gene6_OL2_for		gctcggtagccttgcGATAATCAGATCACTCAAC	-	
BCM300_gene6_BamHI_rev		ataggatccCCCGTCTTCAATGGTTTTG	BamHI	
BCM300_g10_RM_PstI_for	BCM300_00688	atactgcagATGCAAGAAATCAGCGCCTAC	PstI	Insertion mutagenesis
BCM300_g10_RM_OL1_rev		cggatcggacagtgcGGGCGCGGCCTTTTCTTAT	-	
BCM300_g10_RM_OL2_for		gctcggtagccttgcATCGCTCGGTTGCTTTCCAA	-	
BCM300_g10_RM_BamHI_rev		ataggatccTAACAACATCGCTCGCCTGC	BamHI	
BCM300_g10_S_PstI_for	BCM300_00689	atactgcagGTGATAGGCTTTGTGGTGA	PstI	Insertion mutagenesis
BCM300_g10_S_OL1_rev		cggatcggacagtgcACCCCATTAGAATTGAGATC	-	
BCM300_g10_S_OL2_for		gctcggtagccttgcTCCACCATAAAGGCTATTTG	-	
BCM300_g10_S_BamHI_rev		ataggatccTCAAACCTCTTCTCGCTGAT	BamHI	
BCM300_g11_S_EcoRI_for	BCM300_01392	atagaattcTGATTGGCCCCCTTAGTAG	EcoRI	Insertion mutagenesis
BCM300_g11_S_OL1_rev		cggatcggacagtgcCTCAAGTTCGGCTATGAAT	-	
BCM300_g11_S_OL2_for		gctcggtagccttgcATCAGATCACTTATAGCGATA	-	
BCM300_g11_S_BamHI_rev		ataggatccTTAAAATAACGAGTCTTTTTGAAC	BamHI	

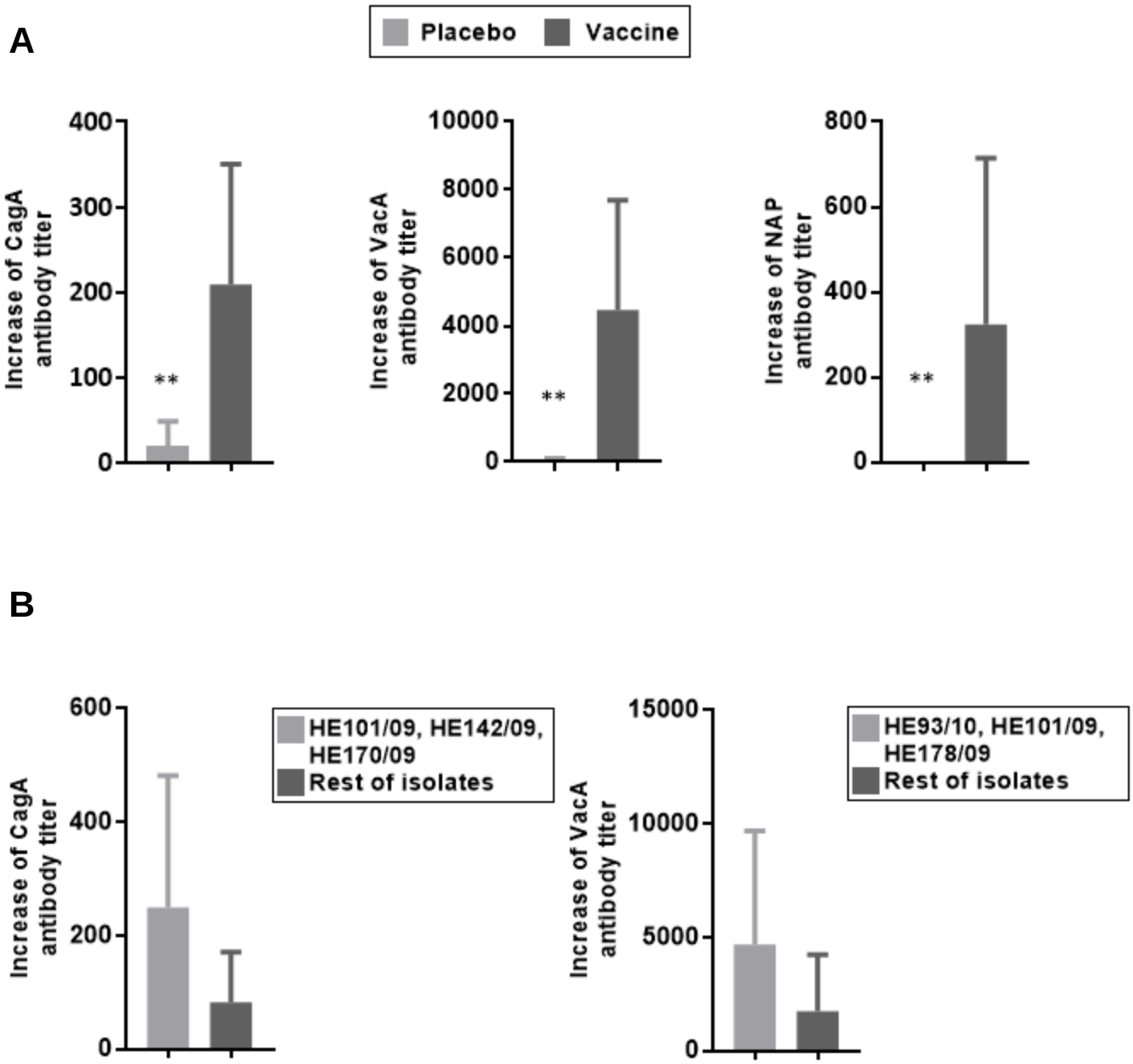
qPCR, quantitative PCR; RS, restriction site.

Supplementary Table 3. Plasmids

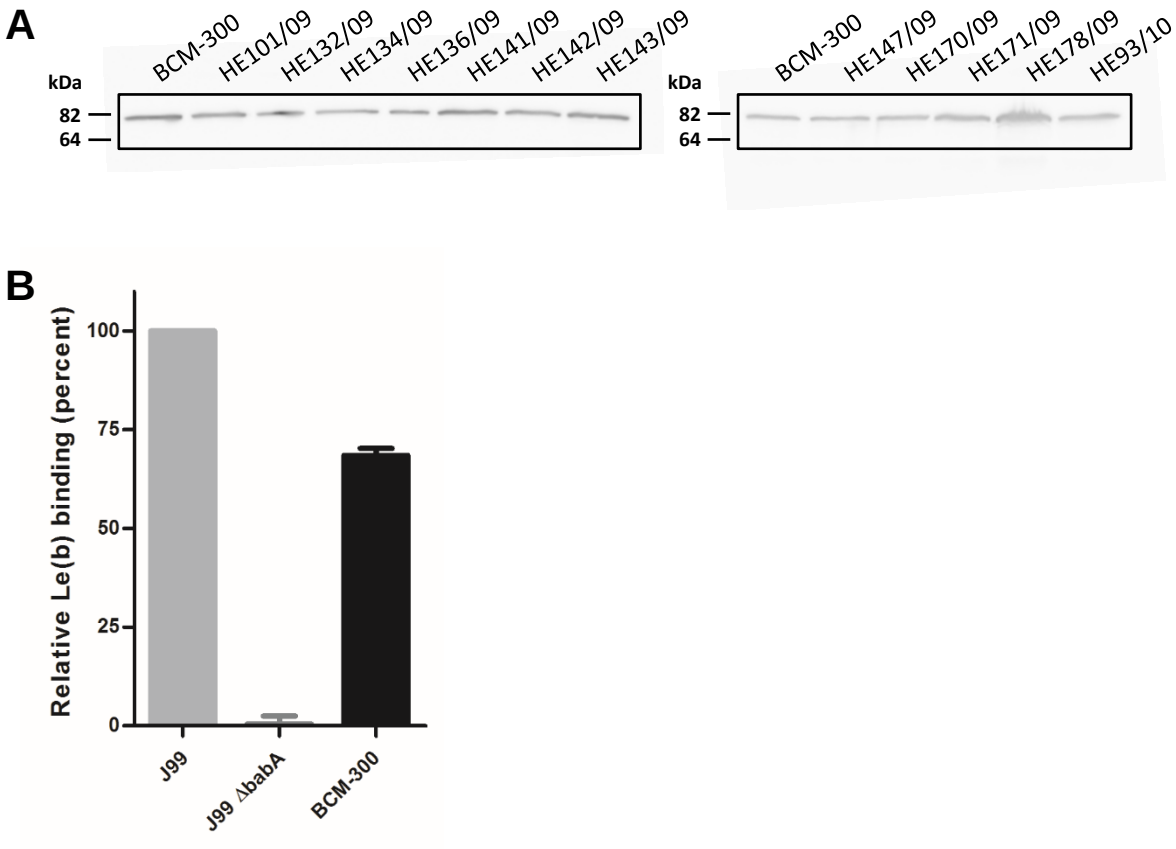
Plasmids	Genotype	Source
pBHpC8	Cm^r, Source of the *cat* cassette	Reference 3
pILL600	Amp^r, Km^r, Source of the *aphA-3* cassette	Reference 4
pRRS	pUC19 derivative, retroregulator structure from *B. thuringiensis* inserted into the BglII linker of pBBO (pUC19 modified by insertion of a BglII linker at the SacI site)	Reference 5
pSUS3133	Amp^r, Km^r, pUC19 derivative with a BCM300_01060::*ahpA3* disruption	This study
pSUS3135	Amp^r, Km^r, pUC19 derivative with a BCM300_01574::*ahpA3* disruption	This study
pSUS3136	Amp^r, Km^r, pUC19 derivative with a BCM300_01342::*ahpA3* disruption	This study
pSUS3138	Amp^r, Km^r, pUC19 derivative with a BCM300_01297::*ahpA3* disruption	This study
pSUS3139	Amp^r, pRRS derivative containing BCM300_01297	This study
pSUS3140	Amp^r, Km^r, pUC19 derivative with a BCM300_01346::*ahpA3* disruption	This study
pSUS3141	Amp^r, Km^r, pUC19 derivative with BCM300_01362-01363::*ahpA3* disruption (gene 2RM)	This study
pSUS3142	Amp^r, pSUS3139 derivative with frameshift 1 corrected	This study
pSUS3144	Amp^r, Cm^r, pUC19 derivative with a BCM300_01364::*cat* disruption	This study
pSUS3145	Amp^r, pSUS3142 derivative with both frameshifts corrected	This study
pSUS3146	Amp^r, pRRS derivative containing BCM300_01574)	This study
pSUS3151	Amp^r, pSUS3142 derivative with frameshift 1 corrected and frameshift 2 mutation stabilised	This study
pSUS3158	Amp^r, Cm^r, pUC19 derivative with a BCM300_00688::*cat* disruption	This study
pSUS3159	Amp^r, Km^r, pUC19 derivative with a BCM300_00689::*ahpA3* disruption	This study
pSUS3160	Amp^r, Km^r, pUC19 derivative with a BCM300_01392::*ahpA3* disruption	This study
pUC19	Amp^r, Colx101, MCS within *lacZ*: blue/white selection	Reference 6

References

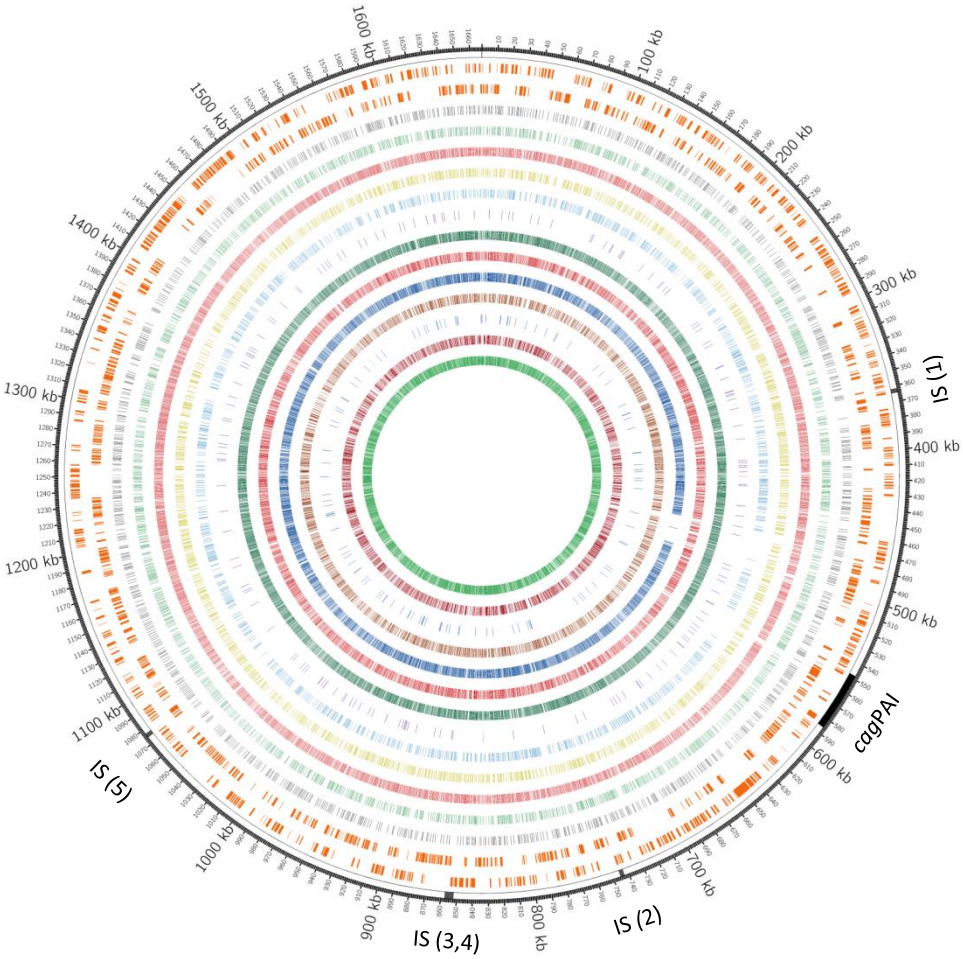
1. Sibley MH, Raleigh EA. Cassette-like variation of restriction enzyme genes in *Escherichia coli* C and relatives. Nucleic Acids Res 2004;32:522–534.
2. Casadaban MJ, Cohen SN. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J Mol Biol 1980;138:179–207.
3. Ge Z, Hiratsuka K, Taylor DE. Nucleotide sequence and mutational analysis indicate that two *Helicobacter pylori* genes encode a P-type ATPase and a cation-binding protein associated with copper transport. Mol Microbiol 1995;15:97–106.
4. Labigne-Roussel A, Courcoux P, Tompkins L. Gene disruption and replacement as a feasible approach for mutagenesis of *Campylobacter jejuni*. J Bacteriol 1988; 170:1704–1708.
5. Skoglund CM, Smith HO, Chandrasegaran S. Construction of an efficient overproducer clone of HinfI restriction endonuclease using the polymerase chain reaction. Gene 1990;88:1–5.
6. Yanisch-Perron C, Vieira J, Messing J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 1985;33:103–119.



Supplementary Figure 1. IgG antibody responses to vaccine components. (*A*) Increases of antibody titers against each of the three vaccine antigens 12 weeks after challenge over baseline titers. (*B*) Increases of antibody titers against CagA (left) or VacA (right) 12 weeks after challenge over baseline titers for volunteers harboring the three isolates not able to induce IL-8 *vs*. the rest of the isolates (left), or the three isolates with an inactivation of *vacA vs*. the rest of the isolates (right). Statistical significance was analyzed by Mann Whitney test (**$P < 0.01$).



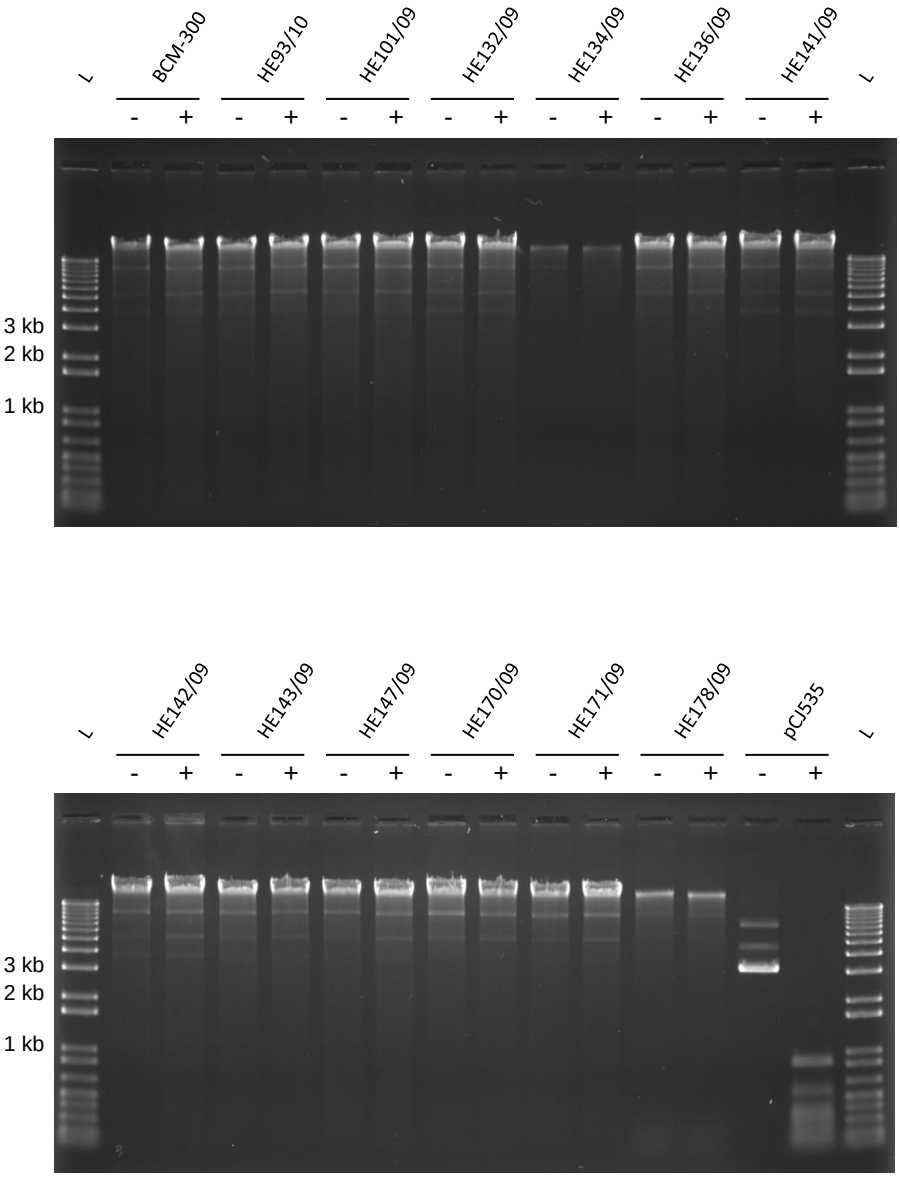
Supplementary Figure 2. BabA protein expression and Le(b) binding. (*A*) Whole cell extracts of *H. pylori* strains were analyzed for protein expression of BabA. (*B*) Le(b) binding of BCM-300 was determined by ELISA. *H. pylori* strain J99 and its isogenic *babA* mutant were used as positive and negative controls, respectively. Le(b) binding is depicted relative to J99 set to 100%. Data are presented as mean ± SD from three independent experiments.



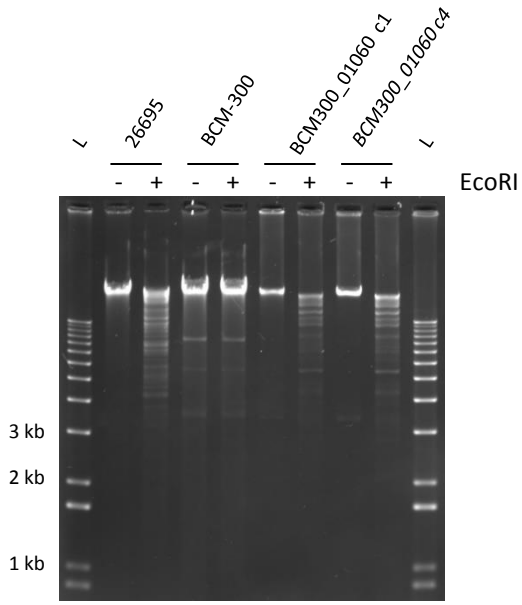
Outwards to inwards

1. *cag*PAI, IS*Hp609*
2. ORFs plus strand
3. ORFs minus strand
4. AGG^{m6}**A**G
5. CAC^{m6}**A**T
6. C^{m6}**A**TG
7. CC^{m6}**A**TC
8. ^{m4}**C**CGG
9. GA^{m6}**A**TTC
10. G^{m6}**A**CY
11. G^{m6}**A**GG
12. G^{m6}**A**TC
13. G^{m6}**A**TGG
14. GTS^{m6}**A**C
15. TCNNG^{m6}**A**
16. TGC^{m6}**A**

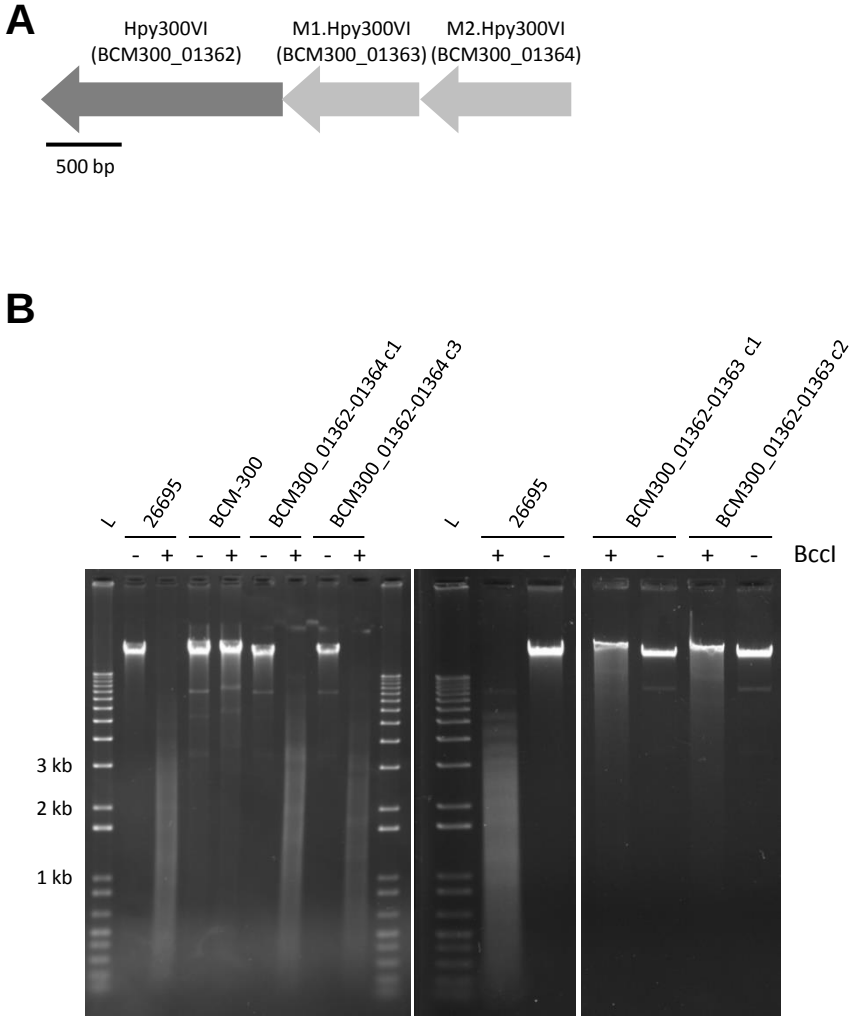
Supplementary Figure 3. Circular representation of *H. pylori* BCM-300 genome displaying the distribution of methylated bases. The representation was generated using Circos[1]. The outmost circle displays the localization of the *cag*PAI (black) and the five copies of IS*Hp609* (dark grey), followed by circles depicting ORFs on plus and minus strand (both dark orange), and methylated motifs (see legend). Please note, that methylation of two motifs is not depicted: 5´-CCTYN^{m6}**A**-3´ was only detected in strain HE143/09 and 5'-G^{m5}**C**GC-3' was not detected by SMRT sequencing.



Supplementary Figure 4. Confirmation of 5´-G^{m5}CGC-3´ MTase activity. Genomic DNA (300 ng) of *H. pylori* BCM-300 and the reisolates were either incubated without (-) or with (+) the restriction endonuclease HhaI cleaving non-methylated GCGC sites. Plasmid DNA of pCJ535 lacking 5´-G^{m5}**C**GC-3´ methylation was used as control for HhaI activity. L, 1 kb Plus DNA Ladder (Invitrogen).

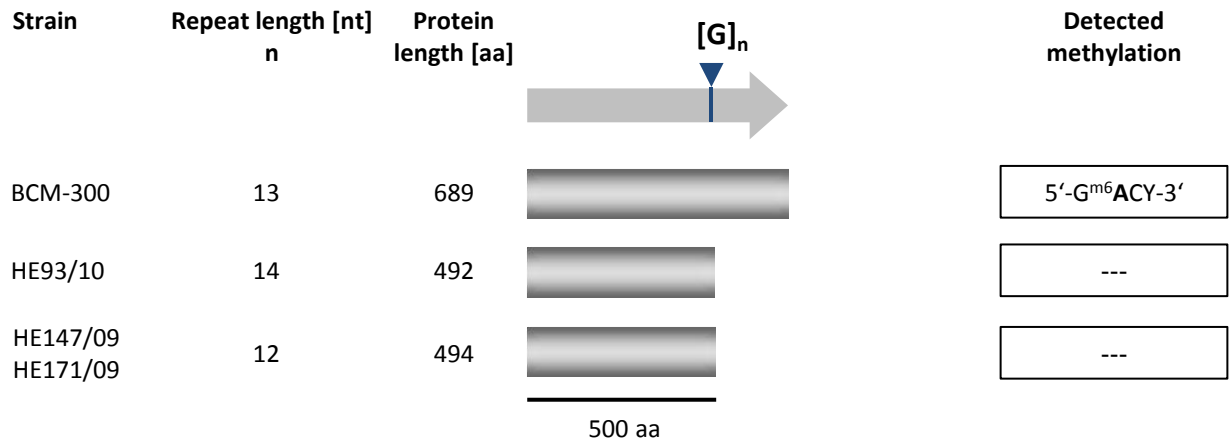


Supplementary Figure 5. Functionality of the Hpy300III R-M system (EcoRI homolog). Analysis of BCM300_01060 MTase activity was determined by DNA restriction digestion experiment. Genomic DNA (300 ng) of *H. pylori* wild type and isogenic mutant strains was either incubated without (-) or with (+) EcoRI (GAATTC). *H. pylori* 26695 was used as control as it is known to be susceptible to EcoRI cleavage. Two independent clones were assayed for MTase activity. L, 1 kb Plus DNA Ladder (Invitrogen).

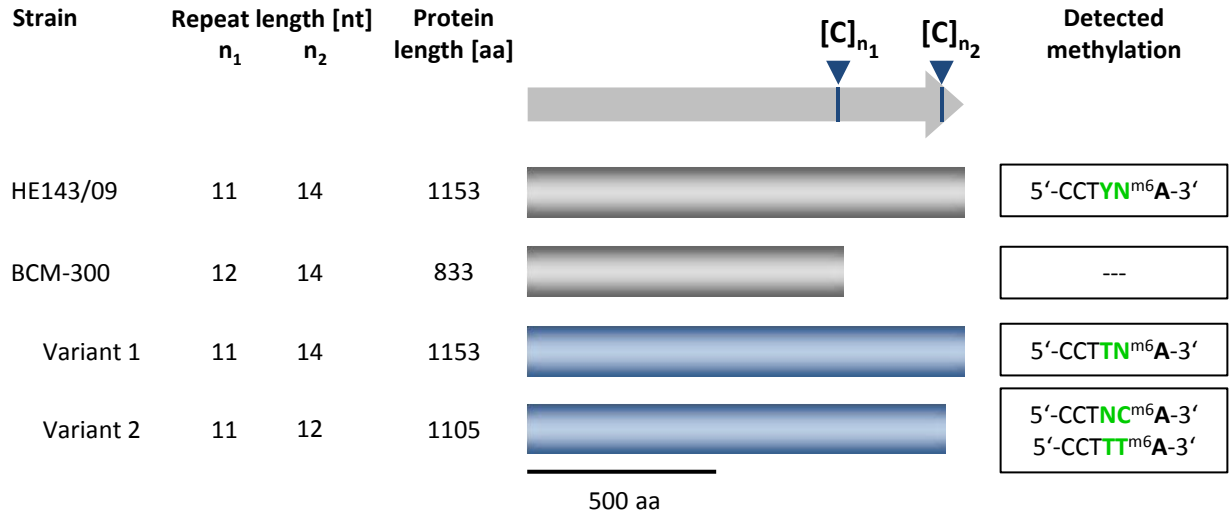


Supplementary Figure 6. Organization and functionality of the Hpy300VI R-M system (BccI homolog). (*A*) Genetic organization of the BccI locus: MTase genes (M) are depicted in light grey and the REase gene in dark grey. (*B*) Analysis of MTase activity was determined by DNA restriction digestion experiments. Genomic DNA (300 ng) of *H. pylori* wild type and isogenic mutant strains was either incubated without (-) or with (+) BccI (CCATC). *H. pylori* 26695 was used as control as it is known to be susceptible to BccI cleavage. Two independent clones were assayed for MTase activity. L, 1 kb Plus DNA Ladder (Invitrogen).

A **Type III MTase M.Hpy300X**



B **Type IIG R-M system Hpy300XI**



Supplementary Figure 7. Functional characterization of phase-variable R-M systems M.Hpy300X and Hpy300XI. Genes are depicted as grey arrows and proteins as grey (expressed in *H. pylori*) or blue bars (expressed in *E. coli*). Triangles plus vertical lines highlight the position of homopolymeric repeat tracts and the letter above indicates the type of nucleotide. MTase activities were analyzed by SMRT sequencing of either *H. pylori* or *E. coli* genomic DNA (dashed lines indicate that methylation of the relevant motif was not detected). (*A*) The phase-variable expression of the MTase M.Hpy300X is mediated by length variation of a homopolymeric G repeat. Deviations in nucleotide number inhibit translation of a full-length protein for isolates HE93/10, HE147/09 and HE171/09, for which methylation of 5´-G^{m6}**A**CY-3´ was not detected by SMRT sequencing of the *H. pylori* genomic DNAs. (*B*) For HE143/09 both repeat tracts are in frame and the full-length MTase mediates methylation of 5´-CCTYN^{m6}**A**-3´. In BCM-300, an authentic frameshift within the first repeat prevents translation of a full-length protein resulting in a lack of methylation. Two different alleles of BCM-300 Hpy300XI were constructed by frameshift correction and repeat stabilization through site-directed mutagenesis. Both alleles were recombinantly expressed and methylation pattern of the *E. coli* genomic DNAs was analyzed by SMRT sequencing confirming the specificity switch at positions -1 and -2 relative to the modified A (indicated by green color).

References

1. Krzywinski M, Schein J, Birol I, *et al.* Circos: an information aesthetic for comparative genomics. Genome Res 2009;19:1639-1645.

Supplementary Dataset 1:
List of genomic differences between *H. pylori* challenge strain BCM-300 and reisolates

[1]The genome annotation was automatically generated by Kodon based on annotation of H. pylori strain 26695 (NC_000915.1), and then manually curated if necessary.
[2]The genome annotation of BCM-300 was generated by Prokka (v1.7).

HE101/09

No.	Position in BCM-300		Sequence in		Type of difference		Annotation[1]			BCM-300[2]
	Start	End	BCM-300	Reisolate			Gene description	Gene name	Locus tag	Locus tag
1	25796	25800	5 Cs	4 Cs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
2	28076	28091	16 As	15 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
3	32486	32490	5 Gs	6 Gs	indel	1 bp insertion	ATP-dependent dethiobiotin synthetase BioD	bioD	hp0029	BCM300_00035
4	71801	71806	6 Gs	5 Gs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
5	98334	98339	6 As	5 As	indel	1 bp deletion	type II restriction enzyme M protein (hsdM)	M.HpyAIII	hp0092	BCM300_00102
6	99174	99185	12 Gs	11 Gs	indel	1 bp deletion	alpha-1,2-fucosyltransferase	futC	hp0094	BCM300_00103
7	169575	169588	14 Gs	13 Gs	indel	1 bp deletion	signal-transducing protein, histidine kinase	-	hp0164	BCM300_00169
8	177965	177971	7 As	6 As	indel	1 bp deletion	hypothetical protein	-	hp0174	BCM300_00178
9	178003	178009	7 As	6 As	indel	1 bp deletion	hypothetical protein	-	hp0174	BCM300_00178
10	193058	193066	9 As	8 As	indel	1 bp deletion	hypothetical protein	-	hp0190	BCM300_00191
11	222870	222881	12 Gs	11 Gs	indel	1 bp deletion	beta-1,4-N-acetylgalactosamyltransferase	cgtA	hp0217	BCM300_00220
12	229250	229255	6 Ts	5 Ts	indel	1 bp deletion	bifunctional methionine sulfoxide reductase subu	msrA	hp0224	intergenic: no annotation
13	231460	231466	7 Cs	6 Cs	indel	1 bp deletion	membrane protein	-	hp0226	BCM300_00228
14	350265	350279	15 As	14 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
15	450369	450379	11 Gs	12 Gs	indel	1 bp insertion	type I restriction-modification system endonuclea	hsdR	hp0464	BCM300_00442
16	486962	486966	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	BCM300_00476
17	504277	504281	5 Ts	4 Ts	indel	1 bp deletion	virB4 homolog (virB4)	virB4	hp0459	BCM300_00493
18	516061	516067	7 As	6 As	indel	1 bp deletion	hypothetical protein	-	hp0488	BCM300_00507
19	526366	526372	7 Gs	8 Gs	indel	1 bp insertion	phospholipase A1 precursor (DR-phospholipase A	pldA	hp0499	BCM300_00519
20	559284	559291	8 Ts	7 Ts	indel	1 bp deletion	cag pathogenicity island protein (cag7)	cagY	hp0527	BCM300_00549
21	604505	604511	7 Gs	6 Gs	indel	1 bp deletion	aminopeptidase a/i (pepA)	pepA	hp0570	BCM300_00591
22	663718	663732	15 Cs	14 Cs	indel	1 bp deletion	predicted LPS biosynthesis protein	-	hp0619	BCM300_00638
23	727943	727947	5 Cs	6 Cs	indel	1 bp insertion	methylated-DNA--protein-cysteine methyltransfe	dat1	hp0676	BCM300_00697
24	780145	780159	15 Ts	14 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
25	786911	786928	18 As	17 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
26	862628	862631	4 Gs	5 Gs	indel	1 bp insertion	sigma-54 interacting protein	comM	hp0792	BCM300_00832
27	864414	864418	5 As	6 As	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00836
28	864639		1 C	2 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	intergenic: no annotation
29	864688	864691	4 Ts	5 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00837
30	864943	864944	2 Cs	3 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00838
31	865147	865148	2 Ts	3 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00839
32	877934	877946	13 Ts	12 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	BCM300_00856
33	936941	936954	14 Ts	13 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
34	948014		C	T	SNP	non-synonymous	vacuolating cytotoxin	vacA	hp0887	BCM300_00927
35	960612	960615	4 Cs	5 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
36	979321	979335	15 Ts	14 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
37	985093	985106	14 As	13 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
38	1081106	1081409	38 x 8-nt repeat	39 x 8-nt repeat	indel	8 bp insertion [repeat: CATTCTAT]	hypothetical protein	-	hp0427	BCM300_01049
39	1081414	1081415		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
40	1081419	1081420		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
41	1081421		1 A	2 As	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
42	1222342	1222346	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
43	1222580	1222585	6 Cs	5 Cs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
44	1224971	1224985	15 Ts	14 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
45	1225114	1225253	14 x 10-nt repeat	13 x 10-nt repeat	indel	10 bp deletion [repeat:	intergenic: no annotation	-	-	intergenic: no annotation
46	1236530	1236534	5 Cs	4 Cs	indel	1 bp deletion	protein translocation protein, low temperature (s	secG	hp1255	BCM300_01197
47	1256981	1256995	15 As	14 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
48	1287348	1287387	5 x 8-nt repeat	6 x 8-nt repeat	indel	8 bp insertion [repeat: GTATTTAT]	intergenic: no annotation	-	-	intergenic: no annotation
49	1304687	1304701	15 As	14 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
50	1317399	1317412	14 Gs	15 Gs	indel	1 bp insertion	putative adenine specific DNA methyltransferase	-	hp1354	BCM300_01297
51	1318224	1318234	11 Gs	12 Gs	indel	1 bp insertion	putative adenine specific DNA methyltransferase	-	hp1354	BCM300_01297
52	1343408	1343418	11 Ts	10 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
53	1419748	1419752	5 Gs	4 Gs	indel	1 bp deletion	hypothetical protein	-	hp1479	BCM300_01384
54	1478232	1478238	7 As	6 As	indel	1 bp deletion	hypothetical protein	-	hp1127	BCM300_01439
55	1488920	1488926	7 As	8 As	indel	1 bp insertion	Sel1 repeat-containing protein	hcpX	hp1117	BCM300_01451
56	1503023	1503044	22 Ts	20 Ts	indel	2 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
57	1517648	1517721		-74 bp	indel	74 bp deletion	UDP-glucose 4-epimerase	galE	hp0360	BCM300_01480
58	1536449	1536458	10 As	9 As	indel	1 bp deletion	fucosyltransferase	futA	hp0379	BCM300_01500
59	1562284	1562289	6 Ts	5 Ts	indel	1 bp deletion	phenylalanyl-tRNA synthetase, beta subunit (pheT	pheT	hp0402	BCM300_01523
60	1587896	1588151	32 x 8-nt repeat	35 x repeat	indel	24 bp insertion [repeat: GAATGATA]	hypothetical protein	-	hp0427	BCM300_01551
61	1652479	1652486	8 As	7 As	indel	1 bp deletion	GTP-binding protein EngB	engB	hp1567	BCM300_01607

HE132/09

No.	Position in BCM-300		Sequence in		Type of difference		Annotation[1]			BCM-300[2]
	Start	End	BCM-300	Reisolate			Gene description	Gene name	Locus tag	Locus tag
1	9816	9829	14 Ts	15 Ts	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
2	32486	32490	5 Gs	6 Gs	indel	1 bp insertion	ATP-dependent dethiobiotin synthetase BioD	bioD	hp0029	BCM300_00035
3	61680	61688	9 Cs	10 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	BCM300_00061
4	71831	71835	5 Gs	6 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
5	110540	110553	14 Cs	15 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
6	166606	166821		+216 bp	indel	duplication of 216-bp sequence	beta-lactamase HcpD	hcpD	hp0160	BCM300_00166
7	222870	222881	12 Gs	11 Gs	indel	1 bp deletion	beta-1,4-N-acetylgalactosamyltransferase	cgtA	hp0217	BCM300_00220
8	450369	450379	11 Gs	13 Gs	indel	2 bp insertion	type I restriction-modification system endonucleas	hsdR	hp0464	BCM300_00442
9	486962	486966	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	BCM300_00476
10	526366	526372	7 Gs	8 Gs	indel	1 bp insertion	phospholipase A1 precursor (DR-phospholipase A)	pldA	hp0499	BCM300_00519
11	581663		C	T	SNP	non-synonymous	cag pathogenicity island protein (cag26)	cagA	hp0547	BCM300_00568
12	662004	662235			intrachromosomal	232-bp fragment recombined; source: 663338-663572	intergenic: no annotation	-	-	BCM300_00637
13	697107		A	G	SNP	synonymous	fucosyltransferase	futB	hp0651	BCM300_00672
14	727943	727947	5 Cs	6 Cs	indel	1 bp insertion	methylated-DNA--protein-cysteine methyltransfer	dat1	hp0676	BCM300_00697
15	780276	780291	8 x CT repeat	9 x CT repeat	indel	2 bp insertion	outer membrane protein (omp17)	sabA	hp0725	BCM300_00748
16	786776	786793	9 x GA repeat	8 x GA repeat	indel	2 bp deletion	outer membrane protein (omp16)	sabB	hp0722	BCM300_00751
17	862628	862631	4 Gs	5 Gs	indel	1 bp insertion	sigma-54 interacting protein	comM	hp0792	BCM300_00832
18	864414	864418	5 As	6 As	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00836
19	864639		1 C	2 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	intergenic: no annotation
20	864688	864691	4 Ts	5 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00837
21	864943	864944	2 Cs	3 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00838
22	865147	865148	2 Ts	3 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00839
23	875914		G	A	SNP	intergenic	intergenic: no annotation	-	-	intergenic: no annotation
24	877934	877946	13 Ts	14 Ts	indel	1 bp insertion	intergenic: no annotation	-	-	BCM300_00856
25	960612	960615	4 Cs	5 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
26	1081369	1081370		+37 bp	indel	37 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
27	1081414	1081415		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
28	1081419	1081420		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
29	1081421		1 A	2 As	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
30	1222342	1222346	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
31	1317399	1317412	14 Gs	13 Gs	indel	1 bp deletion	putative adenine specific DNA methyltransferase	-	hp1354	BCM300_01297
32	1318224	1318234	11 Gs	12 Gs	indel	1 bp insertion	putative adenine specific DNA methyltransferase	-	hp1354	BCM300_01297
33	1370527	1370539	13 Gs	14 Gs	indel	1 bp insertion	predicted type III R-M system modification enzyme	-	hp1522	BCM300_01342
34	1488920	1488926	7 As	8 As	indel	1 bp insertion	Sel1 repeat-containing protein	hcpX	hp1117	BCM300_01451
35	1587927	1587928		+12 bp	indel	12 bp insertion	hypothetical protein	-	hp0427	BCM300_01551
36	1592310	1592315	6 Gs	7 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation

HE134/09

No.	Position in BCM-300		Sequence in		Type of difference		Annotation[1]			BCM-300[2]
	Start	End	BCM-300	Reisolate			Gene description	Gene name	Locus tag	Locus tag
1	9816	9829	14 Ts	13 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
2	32486	32490	5 Gs	6 Gs	indel	1 bp insertion	ATP-dependent dethiobiotin synthetase BioD	bioD	hp0029	BCM300_00035
3	71831	71835	5 Gs	6 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
4	99174	99185	12 Gs	13 Gs	indel	1 bp insertion	alpha-1,2-fucosyltransferase	futC	hp0094	BCM300_00103
5	169575	169588	14 Gs	13 Gs	indel	1 bp deletion	signal-transducing protein, histidine kinase	-	hp0164	BCM300_00169
6	222870	222881	12 Gs	11 Gs	indel	1 bp deletion	beta-1,4-N-acetylgalactosamyltransferase	cgtA	hp0217	BCM300_00220
7	260916		T	C	SNP	non-synonymous	outer membrane protein (omp8)	hopG	hp0253-254	BCM300_00257
8	365123	365129	7 Gs	6 Gs	indel	1 bp deletion	cell division protein (ftsK)	ftsK	hp1090	BCM300_00358
9	401200		G	A	SNP	synonymous	membrane protein	-	hp1055	BCM300_00395
10	450369	450379	11 Gs	12 Gs	indel	1 bp insertion	type I restriction-modification system endonucleas	hsdR	hp0464	BCM300_00442
11	486962	486966	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	BCM300_00476
12	526366	526372	7 Gs	8 Gs	indel	1 bp insertion	phospholipase A1 precursor (DR-phospholipase A)	pldA	hp0499	BCM300_00519
13	662803	662807	5 As	3 As	indel	2 bp deletion	intergenic: no annotation	-	-	BCM300_00637
14	663718	663732	15 Cs	13 cs	indel	2 bp deletion	predicted LPS biosynthesis protein	-	hp0619	BCM300_00638
15	705587	705595	9 Gs	8 Gs	indel	1 bp deletion	processing protease (ymxG)	ymxG	hp0657	BCM300_00678
16	727943	727947	5 Cs	6 Cs	indel	1 bp insertion	methylated-DNA--protein-cysteine methyltransfer	dat1	hp0676	BCM300_00697
17	826877	826887	11 Cs	10 Cs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
18	862628	862631	4 Gs	5 Gs	indel	1 bp insertion	sigma-54 interacting protein	comM	hp0792	BCM300_00832
19	864414	864418	5 As	6 As	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00836
20	864639		1 C	2 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	intergenic: no annotation
21	864688	864691	4 Ts	5 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00837
22	864943	864944	2 Cs	3 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00838
23	865147	865148	2 Ts	3 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00839
24	960612	960615	4 Cs	5 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
25	985093	985106	14 As	15 As	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
26	1081106	1081409	38 x 8-nt repeat	37 x 8-nt repeat	indel	8 bp deletion [repeat: CATTCTAT]	hypothetical protein	-	hp0427	BCM300_01049
27	1081414	1081415		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
28	1081419	1081420		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
29	1081421		1 A	2 As	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
30	1221905	1221912	8 Cs	7 Cs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
31	1222342	1222346	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
32	1309317	1309324	8 Cs	7 Cs	indel	1 bp deletion	phosphoglycerate kinase	pgk	hp1345	BCM300_01287
33	1317399	1317412	14 Gs	13 Gs	indel	1 bp deletion	putative adenine specific DNA methyltransferase	-	hp1354	BCM300_01297
34	1347536	1347549	14 As	15 As	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
35	1503023	1503044	22 Ts	21 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
36	1587896	1588151	32 x 8-nt repeat	29 x 8-nt repeat	indel	24 bp deletion [repeat: GAATGATA]	hypothetical protein	-	hp0427	BCM300_01551
37	1592310	1592315	6 Gs	7 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation

HE136/09

No.	Position in BCM-300		Sequence in		Type of difference		Annotation[1]			BCM-300[2]
	Start	End	BCM-300	Reisolate			Gene description	Gene name	Locus tag	Locus tag
1	32486	32490	5 Gs	6 Gs	indel	1 bp insertion	ATP-dependent dethiobiotin synthetase BioD	bioD	hp0029	BCM300_00035
2	61680	61688	9 Cs	10 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	BCM300_00061
3	71831	71835	5 Gs	6 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
4	99174	99185	12 Gs	14 Gs	indel	2 bp insertion	alpha-1,2-fucosyltransferase	futC	hp0094	BCM300_00103
5	110540	110553	14 Cs	15 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
6	230344	230351	8 Gs	9 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
7	344681	344689	9 Gs	10 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
8	381791	381797	7 Gs	8 Gs	indel	1 bp insertion	nickel transport protein (nixA)	nixA	hp1077	BCM300_00373
9	427626	427630	5 Gs	6 Gs	indel	1 bp insertion	recombination factor protein RarA	-	hp1026	BCM300_00423
10	450369	450379	11 Gs	12 Gs	indel	1 bp insertion	type I restriction-modification system endonuclea	hsdR	hp0464	BCM300_00442
11	486962	486966	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	BCM300_00476
12	526366	526372	7 Gs	8 Gs	indel	1 bp insertion	phospholipase A1 precursor (DR-phospholipase A	pldA	hp0499	BCM300_00519
13	535902	535907	6 Gs	7 Gs	indel	1 bp insertion	glycolate oxidase subunit (glcD)	glcD	hp0509	BCM300_00530
14	653515	653521	7 Gs	8 Gs	indel	1 bp insertion	toxin-like outer membrane protein	-	hp0610	BCM300_00631
15	662723	662731	9 As	8 As	indel	1 bp deletion	intergenic: no annotation	-	-	BCM300_00637
16	727943	727947	5 Cs	6 Cs	indel	1 bp insertion	methylated-DNA--protein-cysteine methyltransfe	dat1	hp0676	BCM300_00697
17	729048	729053	6 Cs	7 Cs	indel	1 bp insertion	conserved hypothetical integral membrane protei	-	hp0677	BCM300_00698
18	739831	739838	8 Cs	9 Cs	indel	1 bp insertion	iron(II) transport protein (feoB)	feoB	hp0687	BCM300_00706
19	752910	752916	7 Cs	8 Cs	indel	1 bp insertion	conserved hypothetical integral membrane protei	atoE	hp0693	BCM300_00722
20	839307	839313	7 Gs	8 Gs	indel	1 bp insertion	aconitase B (acnB)	acnB	hp0779	BCM300_00809
21	839961	839967	7 Gs	8 Gs	indel	1 bp insertion	aconitase B (acnB)	acnB	hp0779	BCM300_00809
22	847847	847851	5 Gs	6 Gs	indel	1 bp insertion	preprotein translocase subunit (secA)	secA	hp0786	BCM300_00816
23	862628	862631	4 Gs	5 Gs	indel	1 bp insertion	sigma-54 interacting protein	comM	hp0792	BCM300_00832
24	864414	864418	5 As	6 As	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00836
25	864639		1 C	2 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	intergenic: no annotation
26	864688	864691	4 Ts	5 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00837
27	864943	864944	2 Cs	3 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00838
28	865147	865148	2 Ts	3 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00839
29	924452		G	A	SNP	synonymous	pantothenate kinase	-	hp0862	BCM300_00904
30	960612	960615	4 Cs	5 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
31	985093	985106	14 As	13 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
32	1029483		C	A	SNP	non-synonymous	hypothetical protein	-	hp0953	BCM300_00997
33	1081106	1081409	38 x 8-nt repeat	40 x 8-nt repeat	indel	16 bp insertion [repeat: CATTCTAT]	hypothetical protein	-	hp0427	BCM300_01049
34	1081414	1081415		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
35	1081419	1081420		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
36	1081421		1 A	2 As	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
37	1189877	1189883	7 Cs	8 Cs	indel	1 bp insertion	ulcer associated adenine specific DNA methyltran	M.HpyAI	hp1208	BCM300_01150
38	1222342	1222346	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
39	1304687	1304701	15 As	16 As	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
40	1318224	1318234	11 Gs	12 Gs	indel	1 bp insertion	putative adenine specific DNA methyltransferase	-	hp1354	BCM300_01297
41	1370527	1370539	13 Gs	14 Gs	indel	1 bp insertion	predicted type III R-M system modification enzym	-	hp1522	BCM300_01342
42	1426091	1426096	6 Gs	7 Gs	indel	1 bp insertion	thymidylate kinase (tmk)	tmk	hp1474	BCM300_01389
43	1486540	1486548	9 Gs	10 Gs	indel	1 bp insertion	gamma-glutamyltranspeptidase (ggt)	ggt	hp1118	BCM300_01449
44	1537936	1537942	7 Cs	8 Cs	indel	1 bp insertion	glutamate dehydrogenase (gdhA)	gdhA	hp0380	BCM300_01501
45	1587896	1588151	32 x 8-nt repeat	37 x 8-nt repeat	indel	40 bp insertion [repeat: GAATGATA]	hypothetical protein	-	hp0427	BCM300_01551
46	1592310	1592315	6 Gs	7 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation

HE141/09

No.	Position in BCM-300		Sequence in		Type of difference		Annotation[1]			BCM-300[2]
	Start	End	BCM-300	Reisolate			Gene description	Gene name	Locus tag	Locus tag
1	28076	28091	16 As	15 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
2	32486	32490	5 Gs	6 Gs	indel	1 bp insertion	ATP-dependent dethiobiotin synthetase BioD	bioD	hp0029	BCM300_00035
3	71786	71791	6 Gs	5 Gs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
4	71801	71806	6 Gs	5 Gs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
5	71831	71835	5 Gs	6 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
6	99174	99185	12 Gs	13 Gs	indel	1 bp insertion	alpha-1,2-fucosyltransferase	futC	hp0094	BCM300_00103
7	110540	110553	14 Cs	15 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
8	169575	169588	14 Gs	13 Gs	indel	1 bp deletion	signal-transducing protein, histidine kinase	-	hp0164	BCM300_00169
9	222870	222881	12 Gs	11 Gs	indel	1 bp deletion	beta-1,4-N-acetylgalactosamyltransferase	cgtA	hp0217	BCM300_00220
10	256678	256684	7 Ts	6 Ts	indel	1 bp deletion	oligopeptide ABC transporter, ATP-binding protei	oppD	hp0250	BCM300_00253
11	450369	450379	11 Gs	12 Gs	indel	1 bp insertion	type I restriction-modification system endonuclea	hsdR	hp0464	BCM300_00442
12	486962	486966	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	BCM300_00476
13	526366	526372	7 Gs	8 Gs	indel	1 bp insertion	phospholipase A1 precursor (DR-phospholipase A	pldA	hp0499	BCM300_00519
14	573786	573792	7 Cs	8 Cs	indel	1 bp insertion	cag pathogenicity island protein (cag19)	cagI	hp0540	BCM300_00560
15	581292	581293		+2398 bp	indel	transposon insertion	intergenic: no annotation	-	-	intergenic: no annotation
16	727943	727947	5 Cs	6 Cs	indel	1 bp insertion	methylated-DNA--protein-cysteine methyltransfe	dat1	hp0676	BCM300_00697
17	780276	780291	8 x CT repeat	7 x CT repeat	indel	2 bp deletion	outer membrane protein (omp17)	sabA	hp0725	BCM300_00748
18	786776	786793	9 x GA repeat	8 x GA repeat	indel	2 bp deletion	outer membrane protein (omp16)	sabB	hp0722	BCM300_00751
19	862628	862631	4 Gs	5 Gs	indel	1 bp insertion	sigma-54 interacting protein	comM	hp0792	BCM300_00832
20	864414	864418	5 As	6 As	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00836
21	864639		1 C	2 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	intergenic: no annotation
22	864688	864691	4 Ts	5 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00837
23	864943	864944	2 Cs	3 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00838
24	865147	865148	2 Ts	3 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00839
25	903106	903113	8 Gs	7 Gs	indel	1 bp deletion	outer membrane protein P1 (ompP1)	ompP1	hp0839	BCM300_00884
26	960239	960323	5 x 17-nt repeat	6 x 17-nt repeat	indel	17 bp insertion [repeat: CAAAGAAAAAGGGAGTT]	intergenic: no annotation	-	-	intergenic: no annotation
27	960612	960615	4 Cs	5 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
28	979321	979335	15 Ts	16 Ts	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
29	985093	985106	14 As	13 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
30	1002974	1002980	7 Cs	8 Cs	indel	1 bp insertion	recombinational DNA repair protein (recR)	recR	hp0925	BCM300_00967
31	1029281		G	A	SNP	non-synonymous	hypothetical protein	-	hp0953	BCM300_00997
32	1081106	1081409	38 x 8-nt repeat	47 x 8-nt repeat	indel	72 bp insertion [repeat: CATTCTAT]	hypothetical protein	-	hp0427	BCM300_01049
33	1081414	1081415		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
34	1081419	1081420		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
35	1081421		1 A	2 As	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
36	1218049	1218061	13 Ts	14 Ts	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
37	1222342	1222346	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
38	1304687	1304701	15 As	14 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
39	1318224	1318234	11 Gs	12 Gs	indel	1 bp insertion	putative adenine specific DNA methyltransferase	-	hp1354	BCM300_01297
40	1471570		C	T	SNP	synonymous	ATP synthase F0, subunit b (atpF)	atpF	hp1136	BCM300_01431
41	1503023	1503044	22 Ts	23 Ts	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
42	1592310	1592315	6 Gs	7 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
43	1594330	1594344	15 Cs	16 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation

HE142/09

No.	Position in BCM-300		Sequence in		Type of difference		Annotation[1]			BCM-300[2]
	Start	End	BCM-300	Reisolate			Gene description	Gene name	Locus tag	Locus tag
1	32486	32490	5 Gs	6 Gs	indel	1 bp insertion	ATP-dependent dethiobiotin synthetase BioD	bioD	hp0029	BCM300_00035
2	71786	71859	5 x 15-nt repeat	6 x 15-nt repeat	indel	15 bp insertion [repeat: GGGGGGTTAAAAAAA]	intergenic: no annotation	-	-	intergenic: no annotation
3	71831	71835	5 Gs	6 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
4	99174	99185	12 Gs	13 Gs	indel	1 bp insertion	alpha-1,2-fucosyltransferase	futC	hp0094	BCM300_00103
5	222870	222881	12 Gs	11 Gs	indel	1 bp deletion	beta-1,4-N-acetylgalactosamyltransferase	cgtA	hp0217	BCM300_00220
6	233960	233973	14 As	13 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
7	450369	450379	11 Gs	12 Gs	indel	1 bp insertion	type I restriction-modification system endonucleas	hsdR	hp0464	BCM300_00442
8	486962	486966	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	BCM300_00476
9	557853	557859	7 Ts	8 Ts	indel	1 bp insertion	cag pathogenicity island protein (cag7)	cagY	hp0527	BCM300_00549
10	663718	663732	15 Cs	16 Cs	indel	1 bp insertion	predicted LPS biosynthesis protein	-	hp0619	BCM300_00638
11	727943	727947	5 Cs	6 Cs	indel	1 bp insertion	methylated-DNA--protein-cysteine methyltransfer	dat1	hp0676	BCM300_00697
12	762531		A	G	SNP	intergenic	intergenic: no annotation	-	-	intergenic: no annotation
13	862628	862631	4 Gs	5 Gs	indel	1 bp insertion	sigma-54 interacting protein	comM	hp0792	BCM300_00832
14	864414	864418	5 As	6 As	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00836
15	864639		1 C	2 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	intergenic: no annotation
16	864688	864691	4 Ts	5 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00837
17	864943	864944	2 Cs	3 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00838
18	865147	865148	2 Ts	3 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00839
19	936941	936954	14 Ts	13 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
20	943555	943559	5 Cs	6 Cs	indel	1 bp insertion	hypothetical protein	-	hp0884	BCM300_00924
21	960612	960615	4 Cs	5 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
22	979321	979335	15 Ts	14 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
23	1078938	1078943	6 Cs	7 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	BCM300_01048
24	1081369	1081370		+12 bp	indel	12 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
25	1081414	1081415		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
26	1081419	1081420		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
27	1081421		1 A	2 As	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
28	1222342	1222346	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
29	1236438	1236444	7 As	6 As	indel	1 bp deletion	protein translocation protein, low temperature (se	secG	hp1255	BCM300_01197
30	1287348	1287387	5 x 8-nt repeat	6 x 8-nt repeat	indel	8 bp insertion [repeat: GTATTTAT]	intergenic: no annotation	-	-	intergenic: no annotation
31	1370527	1370539	13 Gs	14 Gs	indel	1 bp insertion	predicted type III R-M system modification enzyme	-	hp1522	BCM300_01342
32	1488893		C	T	SNP	non-synonymous	Sel1 repeat-containing protein	hcpX	hp1117	BCM300_01451
33	1540693		C	T	SNP	non-synonymous	zinc-metallo protease (YJR117W)	YJR117W	hp0382	BCM300_01503
34	1587896	1588151	32 x 8-nt repeat	31 x 8-nt repeat	indel	8 bp deletion [repeat: GAATGATA]	hypothetical protein	-	hp0427	BCM300_01551
35	1592310	1592315	6 Gs	7 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
36	1594330	1594344	15 Cs	16 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation

HE143/09

No.	Position in BCM-300		Sequence in		Type of difference		Annotation[1]			BCM-300[2]
	Start	End	BCM-300	Reisolate			Gene description	Gene name	Locus tag	Locus tag
1	7503	7510	8 As	7 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
2	7558	7563	6 Cs	5 Cs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
3	32486	32490	5 Gs	6 Gs	indel	1 bp insertion	ATP-dependent dethiobiotin synthetase BioD	bioD	hp0029	BCM300_00035
4	71816	71821	6 Gs	5 Gs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
5	137983		1 T	2 Ts	indel	1 bp insertion	hypothetical protein	-	hp0130	BCM300_00136
6	157600	157606	7 As	6 As	indel	1 bp deletion	hypothetical protein	-	hp0150	BCM300_00156
7	167711	167716	6 Cs	5 Cs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
8	169575	169588	14 Gs	13 Gs	indel	1 bp deletion	signal-transducing protein, histidine kinase	-	hp0164	BCM300_00169
9	220712	220718	7 Gs	6 Gs	indel	1 bp deletion	CDP-diglyceride synthetase (cdsA)	cdsA	hp0215	BCM300_00218
10	221283	221285	3 Gs	2 Gs	indel	1 bp deletion	CDP-diglyceride synthetase (cdsA)	cdsA	hp0215	BCM300_00218
11	228036	228042	7 Gs	6 Gs	indel	1 bp deletion	ATP-dependent protease (sms)	sms	hp0223	BCM300_00226
12	233960	233973	14 As	13 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
13	259853		G	A	SNP	non-synonymous	outer membrane protein (omp7)	hopF	hp0252	BCM300_00256
14	263848	263853	6 Cs	5 Cs	indel	1 bp deletion	hypothetical protein	-	hp0257	BCM300_00260
15	289362	289368	7 As	6 As	indel	1 bp deletion	3-dehydroquinate synthase (aroB)	aroB	hp0283	BCM300_00286
16	344826	344829	4 Gs	3 Gs	indel	1 bp deletion	beta-lactamase HcpC	hcpC	hp1098	intergenic: no annotation
17	377230	377235	6 Gs	5 Gs	indel	1 bp deletion	multidrug resistance protein (msbA)	msbA	hp1082	BCM300_00366
18	439295	439300	6 Cs	5 Cs	indel	1 bp deletion	4-hydroxy-tetrahydrodipicolinate synthase	dapA	hp1013	BCM300_00435
19	454861	454866	6 Gs	5 Gs	indel	1 bp deletion	hypothetical protein HP0468	-	hp0468	BCM300_00446
20	486962	486966	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	BCM300_00476
21	526366	526372	7 Gs	8 Gs	indel	1 bp insertion	phospholipase A1 precursor (DR-phospholipase A)	pldA	hp0499	BCM300_00519
22	535707	535713	7 Gs	6 Gs	indel	1 bp deletion	glycolate oxidase subunit (glcD)	glcD	hp0509	BCM300_00530
23	549054	549061	8 As	7 As	indel	1 bp deletion	intergenic: no annotation	-	-	BCM300_00543
24	581210		1 A	2 As	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
25	595619	595625	7 Cs	6 Cs	indel	1 bp deletion	3-ketoacyl-acyl carrier protein reductase (fabG)	fabG	hp0561	BCM300_00581
26	609222	609226	5 Gs	4 Gs	indel	1 bp deletion	methylene-tetrahydrofolate dehydrogenase (folD)	folD	hp0577	BCM300_00598
27	663718	663732	15 Cs	13 Cs	indel	2 bp deletion	predicted LPS biosynthesis protein	-	hp0619	BCM300_00638
28	695168	695174	7 Gs	6 Gs	indel	1 bp deletion	aspartate ammonia-lyase (aspA)	aspA	hp0649	BCM300_00670
29	705587	705595	9 Gs	8 Gs	indel	1 bp deletion	processing protease (ymxG)	ymxG	hp0657	BCM300_00678
30	780145	780159	15 Ts	14 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
31	782006	782009	TTCC	CTCT	intrachromosomal	4-bp fragment recombined	outer membrane protein (omp17)	sabA	hp0725	BCM300_00748
32	786911	786928	18 As	17 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
33	810228	810235	8 As	7 As	indel	1 bp deletion	cell division membrane protein (ftsX)	ftsX	hp0749	BCM300_00778
34	814668	814676	9 As	8 As	indel	1 bp deletion	flagellar hook-associated protein 2 (fliD)	fliD	hp0752	BCM300_00781
35	826877	826887	11 Cs	10 Cs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
36	862628	862631	4 Gs	5 Gs	indel	1 bp insertion	sigma-54 interacting protein	comM	hp0792	BCM300_00832
37	864414	864418	5 As	6 As	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00836
38	864639		1 C	2 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	intergenic: no annotation
39	864688	864691	4 Ts	5 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00837
40	864943	864944	2 Cs	3 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00838
41	865147	865148	2 Ts	3 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00839
42	867194	867200	7 Ts	6 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
43	867350		C	T	SNP	non-synonymous	molybdenum cofactor biosynthesis protein C (moaC)	moaC	hp0798	BCM300_00842
44	936941	936954	14 Ts	13 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
45	960192	960199	8 As	7 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
46	960612	960615	4 Cs	5 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
47	1081106	1081409	38 x 8-nt repeat	39 x 8-nt repeat	indel	8 bp insertion [repeat: CATTCTAT]	hypothetical protein	-	hp0427	BCM300_01049
48	1081414	1081415		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
49	1081419	1081420		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
50	1081421		1 A	2 As	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
51	1089108	1089112	5 Cs	4 Cs	indel	1 bp deletion	tRNA delta(2)-isopentenylpyrophosphate transfera	miaA	hp1415	BCM300_01058
52	1116806	1116812	7 As	6 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
53	1168491	1168652	3 x 54-nt repeat	2 x 54-nt repeat	indel	54 bp deletion [repeat: TCTTTGTGAGCCACTAATTGAGGGCTT AACTCAGGTTTTTTGGCGTCTTTTTA]	hypothetical protein	-	hp1192	BCM300_01130
54	1218049	1218061	13 Ts	12 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
55	1221905	1221912	8 Cs	7 Cs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
56	1222342	1222346	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
57	1256981	1256995	15 As	14 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
58	1287348	1287387	5 x 8-nt repeat	6 x 8-nt repeat	indel	8 bp insertion [repeat: GTATTTAT]	intergenic: no annotation	-	-	intergenic: no annotation
59	1296246	1296250	5 Gs	4 Gs	indel	1 bp deletion	cation efflux system protein (czcA)	czcA	hp1329	BCM300_01272
60	1304687	1304701	15 As	14 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
61	1317399	1317412	14 Gs	13 Gs	indel	1 bp deletion	putative adenine specific DNA methyltransferase	-	hp1354	BCM300_01297
62	1318224	1318234	11 Gs	10 Gs	indel	1 bp deletion	putative adenine specific DNA methyltransferase	-	hp1354	BCM300_01297
63	1347536	1347549	14 As	13 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
64	1370527	1370539	13 Gs	12 Gs	indel	1 bp deletion	predicted type III R-M system modification enzyme	-	hp1522	BCM300_01342
65	1503023	1503044	22 Ts	19 Ts	indel	3 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
66	1512216	1512222	7 Gs	6 Gs	indel	1 bp deletion	2-keto-3-deoxy-6-phosphogluconate aldolase (eda)	eda	hp1099	BCM300_01471
67	1539899	1539906	8 Cs	7 Cs	indel	1 bp deletion	SAM-dependent methyltransferase	hemK	hp0381	BCM300_01502
68	1578463	1578467	5 As	4 As	indel	1 bp deletion	methionyl-tRNA synthetase (metS)	metS	hp0417	BCM300_01539
69	1592191	1592246	8 x 7-nt repeat	7 x 7-nt repeat	indel	7 bp deletion [repeat: TTATTCT]	intergenic: no annotation	-	-	intergenic: no annotation
70	1628666	1628673	8 Ts	7 Ts	indel	1 bp deletion	lipoprotein	-	hp1546	BCM300_01585

HE147/09

No.	Position in BCM-300		Sequence in		Type of difference		Annotation[1]			BCM-300[2]
	Start	End	BCM-300	Reisolate			Gene description	Gene name	Locus tag	Locus tag
1	23007	23084	6 x 13-nt repeat	7 x 13-nt repeat	indel	13 bp insertion [repeat: AAGGTTTTTAATT]	intergenic: no annotation	-	-	intergenic: no annotation
2	32486	32490	5 Gs	6 Gs	indel	1 bp insertion	ATP-dependent dethiobiotin synthetase BioD	bioD	hp0029	BCM300_00035
3	71831	71835	5 Gs	6 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
4	99174	99185	12 Gs	11 Gs	indel	1 bp deletion	alpha-1,2-fucosyltransferase	futC	hp0094	BCM300_00103
5	169575	169588	14 Gs	13 Gs	indel	1 bp deletion	signal-transducing protein, histidine kinase	-	hp0164	BCM300_00169
6	220712	220718	7 Gs	6 Gs	indel	1 bp deletion	CDP-diglyceride synthetase (cdsA)	cdsA	hp0215	BCM300_00218
7	222870	222881	12 Gs	11 Gs	indel	1 bp deletion	beta-1,4-N-acetylgalactosamyltransferase	cgtA	hp0217	BCM300_00220
8	242241		G	T	SNP	non-synonymous	beta-lactamase HcpE	hcpE	hp0235	BCM300_00238
9	283039	283045	7 Cs	6 Cs	indel	1 bp deletion	guanosine pentaphosphate phosphohydrolase (gp	gppA	hp0278	BCM300_00281
10	340352	340356	5 Gs	4 Gs	indel	1 bp deletion	NH(3)-dependent NAD+ synthetase (nadE)	nadE	hp0329	BCM300_00331
11	366392	366395	4 Gs	3 Gs	indel	1 bp deletion	cell division protein (ftsK)	ftsK	hp1090	BCM300_00358
12	393772	393776	5 Gs	4 Gs	indel	1 bp deletion	hypothetical protein	-	hp1064	BCM300_00386
13	486962	486966	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	BCM300_00476
14	581568		C	T	SNP	non-synonymous	cag pathogenicity island protein (cag26)	cagA	hp0547	BCM300_00568
15	615059	615063	5 Gs	4 Gs	indel	1 bp deletion	periplasmic protein TonB	tonB1	hp0582	BCM300_00603
16	615076	615081	6 Gs	5 Gs	indel	1 bp deletion	periplasmic protein TonB	tonB1	hp0582	BCM300_00603
17	615086	615090	5 Gs	4 Gs	indel	1 bp deletion	periplasmic protein TonB	tonB1	hp0582	BCM300_00603
18	663718	663732	15 Cs	14 Cs	indel	1 bp deletion	predicted LPS biosynthesis protein	-	hp0619	BCM300_00638
19	727943	727947	5 Cs	6 Cs	indel	1 bp insertion	methylated-DNA--protein-cysteine methyltransfer	dat1	hp0676	BCM300_00697
20	785086	785089	CTCT	TTCC	intrachromosomal	4-bp fragment recombined	outer membrane protein (omp16)	sabB	hp0722	BCM300_00751
21	786553	786637			intrachromosomal	85-bp fragment recombined; source: 780431-780515	outer membrane protein (omp16)	sabB	hp0722	BCM300_00751
22	786776	786793	9 x GA repeat	8 x GA repeat	indel	2 bp deletion	outer membrane protein (omp16)	sabB	hp0722	BCM300_00751
23	862628	862631	4 Gs	5 Gs	indel	1 bp insertion	sigma-54 interacting protein	comM	hp0792	BCM300_00832
24	864414	864418	5 As	6 As	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00836
25	864639		1 C	2 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	intergenic: no annotation
26	864688	864691	4 Ts	5 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00837
27	864943	864944	2 Cs	3 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00838
28	865147	865148	2 Ts	3 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00839
29	875404	875408	5 Cs	4 Cs	indel	1 bp deletion	iron(III) dicitrate transport protein (fecA)	fecA	hp0807	BCM300_00851
30	934310		C	T	SNP	synonymous	hypothetical protein	-	hp0874	BCM300_00916
31	953010	953022	13 Gs	11 Gs	indel	2 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
32	960612	960615	4 Cs	5 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
33	979321	979335	15 Ts	14 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
34	1042698	1042715	9 x GA repeat	8 x GA repeat	indel	2 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
35	1081106	1081409	38 x 8-nt repeat	37 x 8-nt repeat	indel	8 bp deletion [repeat: CATTCTAT]	hypothetical protein	-	hp0427	BCM300_01049
36	1081414	1081415		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
37	1081419	1081420		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
38	1081421		1 A	2 As	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
39	1083785	1083789	5 Gs	4 Gs	indel	1 bp deletion	hypothetical protein	-	hp0425	BCM300_01052
40	1113391		A	G	SNP	intergenic	intergenic: no annotation	-	-	intergenic: no annotation
41	1198194	1198199	6 Cs	5 Cs	indel	1 bp deletion	phosphoribosylamine--glycine ligase	purD	hp1218	BCM300_01160
42	1218049	1218061	13 Ts	12 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
43	1222342	1222346	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
44	1317399	1317412	14 Gs	13 Gs	indel	1 bp deletion	putative adenine specific DNA methyltransferase	-	hp1354	BCM300_01297
45	1318224	1318234	11 Gs	12 Gs	indel	1 bp insertion	putative adenine specific DNA methyltransferase	-	hp1354	BCM300_01297
46	1370527	1370539	13 Gs	12 Gs	indel	1 bp deletion	predicted type III R-M system modification enzyme	-	hp1522	BCM300_01342
47	1408373	1408377	5 Gs	4 Gs	indel	1 bp deletion	phosphate permease	-	hp1491	BCM300_01372
48	1408833	1408836	4 Gs	3 Gs	indel	1 bp deletion	phosphate permease	-	hp1491	BCM300_01372
49	1409245	1409251	7 As	6 As	indel	1 bp deletion	phosphate permease	-	hp1491	BCM300_01372
50	1462751	1462934		-184 bp	indel	184 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
51	1488895		1 G	2 Gs	indel	1 bp insertion	Sel1 repeat-containing protein	hcpX	hp1117	BCM300_01451
52	1570250	1570256	7 Gs	6 Gs	indel	1 bp deletion	GMP synthase (guaA)	guaA	hp0409	BCM300_01530
53	1587896	1588151	32 x 8-nt repeat	34 x 8-nt repeat	indel	16 bp insertion [repeat: GAATGATA]	hypothetical protein	-	hp0427	BCM300_01551
54	1588129	1588130		ATG	indel	3 bp insertion	hypothetical protein	-	hp0427	BCM300_01551
55	1592310	1592315	6 Gs	7 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation

HE170/09

No.	Position in BCM-300		Sequence in		Type of difference		Annotation[1]			BCM-300[2]
	Start	End	BCM-300	Reisolate			Gene description	Gene name	Locus tag	Locus tag
1	32486	32490	5 Gs	6 Gs	indel	1 bp insertion	ATP-dependent dethiobiotin synthetase BioD	bioD	hp0029	BCM300_00035
2	71816	71821	6 Gs	5 Gs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
3	99174	99185	12 Gs	13 Gs	indel	1 bp insertion	alpha-1,2-fucosyltransferase	futC	hp0094	BCM300_00103
4	222870	222881	12 Gs	10 Gs	indel	2 bp deletion	beta-1,4-N-acetylgalactosamyltransferase	cgtA	hp0217	BCM300_00220
5	257813	257817	5 Cs	4 Cs	indel	1 bp deletion	oligopeptide ABC transporter, permease protein (c	oppC	hp0251	BCM300_00254
6	450369	450379	11 Gs	12 Gs	indel	1 bp insertion	type I restriction-modification system endonucleas	hsdR	hp0464	BCM300_00442
7	486962	486966	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	BCM300_00476
8	577318	577319		+2398 bp	indel	transposon insertion	cag pathogenicity island protein (cag23)	cagE	hp0544	BCM300_00564
9	615096	615100	5 Gs	4 Gs	indel	1 bp deletion	periplasmic protein TonB	tonB1	hp0582	BCM300_00603
10	727943	727947	5 Cs	6 Cs	indel	1 bp insertion	methylated-DNA--protein-cysteine methyltransfer	dat1	hp0676	BCM300_00697
11	804754	804757	4 Cs	3 Cs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
12	862628	862631	4 Gs	5 Gs	indel	1 bp insertion	sigma-54 interacting protein	comM	hp0792	BCM300_00832
13	864414	864418	5 As	6 As	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00836
14	864639		1 C	2 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	intergenic: no annotation
15	864688	864691	4 Ts	5 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00837
16	864943	864944	2 Cs	3 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00838
17	865147	865148	2 Ts	3 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00839
18	960612	960615	4 Cs	5 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
19	960782	960786	5 Cs	4 Cs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
20	1081106	1081409	38 x 8-nt repeat	40 x 8-nt repeat	indel	16 bp insertion [repeat: CATTCTAT]	hypothetical protein	-	hp0427	BCM300_01049
21	1081414	1081415		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
22	1081419	1081420		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
23	1081421		1 A	2 As	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
24	1221905	1221912	8 Cs	7 Cs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
25	1222342	1222346	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
26	1318224	1318234	11 Gs	12 Gs	indel	1 bp insertion	putative adenine specific DNA methyltransferase	-	hp1354	BCM300_01297
27	1340881		G	A	SNP	intergenic	intergenic: no annotation	-	-	intergenic: no annotation
28	1347536	1347549	14 As	13 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
29	1385328	1385332	5 Cs	4 Cs	indel	1 bp deletion	iron-regulated outer membrane protein (frpB)	frpB	hp1512	BCM300_01350
30	1503023	1503044	22 Ts	21 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
31	1587896	1588151	32 x 8-nt repeat	34 x 8-nt repeat	indel	16 bp insertion [repeat: GAATGATA]	hypothetical protein	-	hp0427	BCM300_01551
32	1592310	1592315	6 Gs	7 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
33	1594330	1594344	15 Cs	16 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation

HE171/09

No.	Position in BCM-300		Sequence in		Type of difference		Annotation[1]			BCM-300[2]
	Start	End	BCM-300	Reisolate			Gene description	Gene name	Locus tag	Locus tag
1	9816	9829	14 Ts	15 Ts	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
2	32486	32490	5 Gs	6 Gs	indel	1 bp insertion	ATP-dependent dethiobiotin synthetase BioD	bioD	hp0029	BCM300_00035
3	71831	71835	5 Gs	6 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
4	99174	99185	12 Gs	13 Gs	indel	1 bp insertion	alpha-1,2-fucosyltransferase	futC	hp0094	BCM300_00103
5	169575	169588	14 Gs	12 Gs	indel	2 bp deletion	signal-transducing protein, histidine kinase	-	hp0164	BCM300_00169
6	222870	222881	12 Gs	10 Gs	indel	2 bp deletion	beta-1,4-N-acetylgalactosamyltransferase	cgtA	hp0217	BCM300_00220
7	259244		G	A	SNP	non-synonymous	outer membrane protein (omp7)	hopF	hp0252	BCM300_00256
8	350265	350279	15 As	16 As	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
9	450369	450379	11 Gs	12 Gs	indel	1 bp insertion	type I restriction-modification system endonucleas	hsdR	hp0464	BCM300_00442
10	486962	486966	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	BCM300_00476
11	662731	663008		-278 bp	indel	278 bp deletion	intergenic: no annotation	-	-	BCM300_00637
12	663718	663732	15 Cs	13 Cs	indel	2 bp deletion	predicted LPS biosynthesis protein	-	hp0619	BCM300_00638
13	727943	727947	5 Cs	6 Cs	indel	1 bp insertion	methylated-DNA--protein-cysteine methyltransfer	dat1	hp0676	BCM300_00697
14	781175	781235			intrachromosomal	61-bp fragment recombined; source: c785860-785920 (sabB)	outer membrane protein (omp17)	sabA	hp0725	BCM300_00748
15	782006	782009	TTCC	CTCT	intrachromosomal	4-bp fragment recombined	outer membrane protein (omp17)	sabA	hp0725	BCM300_00748
16	862628	862631	4 Gs	5 Gs	indel	1 bp insertion	sigma-54 interacting protein	comM	hp0792	BCM300_00832
17	864414	864418	5 As	6 As	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00836
18	864639		1 C	2 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	intergenic: no annotation
19	864688	864691	4 Ts	5 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00837
20	864943	864944	2 Cs	3 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00838
21	865147	865148	2 Ts	3 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00839
22	960612	960615	4 Cs	5 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
23	979321	979335	15 Ts	16 Ts	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
24	1081106	1081409	38 x 8-nt repeat	37 x 8-nt repeat	indel	8 bp deletion [repeat: CATTCTAT]	hypothetical protein	-	hp0427	BCM300_01049
25	1081414	1081415		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
26	1081419	1081420		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
27	1081421		1 A	2 As	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
28	1222342	1222346	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
29	1224971	1224985	15 Ts	16 Ts	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
30	1256981	1256995	15 As	14 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
31	1370527	1370539	13 Gs	12 Gs	indel	1 bp deletion	predicted type III R-M system modification enzyme	-	hp1522	BCM300_01342
32	1503023	1503044	22 Ts	20 Ts	indel	2 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
33	1587896	1588151	32 x 8-nt repeat	34 x 8-nt repeat	indel	16 bp insertion [repeat: GAATGATA]	hypothetical protein	-	hp0427	BCM300_01551
34	1592310	1592315	6 Gs	7 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation

HE178/09

No.	Position in BCM-300		Sequence in		Type of difference		Annotation[1]			BCM-300[2]
	Start	End	BCM-300	Reisolate			Gene description	Gene name	Locus tag	Locus tag
1	32486	32490	5 Gs	6 Gs	indel	1 bp insertion	ATP-dependent dethiobiotin synthetase BioD	bioD	hp0029	BCM300_00035
2	71831	71835	5 Gs	6 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
3	75512	75516	5 As	4 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
4	99174	99185	12 Gs	11 Gs	indel	1 bp deletion	alpha-1,2-fucosyltransferase	futC	hp0094	BCM300_00103
5	110540	110553	14 Cs	13 Cs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
6	169575	169588	14 Gs	13 Gs	indel	1 bp deletion	signal-transducing protein, histidine kinase	-	hp0164	BCM300_00169
7	210809	210824	8 x CT repeat	9 x CT repeat	indel	2 bp insertion	lipopolysaccharide 1,2-glucosyltransferase (rfaJ)	rfaJ	hp0159	BCM300_00211
8	210990	211003	14 Ts	15 Ts	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
9	222870	222881	12 Gs	11 Gs	indel	1 bp deletion	beta-1,4-N-acetylgalactosamyltransferase	cgtA	hp0217	BCM300_00220
10	255365	255368	4 Cs	3 Cs	indel	1 bp deletion	hypothetical protein	-	hp0249	BCM300_00252
11	450369	450379	11 Gs	12 Gs	indel	1 bp insertion	type I restriction-modification system endonucleas	hsdR	hp0464	BCM300_00442
12	479548	479552	5 As	4 As	indel	1 bp deletion	hypothetical protein	-	hp0990	BCM300_00468
13	482354	482358	5 Ts	4 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	BCM300_00471
14	486962	486966	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	BCM300_00476
15	542103	542111	9 As	8 As	indel	1 bp deletion	hypothetical protein	-	hp0513	BCM300_00535
16	563321		C	T	SNP	non-synonymous	cag pathogenicity island protein (cag9)	cagW	hp0529	BCM300_00551
17	580121	580128	8 As	7 As	indel	1 bp deletion	cag pathogenicity island protein (cag24)	cagD	hp0545	BCM300_00565
18	727943	727947	5 Cs	6 Cs	indel	1 bp insertion	methylated-DNA--protein-cysteine methyltransfer	dat1	hp0676	BCM300_00697
19	780276	780291	8 x CT repeat	9 x CT repeat	indel	2 bp insertion	outer membrane protein (omp17)	sabA	hp0725	BCM300_00748
20	862628	862631	4 Gs	5 Gs	indel	1 bp insertion	sigma-54 interacting protein	comM	hp0792	BCM300_00832
21	864414	864418	5 As	6 As	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00836
22	864639		1 C	2 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	intergenic: no annotation
23	864688	864691	4 Ts	5 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00837
24	864943	864944	2 Cs	3 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00838
25	865147	865148	2 Ts	3 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00839
26	949904		C	T	SNP	non-synonymous	vacuolating cytotoxin	vacA	hp0887	BCM300_00927
27	960192	960199	8 As	7 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
28	960208	960215	8 As	7 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
29	960224	960231	8 As	7 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
30	960612	960615	4 Cs	5 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
31	985093	985106	14 As	13 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
32	1049275	1049282	8 Ts	7 Ts	indel	1 bp deletion	glycyl-tRNA synthetase, beta subunit (glyS)	glyS	hp0972	BCM300_01019
33	1081106	1081409	38 x 8-nt repeat	39 x 8-nt repeat	indel	8 bp insertion [repeat: CATTCTAT]	hypothetical protein	-	hp0427	BCM300_01049
34	1081414	1081415		A	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
35	1081419	1081420		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
36	1081421		1 A	2 As	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
37	1081422	1081425	TCAT	GCAA	intrachromosomal	4-bp fragment recombined	hypothetical protein	-	hp0427	BCM300_01049
38	1222342	1222346	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
39	1257481	1257487	7 As	6 As	indel	1 bp deletion	intergenic: no annotation	-	-	BCM300_01220
40	1304687	1304701	15 As	14 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
41	1309317	1309324	8 Cs	7 Cs	indel	1 bp deletion	phosphoglycerate kinase	pgk	hp1345	BCM300_01287
42	1318224	1318234	11 Gs	12 Gs	indel	1 bp insertion	putative adenine specific DNA methyltransferase	-	hp1354	BCM300_01297
43	1347536	1347549	14 As	15 As	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
44	1468210	1468215	6 Ts	5 Ts	indel	1 bp deletion	methionyl-tRNA formyltransferase (fmt)	fmt	hp1141	BCM300_01426
45	1502614	1502619	6 Cs	5 Cs	indel	1 bp deletion	hypothetical protein	-	hp1106	BCM300_01463
46	1503023	1503044	22 Ts	21 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
47	1504362	1504373	12 Ts	13 Ts	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
48	1544143	1544150	8 As	7 As	indel	1 bp deletion	primosomal protein replication factor (priA)	priA	hp0387	BCM300_01508
49	1592310	1592315	6 Gs	7 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation

HE93/10

No.	Position in BCM-300		Sequence in		Type of difference		Annotation[1]			BCM-300[2]
	Start	End	BCM-300	Reisolate			Gene description	Gene name	Locus tag	Locus tag
1	23007	23084	6 x 13-nt repeat	5 x 13-nt repeat	indel	13 bp deletion [repeat: AAGGTTTTTAATT]	intergenic: no annotation	-	-	intergenic: no annotation
2	28076	28091	16 As	15 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
3	32486	32490	5 Gs	6 Gs	indel	1 bp insertion	ATP-dependent dethiobiotin synthetase BioD	bioD	hp0029	BCM300_00035
4	71831	71835	5 Gs	6 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
5	108138	108144	7 Cs	8 Cs	indel	1 bp insertion	glycosyltransferase	-	hp0102	BCM300_00111
6	110540	110553	14 Cs	13 Cs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
7	169575	169588	14 Gs	12 Gs	indel	2 bp deletion	signal-transducing protein, histidine kinase	-	hp0164	BCM300_00169
8	222870	222881	12 Gs	11 Gs	indel	1 bp deletion	beta-1,4-N-acetylgalactosamyltransferase	cgtA	hp0217	BCM300_00220
9	258783		G	A	SNP	non-synonymous	outer membrane protein (omp7)	hopF	hp0252	BCM300_00256
10	301606		C	T	SNP	synonymous	toxin-like outer membrane protein	-	hp0289	BCM300_00292
11	450369	450379	11 Gs	12 Gs	indel	1 bp insertion	type I restriction-modification system endonucleas	hsdR	hp0464	BCM300_00442
12	486962	486966	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	BCM300_00476
13	526366	526372	7 Gs	8 Gs	indel	1 bp insertion	phospholipase A1 precursor (DR-phospholipase A)	pldA	hp0499	BCM300_00519
14	580805	585875	5071 bp	+19950 bp	indel	insertion in cagPAI				
15	727943	727947	5 Cs	6 Cs	indel	1 bp insertion	methylated-DNA--protein-cysteine methyltransfera	dat1	hp0676	BCM300_00697
16	780145	780159	15 Ts	14 Ts	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
17	786911	786928	18 As	17 As	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
18	826877	826887	11 Cs	10 Cs	indel	1 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
19	857371		C	T	SNP	non-synonymous	hypothetical protein	-	hp1005	BCM300_00828
20	862628	862631	4 Gs	5 Gs	indel	1 bp insertion	sigma-54 interacting protein	comM	hp0792	BCM300_00832
21	864414	864418	5 As	6 As	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00836
22	864639		1 C	2 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	intergenic: no annotation
23	864688	864691	4 Ts	5 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00837
24	864943	864944	2 Cs	3 Cs	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00838
25	865147	865148	2 Ts	3 Ts	indel	1 bp insertion	trigger factor (tig)	tig	hp0795	BCM300_00839
26	940491	940498	ATATTTTC		indel	8 bp deletion	hypothetical protein	-	hp0879	BCM300_00920
27	949223		C	T	SNP	non-synonymous	vacuolating cytotoxin	vacA	hp0887	BCM300_00927
28	953010	953022	13 Gs	14 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
29	960612	960615	4 Cs	5 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
30	979321	979335	15 Ts	13 Ts	indel	2 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
31	1081106	1081409	38 x 8-nt repeat	39 x 8-nt repeat	indel	8 bp insertion [repeat: CATTCTAT]	hypothetical protein	-	hp0427	BCM300_01049
32	1081414	1081415		A	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
33	1081419	1081420		T	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
34	1081421		1 A	2 As	indel	1 bp insertion	hypothetical protein	-	hp0427	BCM300_01049
35	1081422	1081425	TCAT	GCAA	intrachromosomal	4-bp fragment recombined	hypothetical protein	-	hp0427	BCM300_01049
36	1218049	1218061	13 Ts	14 Ts	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
37	1222342	1222346	5 Cs	6 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
38	1224868	1224885	9 x GA repeat	10 x GA repeat	indel	2 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
39	1224971	1224985	15 Ts	13 Ts	indel	2 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
40	1225114	1225253	14 x 10-nt repeat	15 x 10-nt repeat	indel	10 bp insertion [repeat: AAGATTAAAC]	intergenic: no annotation	-	-	intergenic: no annotation
41	1293053	1293059	7 Gs	8 Gs	indel	1 bp insertion	predicted cobalt-zinc-cadmium resistance protein/	czcB	hp1328	BCM300_01271
42	1318224	1318234	11 Gs	12 Gs	indel	1 bp insertion	putative adenine specific DNA methyltransferase	-	hp1354	BCM300_01297
43	1370527	1370539	13 Gs	14 Gs	indel	1 bp insertion	predicted type III R-M system modification enzyme	-	hp1522	BCM300_01342
44	1488899	1488901	3 Gs	4 Gs	indel	1 bp insertion	Sel1 repeat-containing protein	hcpX	hp1117	BCM300_01451
45	1503023	1503044	22 Ts	20 Ts	indel	2 bp deletion	intergenic: no annotation	-	-	intergenic: no annotation
46	1587943	1587944		+26 bp	indel	26 bp insertion	hypothetical protein	-	hp0427	BCM300_01551
47	1592310	1592315	6 Gs	7 Gs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
48	1594330	1594344	15 Cs	16 Cs	indel	1 bp insertion	intergenic: no annotation	-	-	intergenic: no annotation
49	1600696	1600702	7 Cs	8 Cs	indel	1 bp insertion	prephenate dehydrogenase (tyrA)	tyrA	hp1380	BCM300_01560

Supplementary Dataset 2: Percentages of methyltransferase recognition sequences detected as methylated by SMRT sequencing

Motif	Modification type	% of motifs detected																
		BCM-300	HE93/10	HE101/09	HE132/09	HE134/09	HE136/09	HE141/09	HE142/09	HE143/09	HE147/09	HE170/09	HE171/09	HE178/09	Min	Max	mean	SD
AGGAG	m6A	99.3	99.3	98.9	99.7	99.9	99.3	99.5	98.8	99.5	99.3	99.4	99.8	99.8	98.8	99.9	99.4	0.3
CACAT	m6A	98.0	98.7	97.9	99.4	100.0	98.8	98.7	97.8	98.5	98.7	98.0	99.8	100.0	97.8	100.0	98.8	0.7
CATG	m6A	99.9	99.9	99.9	100.0	100.0	100.0	100.0	99.8	100.0	100.0	99.9	100.0	100.0	99.8	100.0	99.9	0.1
CCATC	m6A	99.6	99.8	99.6	99.8	100.0	99.9	99.8	99.1	99.8	99.7	99.5	99.9	100.0	99.1	100.0	99.7	0.2
CCGG	m4C	72.3	92.0	72.1	88.2	99.9	85.3	87.8	62.2	88.1	88.0	72.9	93.2	99.9	62.2	99.9	84.8	11.1
CCTYNA[1]	m6A	3.1	2.4	3.1	3.3	3.3	3.2	3.2	3.1	89.5	2.4	3.1	2.8	3.2	2.4	89.5	9.7	23.1
GAATTC	m6A	99.1	100.0	98.9	99.4	100.0	99.7	99.1	98.6	99.1	99.1	99.1	100.0	100.0	98.6	100.0	99.4	0.5
GACY[2]	m6A	99.4	4.0	98.9	99.7	100.0	99.5	99.7	98.7	99.7	3.8	99.3	4.1	100.0	3.8	100.0	77.4	40.3
GAGG	m6A	98.6	98.9	98.4	99.3	99.6	98.9	99.2	98.2	99.3	99.4	98.9	99.8	99.5	98.2	99.8	99.1	0.5
GATC	m6A	99.4	99.8	99.3	99.9	100.0	99.8	99.9	99.1	99.7	99.8	99.5	100.0	100.0	99.1	100.0	99.7	0.3
GATGG	m6A	99.9	100.0	99.8	100.0	100.0	100.0	100.0	99.5	100.0	100.0	99.9	100.0	100.0	99.5	100.0	99.9	0.1
GCGC[3]	m5C	2.3	3.1	1.3	3.7	9.0	3.0	3.7	1.7	3.9	3.2	3.0	6.4	7.7	1.3	9.0	4.0	2.2
GTSAC	m6A	99.6	100.0	99.6	100.0	100.0	99.6	100.0	98.0	99.6	100.0	100.0	100.0	100.0	98.0	100.0	99.7	0.5
TCNNGA	m6A	98.3	98.7	98.3	99.3	99.9	99.0	99.0	97.9	98.9	98.9	98.0	99.6	99.9	97.9	99.9	98.9	0.6
TGCA	m6A	99.1	99.5	99.1	99.7	100.0	99.6	99.5	98.9	99.4	99.4	99.1	99.8	100.0	98.9	100.0	99.5	0.3

[1]CCTYNA only methylated in HE143/09

[2]GACY not methylated in HE93/10, HE147/09 and HE171/09

[3]GCGC not reliably detected

BCM-300

Motif	Modified position	Modification type	No. of motifs detected	No. of motifs in genome	% of motifs detected	Mean Score	Mean coverage
AGGAG	4	m6A	1504	1515	99.27	78.42	53.77
CACAT	4	m6A	2390	2438	98.03	76.74	54.78
CATG	2	m6A	15300	15320	99.87	82.19	55.06
CCATC	3	m6A	2106	2115	99.57	76.29	54.86
CCGG	1	m4C	2566	3550	72.28	44.46	54.94
CCTYNA	6	m6A	261	8388	3.11	77.50	54.92
GAATTC	3	m6A	349	352	99.15	78.11	53.61
GACY	2	m6A	8180	8229	99.40	75.53	54.75
GAGG	2	m6A	4985	5055	98.62	73.69	53.00
GATC	2	m6A	10930	10994	99.42	75.57	54.21
GATGG	2	m6A	2113	2115	99.91	78.56	54.78
GCGC	2	m5C	306	13046	2.35	41.22	56.27
GTSAC	4	m6A	249	250	99.60	77.49	54.73
TCNNGA	6	m6A	3919	3986	98.32	76.37	54.73
TGCA	4	m6A	10371	10466	99.09	77.66	55.01

HE93/10

Motif	Modified position	Modification type	No. of motifs detected	No. of motifs in genome	% of motifs detected	Mean Score	Mean coverage
AGGAG	4	m6A	1517	1527	99.35	121.20	86.38
CACAT	4	m6A	2433	2465	98.70	117.24	86.30
CATG	2	m6A	15393	15404	99.93	127.55	87.46
CCATC	3	m6A	2123	2127	99.81	115.85	86.13
CCGG	1	m4C	3275	3558	92.05	62.84	86.02
CCTYNA	6	m6A	199	8461	2.35	120.11	87.34
GAATTC	3	m6A	364	364	100.00	119.62	84.82
GACY	2	m6A	331	8310	3.98	111.71	89.69
GAGG	2	m6A	5031	5088	98.88	111.90	85.56
GATC	2	m6A	11142	11166	99.79	115.57	86.54
GATGG	2	m6A	2126	2127	99.95	122.23	87.75
GCGC	2	m5C	401	13092	3.06	45.17	88.29
GTSAC	4	m6A	256	256	100.00	122.34	88.48
TCNNGA	6	m6A	4007	4060	98.69	117.20	88.28
TGCA	4	m6A	10471	10528	99.46	119.34	87.38

HE101/09

Motif	Modified position	Modification type	No. of motifs detected	No. of motifs in genome	% of motifs detected	Mean Score	Mean coverage
AGGAG	4	m6A	1498	1515	98.88	75.11	48.21
CACAT	4	m6A	2388	2438	97.95	73.51	48.41
CATG	2	m6A	15302	15320	99.88	80.27	50.38
CCATC	3	m6A	2107	2115	99.62	73.15	48.60
CCGG	1	m4C	2561	3552	72.10	44.84	51.56
CCTYNA	6	m6A	260	8388	3.10	75.47	49.83
GAATTC	3	m6A	348	352	98.86	71.66	45.17
GACY	2	m6A	8140	8229	98.92	72.40	49.16
GAGG	2	m6A	4976	5055	98.44	72.19	49.31
GATC	2	m6A	10913	10992	99.28	72.70	48.80
GATGG	2	m6A	2111	2115	99.81	76.36	50.35
GCGC	2	m5C	176	13046	1.35	39.99	53.16
GTSAC	4	m6A	249	250	99.60	72.96	48.47
TCNNGA	6	m6A	3920	3986	98.34	74.09	49.81
TGCA	4	m6A	10374	10464	99.14	75.55	50.26

HE132/09

Motif	Modified position	Modification type	No. of motifs detected	No. of motifs in genome	% of motifs detected	Mean Score	Mean coverage
AGGAG	4	m6A	1511	1516	99.67	127.75	89.92
CACAT	4	m6A	2423	2438	99.38	124.72	91.78
CATG	2	m6A	15318	15320	99.99	135.48	92.32
CCATC	3	m6A	2111	2115	99.81	123.47	90.68
CCGG	1	m4C	3133	3552	88.20	66.07	91.51
CCTYNA	6	m6A	279	8387	3.33	118.06	92.90
GAATTC	3	m6A	352	354	99.44	124.84	87.64
GACY	2	m6A	8200	8227	99.67	120.65	92.06
GAGG	2	m6A	5018	5055	99.27	119.76	90.07
GATC	2	m6A	10986	10992	99.95	122.88	91.41
GATGG	2	m6A	2115	2115	100.00	128.51	91.62
GCGC	2	m5C	478	13044	3.66	46.68	92.19
GTSAC	4	m6A	250	250	100.00	124.60	90.81
TCNNGA	6	m6A	3959	3988	99.27	123.87	92.03
TGCA	4	m6A	10432	10468	99.66	127.53	92.13

HE134/09

Motif	Modified position	Modification type	No. of motifs detected	No. of motifs in genome	% of motifs detected	Mean Score	Mean coverage
AGGAG	4	m6A	1513	1515	99.87	128.05	93.01
CACAT	4	m6A	2437	2438	99.96	136.09	93.45
CATG	2	m6A	15320	15320	100.00	137.92	93.46
CCATC	3	m6A	2115	2115	100.00	139.85	92.94
CCGG	1	m4C	3547	3552	99.86	86.19	94.19
CCTYNA	6	m6A	278	8388	3.31	133.42	95.25
GAATTC	3	m6A	352	352	100.00	132.54	88.46
GACY	2	m6A	8229	8229	100.00	130.49	93.64
GAGG	2	m6A	5035	5055	99.60	118.91	93.09
GATC	2	m6A	10994	10994	100.00	142.55	93.48
GATGG	2	m6A	2115	2115	100.00	134.67	93.28
GCGC	2	m5C	1179	13046	9.04	52.29	95.79
GTSAC	4	m6A	250	250	100.00	132.24	94.31
TCNNGA	6	m6A	3984	3986	99.95	129.54	93.33
TGCA	4	m6A	10465	10468	99.97	133.69	93.55

HE136/09

Motif	Modified position	Modification type	No. of motifs detected	No. of motifs in genome	% of motifs detected	Mean Score	Mean coverage
AGGAG	4	m6A	1505	1515	99.34	101.94	70.93
CACAT	4	m6A	2409	2438	98.81	100.16	71.89
CATG	2	m6A	15318	15320	99.99	107.48	72.05
CCATC	3	m6A	2112	2115	99.86	99.30	71.63
CCGG	1	m4C	3030	3552	85.30	54.37	71.94
CCTYNA	6	m6A	265	8388	3.16	98.39	71.56
GAATTC	3	m6A	351	352	99.72	101.95	71.10
GACY	2	m6A	8187	8229	99.49	97.15	71.81
GAGG	2	m6A	4998	5055	98.87	96.11	70.61
GATC	2	m6A	10970	10994	99.78	98.54	71.52
GATGG	2	m6A	2115	2115	100.00	103.33	71.98
GCGC	2	m5C	386	13044	2.96	43.34	74.42
GTSAC	4	m6A	249	250	99.60	99.57	69.86
TCNNGA	6	m6A	3946	3986	99.00	99.78	72.52
TGCA	4	m6A	10424	10466	99.60	101.36	72.32

HE141/09

Motif	Modified position	Modification type	No. of motifs detected	No. of motifs in genome	% of motifs detected	Mean Score	Mean coverage
AGGAG	4	m6A	1507	1515	99.47	107.78	75.10
CACAT	4	m6A	2409	2441	98.69	104.67	76.18
CATG	2	m6A	15333	15334	99.99	114.42	77.19
CCATC	3	m6A	2116	2120	99.81	103.38	75.35
CCGG	1	m4C	3120	3552	87.84	59.07	77.18
CCTYNA	6	m6A	267	8399	3.18	104.51	77.31
GAATTC	3	m6A	349	352	99.15	104.18	72.86
GACY	2	m6A	8216	8243	99.67	102.33	76.69
GAGG	2	m6A	5021	5060	99.23	103.01	75.88
GATC	2	m6A	10989	11000	99.90	102.99	75.97
GATGG	2	m6A	2120	2120	100.00	110.48	77.13
GCGC	2	m5C	484	13058	3.71	45.70	77.20
GTSAC	4	m6A	250	250	100.00	105.87	76.13
TCNNGA	6	m6A	3943	3984	98.97	104.85	77.31
TGCA	4	m6A	10440	10488	99.54	106.33	76.71

HE142/09

Motif	Modified position	Modification type	No. of motifs detected	No. of motifs in genome	% of motifs detected	Mean Score	Mean coverage
AGGAG	4	m6A	1497	1515	98.81	69.10	44.03
CACAT	4	m6A	2385	2438	97.83	68.50	44.62
CATG	2	m6A	15288	15320	99.79	72.47	45.46
CCATC	3	m6A	2096	2115	99.10	67.28	44.23
CCGG	1	m4C	2208	3552	62.16	41.98	47.26
CCTYNA	6	m6A	264	8388	3.15	68.14	44.87
GAATTC	3	m6A	347	352	98.58	66.65	41.76
GACY	2	m6A	8126	8229	98.75	66.80	44.72
GAGG	2	m6A	4963	5055	98.18	65.88	44.22
GATC	2	m6A	10889	10992	99.06	67.61	44.81
GATGG	2	m6A	2105	2115	99.53	68.64	45.19
GCGC	2	m5C	216	13048	1.66	39.28	47.86
GTSAC	4	m6A	245	250	98.00	66.36	43.14
TCNNGA	6	m6A	3904	3986	97.94	68.14	45.27
TGCA	4	m6A	10355	10466	98.94	69.27	45.27

HE143/09

Motif	Modified position	Modification type	No. of motifs detected	No. of motifs in genome	% of motifs detected	Mean Score	Mean coverage
AGGAG	4	m6A	1507	1515	99.47	102.59	71.28
CACAT	4	m6A	2402	2438	98.52	98.83	71.58
CATG	2	m6A	15315	15322	99.95	109.13	73.52
CCATC	3	m6A	2110	2115	99.76	97.92	71.97
CCGG	1	m4C	3130	3552	88.12	58.24	72.30
CCTYNA	6	m6A	7510	8389	89.52	90.42	72.86
GAATTC	3	m6A	349	352	99.15	97.58	67.57
GACY	2	m6A	8202	8229	99.67	96.67	72.43
GAGG	2	m6A	5019	5056	99.27	97.81	71.94
GATC	2	m6A	10966	10994	99.75	97.25	71.76
GATGG	2	m6A	2115	2115	100.00	105.23	73.76
GCGC	2	m5C	503	13046	3.86	44.73	71.97
GTSAC	4	m6A	249	250	99.60	98.75	70.13
TCNNGA	6	m6A	3939	3984	98.87	98.83	72.38
TGCA	4	m6A	10404	10466	99.41	101.30	73.45

HE147/09

Motif	Modified position	Modification type	No. of motifs detected	No. of motifs in genome	% of motifs detected	Mean Score	Mean coverage
AGGAG	4	m6A	1505	1515	99.34	91.88	62.33
CACAT	4	m6A	2406	2438	98.69	88.15	62.72
CATG	2	m6A	15315	15318	99.98	96.33	63.79
CCATC	3	m6A	2109	2115	99.72	87.98	62.46
CCGG	1	m4C	3126	3552	88.01	51.84	63.48
CCTYNA	6	m6A	203	8388	2.42	91.51	63.60
GAATTC	3	m6A	349	352	99.15	88.81	58.95
GACY	2	m6A	312	8230	3.79	88.34	62.81
GAGG	2	m6A	5023	5054	99.39	86.26	62.01
GATC	2	m6A	10968	10994	99.76	87.65	62.70
GATGG	2	m6A	2115	2115	100.00	92.45	63.00
GCGC	2	m5C	414	13046	3.17	42.62	64.11
GTSAC	4	m6A	250	250	100.00	90.14	62.63
TCNNGA	6	m6A	3943	3986	98.92	88.53	63.05
TGCA	4	m6A	10407	10466	99.44	89.95	63.52

HE170/09

Motif	Modified position	Modification type	No. of motifs detected	No. of motifs in genome	% of motifs detected	Mean Score	Mean coverage
AGGAG	4	m6A	1506	1515	99.41	78.42	53.33
CACAT	4	m6A	2392	2441	97.99	76.53	54.87
CATG	2	m6A	15322	15334	99.92	81.34	54.71
CCATC	3	m6A	2109	2120	99.48	76.42	54.66
CCGG	1	m4C	2591	3552	72.94	44.96	55.38
CCTYNA	6	m6A	264	8399	3.14	75.43	54.71
GAATTC	3	m6A	349	352	99.15	76.91	52.86
GACY	2	m6A	8183	8243	99.27	75.65	54.46
GAGG	2	m6A	5006	5060	98.93	73.63	52.99
GATC	2	m6A	10947	11000	99.52	74.94	54.11
GATGG	2	m6A	2118	2120	99.91	78.47	54.90
GCGC	2	m5C	394	13060	3.02	42.16	54.50
GTSAC	4	m6A	250	250	100.00	76.80	53.29
TCNNGA	6	m6A	3905	3986	97.97	75.52	54.65
TGCA	4	m6A	10389	10488	99.06	76.66	54.52

HE171/09

Motif	Modified position	Modification type	No. of motifs detected	No. of motifs in genome	% of motifs detected	Mean Score	Mean coverage
AGGAG	4	m6A	1510	1513	99.80	144.06	98.33
CACAT	4	m6A	2433	2438	99.79	143.53	100.20
CATG	2	m6A	15320	15320	100.00	156.29	102.16
CCATC	3	m6A	2113	2115	99.91	143.32	100.18
CCGG	1	m4C	3309	3552	93.16	79.43	100.42
CCTYNA	6	m6A	238	8387	2.84	127.51	102.37
GAATTC	3	m6A	350	350	100.00	144.07	96.17
GACY	2	m6A	334	8228	4.06	131.60	101.18
GAGG	2	m6A	5044	5054	99.80	138.28	99.16
GATC	2	m6A	10993	10994	99.99	141.45	100.54
GATGG	2	m6A	2115	2115	100.00	148.45	101.25
GCGC	2	m5C	836	13046	6.41	49.80	100.33
GTSAC	4	m6A	250	250	100.00	146.54	100.69
TCNNGA	6	m6A	3972	3986	99.65	142.17	101.31
TGCA	4	m6A	10450	10466	99.85	146.06	101.16

HE178/09

Motif	Modified position	Modification type	No. of motifs detected	No. of motifs in genome	% of motifs detected	Mean Score	Mean coverage
AGGAG	4	m6A	1512	1515	99.80	112.60	83.42
CACAT	4	m6A	2437	2438	99.96	120.87	82.99
CATG	2	m6A	15320	15320	100.00	122.22	83.38
CCATC	3	m6A	2115	2115	100.00	126.46	83.66
CCGG	1	m4C	3547	3552	99.86	79.26	85.24
CCTYNA	6	m6A	269	8388	3.21	119.43	84.49
GAATTC	3	m6A	352	352	100.00	118.20	79.53
GACY	2	m6A	8225	8228	99.96	114.79	83.87
GAGG	2	m6A	5028	5055	99.47	104.19	83.55
GATC	2	m6A	10994	10994	100.00	128.41	83.91
GATGG	2	m6A	2115	2115	100.00	119.72	83.56
GCGC	2	m5C	1009	13046	7.73	51.04	86.19
GTSAC	4	m6A	250	250	100.00	117.48	85.38
TCNNGA	6	m6A	3983	3986	99.92	115.11	84.08
TGCA	4	m6A	10462	10466	99.96	118.71	84.08

5.2. Manuscript II

The core genome ^{m5}C methyltransferase JHP1050 (M.Hpy99III) plays an important role in orchestrating gene expression in *Helicobacter pylori*

Iratxe Estibariz, Annemarie Overmann, Florent Ailloud, Juliane Krebes, Christine Josenhans, and Sebastian Suerbaum

Published in:

Nucleic Acids Research 2019; Mar 18; 47(5):2336-2348 (doi: 10.1093/nar/gky1307)

Short summary

In this article, we functionally characterized the role of a highly conserved ^{m5}C-MTase in *H. pylori* present and putatively active in all strains. We studied the impact of this MTase on the transcriptomes of two *H. pylori* wild-type and respective MTase mutant strains, and showed that the MTase had both strain-specific and conserved effects. Methylation of motifs overlapping promoter sequences had a direct impact on gene expression. Methylation affected several phenotypic traits like adhesion to host cells, natural competence, copper resistance, and bacterial morphology.

Nucleic Acids Research, 2019 **1**
doi: 10.1093/nar/gky1307

The core genome ^{m5}C methyltransferase JHP1050 (M.Hpy99III) plays an important role in orchestrating gene expression in *Helicobacter pylori*

Iratxe Estibariz[1,2,3], Annemarie Overmann[1], Florent Ailloud[1,2,3], Juliane Krebes[2], Christine Josenhans[1,2,3,*] and Sebastian Suerbaum[1,2,3,*]

[1]Medical Microbiology and Hospital Epidemiology, Max von Pettenkofer Institute, Faculty of Medicine, LMU Munich, München, Germany, [2]Institute of Medical Microbiology and Hospital Epidemiology, Hannover Medical School, Hannover, Germany and [3]German Center for Infection Research (DZIF), Munich Site, Munich, Germany

Received August 15, 2018; Revised December 18, 2018; Editorial Decision December 19, 2018; Accepted December 21, 2018

ABSTRACT

Helicobacter pylori **encodes a large number of restriction–modification (R–M) systems despite its small genome. R–M systems have been described as 'primitive immune systems' in bacteria, but the role of methylation in bacterial gene regulation and other processes is increasingly accepted. Every *H. pylori* strain harbours a unique set of R–M systems resulting in a highly diverse methylome. We identified a highly conserved GCGC-specific ^{m5}C MTase (JHP1050) that was predicted to be active in all of 459 *H. pylori* genome sequences analyzed. Transcriptome analysis of two *H. pylori* strains and their respective MTase mutants showed that inactivation of the MTase led to changes in the expression of 225 genes in strain J99, and 29 genes in strain BCM-300. Ten genes were differentially expressed in both mutated strains. Combining bioinformatic analysis and site-directed mutagenesis, we demonstrated that motifs overlapping the promoter influence the expression of genes directly, while methylation of other motifs might cause secondary effects. Thus, ^{m5}C methylation modifies the transcription of multiple genes, affecting important phenotypic traits that include adherence to host cells, natural competence for DNA uptake, bacterial cell shape, and susceptibility to copper.**

INTRODUCTION

Epigenetics denotes inheritable mechanisms that regulate gene expression without altering the DNA sequence. In prokaryotes, methyltransferases (MTases) transfer methyl groups from S-adenosyl methionine to adenines or cytosines within a DNA target motif and so contribute to changes of the epigenome (1–3). MTases either belong to restriction–modification (R–M) systems that include MTase and restriction endonuclease (REase) activities, or occur as orphan MTases in the absence of a cognate restriction enzyme (4). Three types of DNA methylation occur in bacteria, N6-methyladenine (^{m6}A), 5-methylcytosine (^{m5}C) and N4-methylcytosine (^{m4}C) (1,2). So far, the major role allocated to bacterial R–M systems is self-DNA protection by restriction of incoming foreign un-methylated DNA (5), and they have thus been described as 'primitive immune systems' (6). Other functions have also been attributed to prokaryotic R–M systems (7–9). For example, methylation marks promoter sequences and alters DNA stability and structure, modifying the affinity of DNA binding proteins and influencing the expression of genes (10,11). Additionally, disturbance of DNA strand separation by methylation can have an effect on gene expression (12).

Methylation can be involved in multiple bacterial functions. In *Escherichia coli*, the Dam adenine MTase plays an essential role in DNA replication (13,14). Another well-studied example is the CcrM MTase from *Caulobacter crescentus* that controls the progression of the cell cycle (15). Furthermore, phase-variable MTases have been shown to control the regulation of multiple genes in several different pathogens, including *Haemophilus influenzae*, *Neisseria meningitidis* and *Helicobacter pylori* (16–18). These MTase-dependent regulons were termed phasevarions (19). As described previously, adenine methylation has been shown to play a key role in transcriptional regulation but the influence of cytosine methylation in gene expression has so far only been investigated in very few studies (20–22).

Helicobacter pylori infection affects half of the world's population and is a major cause of gastric diseases that include ulcers, gastric cancer, and MALT lymphoma (23). This gastric pathogen has coexisted with humans since, at

*To whom correspondence should be addressed. Tel: +4989218072801; Fax: +4989218072802; Email: suerbaum@mvp.uni-muenchen.de
Correspondence may also be addressed to Christine Josenhans. Email: josenhans@mvp.uni-muenchen.de

least, 88 000 years ago (24). *Helicobacter pylori* strains display an extraordinary genetic diversity caused in part by a high mutation rate but especially by DNA recombination occurring during mixed infection with other *H. pylori* strains within the same stomach (25–27). The very high sequence diversity of *H. pylori* and the coevolution of this pathogen with its human host have caused its separation into phylogeographic populations, whose distribution reflects human migrations (28–30).

Despite its small genome, *H. pylori* is one of the pathogens with the highest number of R–M systems (31). The development of Single Molecule, Real-Time (SMRT) Sequencing technology has allowed genome-wide studies of methylation patterns and strongly accelerated the functional elucidation of MTases and their roles in bacterial biology (32,33). Methylome studies of several *H. pylori* strains have revealed that every strain carries a different set of R–M systems leading to highly diverse methylomes (34–37). R–M systems in *H. pylori* were shown to protect the bacterial chromosome against the integration of non-homologous DNA (e.g. antibiotic resistance cassettes), while they had no significant effect on recombination between highly homologous sequences, permitting efficient allelic replacement (9). Despite the diversity of methylation patterns, a small number of target motifs were shown to be methylated in all (one motif, GCGC) or almost all (3 motifs protected in >99% of strains) *H. pylori* strains in a study by Vale *et al.*, who tested genomic DNAs purified from 221 *H. pylori* strains for susceptibility to cleavage by 29 methylation-sensitive restriction enzymes, and in those studies investigating the methylomes of multiple *H. pylori* strains (34,35,37,38). R–M systems have also previously been shown to contribute to gene regulation in *H. pylori*; the phase-variable MTase ModH5 is involved in the control of the expression of virulence-associated genes like *hopG* or *flaA* in strain P12 (39,40).

In the present study, we functionally characterized the role of a highly conserved ^{m5}C MTase (JHP1050, M.Hpy99III) in *H. pylori* (41). We show the MTase gene to be part of the *H. pylori* core genome, present and predicted to be active in all of several hundred *H. pylori* strains representative of all known phylogeographic populations. Transcriptome comparisons of two *H. pylori* wild-type strains and their respective knockout mutants demonstrated that JHP1050 has a strong impact on the *H. pylori* transcriptome that includes both conserved and strain-specific regulatory effects. We show that ^{m5}C methylation of GCGC sequences, among others, affects metabolic pathways, competence and adherence to gastric epithelial cells. Moreover, we provide evidence that methylation of GCGC motifs overlapping with promoter sequences can play a direct role in gene expression, while the regulatory effects of methylated sites outside of promoter regions may be indirect.

MATERIALS AND METHODS

Bacterial culture, growth curves and transformation experiments

H. pylori strains 26695 (42), J99 (43), BCM-300 (35) and H1 (44) were cultured on blood agar plates (45), or in liquid cultures as described (9). Microaerobic conditions were generated in airtight jars (Oxoid, Wesel, Germany) with Anaerocult C gas producing bags (Merck, Darmstadt, Germany). For growth curves, liquid cultures were inoculated with bacteria grown on agar plates for 22–24 h to a starting OD_{600} of ∼0.06 and incubated with shaking (37°C, 140 rpm, microaerobic conditions). The OD_{600} was repeatedly measured until a maximum incubation time of 72 hours. The generation time for *H. pylori* strains J99 and 26695 was calculated to be 3.90 and 4 h respectively, similar to previous calculations (46).

Susceptibility to copper was tested by adding copper sulfate (final concentrations, 0.25 and 0.50 mM) to liquid cultures. The OD_{600} was measured 24 h after inoculation.

For transformation experiments, liquid cultures of the recipient strain were grown overnight (conditions described above). Then, 1 μg/ml of donor bacterial genomic DNA (gDNA) was added to the cultures. The donor gDNA for transformation experiments was purified from isogenic *H. pylori* strains carrying a chloramphenicol (CAT) resistance cassette within the non-essential *rdxA* gene (i.e. J99 *rdxA*::CAT). After gDNA addition, the cultures were incubated for 6–8 h under the same conditions (37°C, 140 rpm, microaerobic atmosphere). Next, the OD_{600} was measured and adjusted to the same number of cells (OD_{600} = 1 as 3×10^8 bacteria). Finally, 100 μl of serial dilutions were plated onto blood agar plates containing chloramphenicol, and incubated at 37°C under microaerobic conditions. Approximately 4–5 days later, colonies were counted and the efficiency of transformation was calculated as cfu/ml.

DNA and RNA extraction

gDNA was isolated from bacteria grown on blood agar plates using the Genomic-tip 100/G kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. The gDNA pellet was dissolved over night at room temperature with EB buffer.

For RNA extraction, 5 ml of bacterial cells grown in liquid medium were pelleted (4°C, 6000 × g, 3 min), snap-frozen in liquid nitrogen and stored at −80°C. Afterwards, bacterial pellets were disrupted with a FastPrep® FP120 Cell Disrupter (Thermo Savant) using Lysing Matrix B 2 ml tubes containing 0.1 mm silica beads (MP Biomedicals, Eschwege, Germany). Isolation of RNA was performed using the RNeasy kit (Qiagen, Hilden, Germany) and on-column DNase digestion with DNase I. A second DNase treatment was carried out using the TURBO DNA-free™ Kit (Ambion, Kaufungen, Germany). Isolated RNA was checked for the absence of DNA contamination by PCR.

DNA and RNA concentrations were measured using a NanoDrop 2000 spectrophotometer (Peqlab Biotechnologies). RNA quality given as RINe number was measured with an Agilent 4200 Tape Station system using RNA Screen Tapes (Agilent, Waldbronn, Germany). All RINe numbers of RNA preparations used for further processing were higher than 8.2, confirming high quality and little RNA degradation.

Construction of mutants and complementation

Inactivation of the MTase or the whole R–M system genes was carried out by insertion of an *aphA3* cassette conferring resistance to kanamycin (Km). A PCR product was

constructed using a combination of primers which added restriction sites and allowed overlap PCR with the *aphA3* cassette (Q5 Polymerase, NEB, Frankfurt am Main, Germany). Ligation of the overlap amplicon with a digested pUC19 vector was done using Quick Ligase (NEB, Frankfurt am Main, Germany). The resulting plasmids were transformed into *E. coli* MC1061. Following plasmid isolation, 750 ng of the plasmids were used for *H. pylori* transformation. Functional complementation of the MTase gene in the strains 26695-mut, J99-mut and BCM-300-mut was achieved by means of the pADC/CAT suicide plasmid approach, as described (47). Transformation of the recipient strains with the resulting plasmid permitted the chromosomal integration of the MTase gene (from strain 26695) into the urease locus, placing the inserted gene under the control of the strong promoter of the *H. pylori* urease operon. The complemented strains were designated 26695-compl, J99-compl and BCM-300-compl, respectively.

Five different methylation motif mutants carrying either a single point mutation in one of the three GCGC motifs of gene *jhp0832*, or a combination of two mutations were constructed using the Multiplex Genome editing (MuGent) technique as described (9,48), with the exception that we used a chloramphenicol resistance cassette within the non-essential *rdxA* locus as selective marker. Sanger sequencing was used to verify the acquisition of the desired mutations within the GCGC motifs. The putative promoter of the gene was predicted within the 50 bp upstream of the transcriptional start site (49) using the BPROM Softberry online tool (50) and verified manually by comparison with *H. pylori* promoter consensus sequences (51). All *H. pylori* mutants were checked via PCR and selected on antibiotic-containing plates. The absence or recovery of methylation was checked by digestion of gDNA with HhaI (NEB, Frankfurt am Main, Germany). All plasmids and primers used in this study are listed in Supplementary Tables S6 and S7.

Microscopy

Live and dead (L/D) staining was performed using the BacLight Bacterial Viability kit (Thermo Fisher Scientific, Darmstadt, Germany) according to the manufacturer's instructions. Bacteria were harvested from plates incubated for 22–24 h, and suspended in 1 ml of BHI medium without serum to an adjusted OD_{600} of ∼0.1. Then, 100 μl of this dilution were mixed with the BacLight dyes, giving green and red fluorescence for live and dead/dying bacteria, respectively. After 30 minutes of incubation at room temperature and in the dark, 0.5 μl of the mix was suspended on slides that were analyzed with an Olympus BX61-UCB microscope equipped with an Olympus DP74 digital camera. Between 80 and 100 pictures from at least two independent biological and technical replicates were obtained and analyzed with the CellSens 1.17 software (Olympus Life Science) and ImageJ (52).

Gram staining was performed as follows: 300 μl of liquid cultures grown over-night were pelleted (6000 × g, 3 min, room temperature) and washed 3 times with PBS (6000 × g, 3 min, room temperature). Afterward, 100 μl of the pellets resuspended in PBS were added to a glass slide that was dried at 37°C during 10–15 min, heat-fixed and Gram-stained.

Bacterial cell adherence assays

Assays for bacterial adherence to the human stomach carcinoma cell line AGS were performed as previously described with slight modifications (53,54). *Helicobacter pylori* strains grown to an OD_{600} ∼1 were suspended in RPMI 1640 medium supplemented with 10% fetal calf serum (FCS). Experiments were executed in 96-well plates containing 2 × 10^5 fixed AGS cells (ATCC CRL-1739) per well. AGS cells were fixed with 2% freshly prepared paraformaldehyde in 100 mM potassium phosphate buffer (pH 7) and subsequently quenched and washed as described (53). Live *H. pylori* bacteria were added to cells at a bacteria:cell ratio of 50 (54), followed by brief centrifugation (300 × g, 5 min), and co-incubated for 1 h at 37°C with 5% CO_2. After this, plates were washed twice with PBS, followed by overnight fixation with 50 μl of fixing solution (see above). Fixing solution was renewed once and incubated for an additional 30 min, and quenched twice with 50 μl of quenching buffer for 15 min. Bacterial adherence to the AGS cells was subsequently quantitated using antibody-based detection as follows: cells were washed three times with washing buffer PBS-T (PBS + 0.05% Tween20), blocked for 30 min with 200 μl of the assay diluent (10% FCS in PBS-T) and washed four times with PBS-T. Then, 100 μl of a 1:2,500 dilution of the primary antibody, α-*H. pylori* (DAKO/Agilent Technologies, Hamburg, Germany) was added and incubated for 2 h. Afterward, cells were washed and incubated with 100 μl of a 1:10 000 dilution of the secondary antibody, goat anti-rabbit HRP-coupled (Jackson ImmunoResearch, Ely, United Kingdom) for 1 h. After four final wash steps, the 96-well plates were finally incubated with 100 μl TMB substrate solution (1:1, Thermo Fisher Scientific, Darmstadt, Germany). The color reaction was developed in the dark for 30 min and stopped with 50 μl of phosphoric acid (1 M). Absorbance was measured at 450/540 nm (Sunrise™ Absorbance Reader). Negative controls (mock-coincubated, fixed AGS cells) were treated the same way with primary and secondary antibody dilutions.

Bioinformatic analyses

To analyze the conservation and the genomic context of the JHP1050 MTase gene in a diverse collection of *H. pylori* strains, we assembled a database consisting of 459 *H. pylori* genomes that included strains from all known phylogeographic populations and subpopulations (Supplementary Table S1). Genomes and methylomes of the four strains investigated in this study have been published previously (34,35,42–44), with the exception of the H1 methylome (own unpublished data). The nucleotide sequence of gene *jhp1050* from the *H. pylori* strain J99 was used to identify and extract the *jhp1050* homologs and the sequences of the flanking genes. The NCBI blastn microbes and StandAlone Blast tools were used to extract the sequences from publicly available genomes and private genomes, respectively.

To study whether the methylated cytosines of the GCGC motifs had a higher tendency to deaminate (^{m5}C>T) than

unmethylated cytosines, we compared the frequency of C>T transitions to either C>A or C>G polymorphisms inside and outside of GCGC motifs among a phylogeographically distinct set of *H. pylori* genomes. GCGC motifs were identified in two *H. pylori* genomes, 26695 and PeCan18, which were subsequently used as reference and aligned separately against 11 other *H. pylori* genomes using BioNumerics version 7.6 (Applied Maths, Sint-Martens-Latem, Belgium). Polymorphisms were called in both alignments and pooled together. The percentage of mutated ^{m5}C positions within GCGC motifs was determined for each possible transition or transversion as follows:

$$\% = \frac{\text{number of m5C} \rightarrow \text{base} * 100}{\text{total of motifs in the reference}}$$

Since the G^{m5}CGC motif is palindromic, the same analysis was performed for the complementary strand, where the position of the second G (^{m5}C in the complementary strand) was compared for each possible mutation and calculated as above.

Finally, the percentage of mutated C outside GCGC motifs calculated as follows for each possible mutation:

$$\% = \frac{(\text{Total number of C} \rightarrow \text{base} - \text{number of m5C} \rightarrow \text{base}) * 100}{\text{Total number of C in the reference genome}}$$

The same analysis was performed for other ^{m5}C motifs and for non-methylated motifs as well as for the non-methylated C of the GCGC motif.

Expected sites

The expected number of motifs per kb was calculated as follows:

$$\text{Expected sites/kb} = \frac{\text{Total observed GCGC motifs} * 1000\,(\text{bp})}{\text{genome length (bp)}}$$

The expected number of motifs/kb was 3.89 (J99), 3.91 (BCM-300), 3.76 (26695) and 3.74 (H1). The expected number of motifs within CDS can then be calculated using the expected number of motifs/kb and the gene length. Finally, the ratio observed/expected (O/E) motifs within CDS was calculated to detect genes enriched for the presence of specific sequence motifs. For example, for a given gene in J99 that is 630 bp long and has two GCGC motifs (observed), the expected number of motifs within that gene would be: 630*3.89/1000. For this example calculation, the O/E ratio would be 0.82, suggesting GCGC motifs are underrepresented in this gene.

The GCGC motif is a 4-mer palindrome. In order to calculate the expected number of motifs that would randomly occur within a genome fragment (either CDS or intergenic region), we took into account the number of 4-mers in a given sequence, $N - K + 1$ (where N means sequence length and K the motif length, in this case 4), and the frequency of G/C (0.2) and A/T (0.3), and calculated the expected number of motifs in a specific fragment as $(N - K + 1)*(0.2)^4$.

RNA-Seq analysis

RNA-Seq analysis was performed on an Illumina HiSeq sequencer obtaining single end reads of 50 bp. Ribosomal RNA (rRNA) depletion was performed prior to cDNA synthesis using a RiboZero Kit (Illumina, Germany). Isolated RNA from a total of 6×10^8 to 1×10^9 bacterial cells corresponding to log phase of growth was used for sequencing. Three biological replicates were used for all the strains, except for J99-mut since one replicate had to be discarded during library preparation. Mapping of reads to a reference genome was done with Geneious 11.0.2 (55). Reads mapping multiple locations or intersecting multiple CDS were counted as partial matches (i.e. 0.5 read). Differential expression was calculated using DESeq2 (56). Fold change (FC) of two and FDR adjusted *P*-value of 0.01 were used as a cut-off.

Quantitative PCR (qPCR)

One μg of RNA was used for cDNA synthesis using the SuperScript™ III Reverse Transcriptase (Thermo Fisher Scientific, Darmstadt, Germany) as described before (54). qPCR was performed with gene specific primers (Supplementary Table S7) and SYBR Green Master Mix (Qiagen, Hilden, Germany). Reactions were run in a BioRad CFX96 system. Standard curves were produced and samples were run as technical triplicates. For quantitative comparisons, samples were normalized to an internal 16S rRNA control qPCR. Details about the reaction conditions in compliance with the MIQE guidelines are specified in Supplementary Methods 1.

RESULTS

Distribution of the G^{m5}CGC R–M system (JHP1049-1050) within a globally representative collection of *H. pylori* genomes

Despite the extensive inter-strain methylome diversity of *H. pylori*, a small number of motifs have been shown to be methylated in all or most of the strains (38). Here, we focused on the MTase JHP1050 (M.Hpy99III), which methylates GCGC sequences, resulting in G^{m5}CGC motifs. Although ^{m5}C methylation is less common in prokaryotes than ^{m6}A methylation, based on the Restriction Enzyme Database (REBASE) (57), this particular motif is highly conserved in many bacterial species.

We therefore hypothesized that the GCGC-specific MTase in *H. pylori* might play an important role apart from self-DNA protection.

We first analyzed the conservation and the genomic context of the MTase gene. The nucleotide sequence of gene *jhp1050* from the *H. pylori* strain J99 was used to identify the *jhp1050* homologs and the sequences of the flanking genes in a collection of 458 *H. pylori* genomes representing all known phylogeographic populations (Supplementary Table S1).

Based on the gene sequences, the M.Hpy99III MTase was predicted to be active in all *H. pylori* strains. The MTase sequence was highly conserved between all 459 strains, with an average nucleotide sequence identity of 94.04 ± 2.03%, and a lowest nucleotide sequence identity of 87% between the most dissimilar alleles. The analyzed region of the chromosome was also highly conserved among the strains and all the flanking genes were present with the exception of

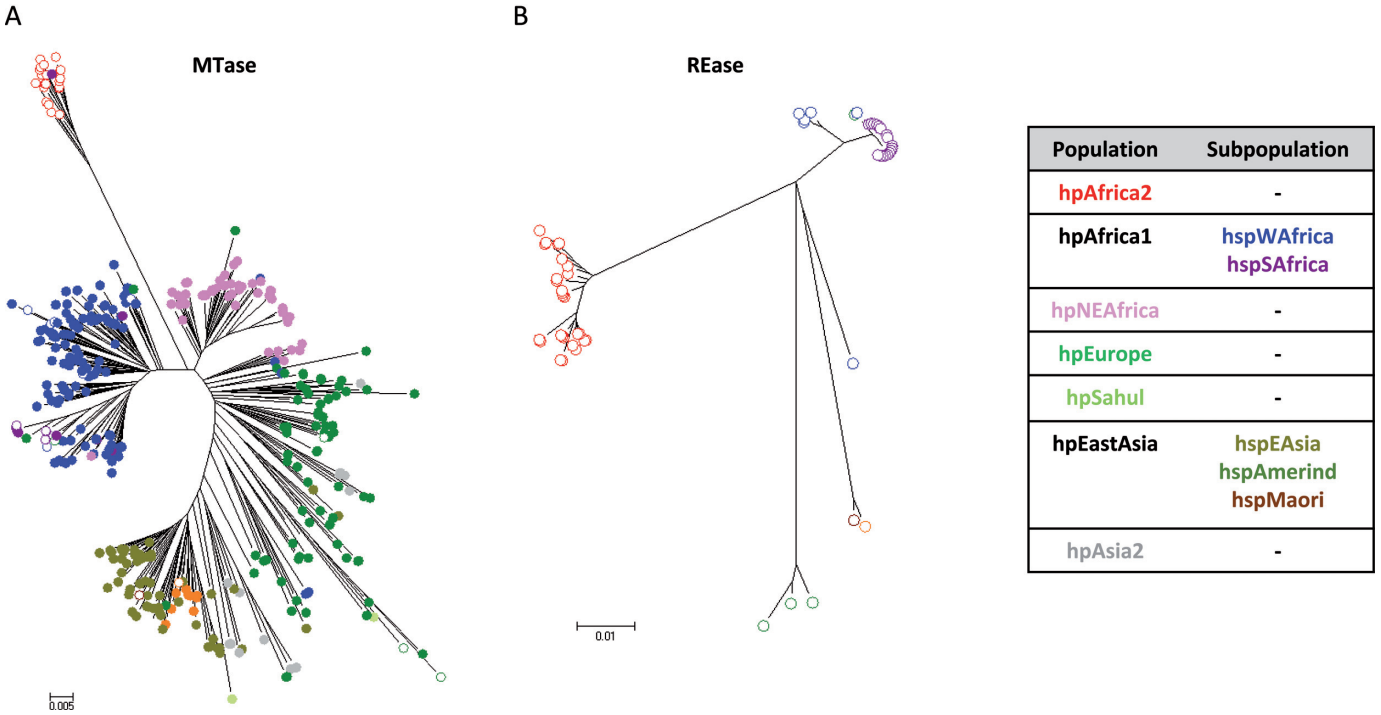


Figure 1. Phylogenetic analysis of the GCGC-specific R–M system JHP1050/1049 (M.Hpy99III/Hpy99III) in *H. pylori*. Neighbour-Joining trees based on the nucleotide sequences of MTase M.Hpy99III (**A**) and REase Hpy99III (**B**). In both cases, strain symbols are colored according to the phylogeographic population assignment based on seven gene MLST and STRUCTURE analysis (see right panel for color coding). Filled circles represent strains without REase gene, while unfilled circles are used for strains containing both MTase and REase genes.

the cognate REase gene (*jhp1049*) which was present in only 61 of the 459 strains. Interestingly, the majority of the REase-positive strains belong to populations with substantial African ancestry, particularly to hpAfrica2, followed by hspSAfrica, hspWAfrica and hpEurope. Furthermore, none of the analyzed hspAsia2 or hspEAsia strains carried the REase gene (Supplementary Table S1). Only 15 REase genes were predicted to be functional, while the others were pseudogenes due to premature stop codons and/or frameshift mutations (Supplementary Table S2). We identified a 10 bp repeat sequence flanking the REase gene. The same sequence was found downstream of the MTase gene and 48 bp upstream of *jhp1048* in 15 of the REase-negative strains. In all cases, the sequence contained a homopolymeric region with a variable number of adenines. This suggests that the REase gene was excised from the genome. The same sequence was found in *H. cetorum* and *H. acinonychis*, the closest known relatives of *H. pylori* (Supplementary Table S3 and Supplementary Figure S1). Moreover, the phylogenetic trees of MTase and REase gene sequences in general were congruent with the global population structure of *H. pylori* (Figure 1) (24). This implies that the R–M system was acquired early in the history of this gastric pathogen. The REase gene appears to have been lost later during species evolution in the majority of the strains, likely before the first modern humans left Africa. Nonetheless, the REase gene could have been reintroduced in some strains (i.e. hpEurope strains) via recombination of the flanking repeats.

Construction of MTase mutants and analysis of target motif abundance

To functionally characterize this highly conserved MTase, we constructed MTase-deficient mutants. The MTase gene was disrupted in the strains 26695 (hpEurope), H1 (hspEAsia) and BCM-300 (hspWAfrica) and the whole R–M system was inactivated in strain J99 (hspWAfrica), the only of the four strains that contained both MTase and REase. Genes were inactivated by insertion of an antibiotic resistance cassette. The loss of methylation was verified by restriction assays using the restriction enzyme HhaI that only cleaves unmethylated GCGC sequences (Supplementary Figure S2). In the following text, mutants are named by the wild type strain name followed by –mut. Complementation of the MTase in strains 26695, J99 and BCM-300 was performed by reintroducing the MTase gene of 26695 (see Materials and Methods). The transcription of the MTase gene was tested in the four wild type strains and in two of the complemented mutant strains (J99-compl and 26695-compl). The transcript amounts of the MTase varied substantially between wild type strains (Supplementary Figure S3). Whether these differences between mRNA amounts have any functional implications is currently unknown.

Methylome comparison of the four strains exhibited only four methylated motifs shared between the strains ($G^{m5}CGC$, $G^{m6}ATC$, $C^{m6}ATG$ and $G^{m6}AGG$) (Supplementary Table S4). All of these motifs occur frequently in the J99 genome (GCGC, 6399 motifs; GATC, 5479; CATG,

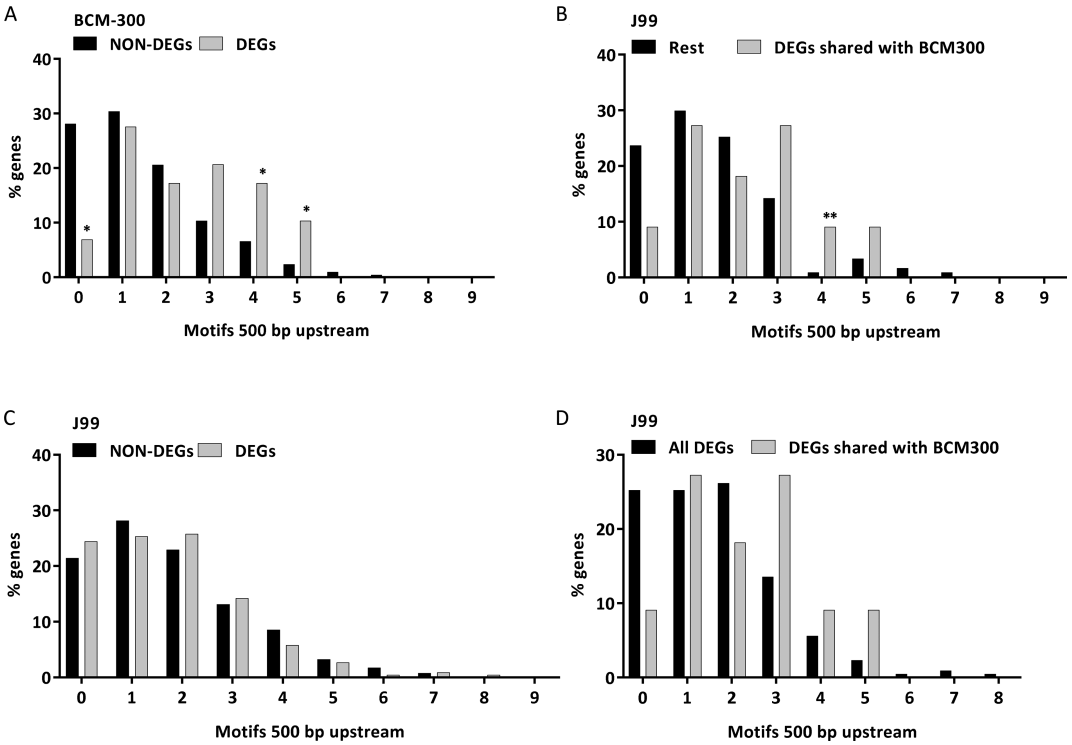


Figure 2. Graphical representation of the percentage of genes with GCGC motifs 500 bp of sequence upstream of the start codon for differentially expressed genes (DEGs) and genes not showing differential expression (Non-DEGs). Non-DEGs versus DEGs in BCM-300 (**A**) and J99 (**C**). DEGs in J99 shared with BCM-300 versus the rest of the J99 genes (**B**). DEGs in J99 shared with BCM-300 versus the rest of the J99 DEGs (all DEGs) (**D**). Statistics: Chi-square, $*P < 0.05$, $**P < 0.01$.

7560; GAGG, 5027). The distribution of GCGC motifs along the genomes was not uniform. We compared this observed distribution to the motif density that would be expected from a random distribution of motifs across the genomes. While the number of motifs was generally higher than expected for a random distribution, fewer motifs than predicted were found in the *cag*PAI and the plasticity zones (PZ) (Supplementary Figure S4A). Finally, we calculated the total number of GCGC motifs that would randomly occur in the complete genomes, the coding regions and the intergenic regions according to the nucleotide composition of *H. pylori.* The observed number of motifs in the coding regions was more than twice the expected number for all four genomes. In contrast, the observed and expected numbers of motifs in the intergenic regions were very similar (Table 1). Therefore, coding sequences appeared to display an over-representation of GCGC motifs.

Comparative RNA-Seq transcriptome analysis of *H. pylori* J99 and BCM-300 and their isogenic MTase mutants

Due to the extraordinary conservation of the $G^{m5}CGC$ MTase in all analyzed strains despite the absence of a cognate REase, we postulated that the function of the enzyme might be more important than simply serving for self-DNA protection. Therefore, in order to study a putative role in gene regulation, we performed comprehensive RNA-Seq analysis in the strains J99, BCM-300 and the two corresponding isogenic MTase mutants.

Whole transcriptome comparison of the J99-mut and J99 wild type strains exhibited 225 differentially expressed genes (DEGs). One hundred fifteen genes were upregulated and 110 downregulated in J99-mut compared with J99 wild type (*P*-adjusted value < 0.01, fold change (FC) > 2). In contrast to J99, the transcriptomes of the BCM-300-mut and wild type strains showed only 29 genes that were differentially expressed in the mutant, all of which were downregulated (*P*-adjusted value < 0.01, FC > 2) (Supplementary Table S5). The two mutants, J99-mut and BCM-300-mut, shared 10 downregulated genes but no upregulated genes (Table 2). Using qPCR, we confirmed that 9 of the 10 shared genes were significantly downregulated as shown by RNA-Seq (Supplementary Figure S5). The gene *jhp1283* showed either upregulation or downregulation in different biological replicates.

In order to understand how the distribution of motifs could play a role in transcriptional regulation, we analyzed the frequencies of GCGC motifs in a 500 bp sequence upstream of each DEG and compared those with sequences upstream of genes that were not differentially regulated (non-DEGs), and with coding sequences (CDS).

In strain BCM-300, the number of GCGC motifs located within 500 bp upstream of the start codon was higher for the 29 DEGs than for non-DEGs (Figure 2A). In contrast, in strain J99, the percentage of genes with three or more

Table 1. Observed and expected frequencies of GCGC motifs in the genome sequences of the four *H. pylori* strains analyzed in this study

Strain	Genome size (bp)	Total length of CDS (bp)	Total length of intergenic sequences (bp)	Predicted no. of GCGC sites/1 kb	No. of motifs in genome	Expected no. of motifs in genome	No. of motifs in CDS	Expected no. of motifs in CDS	No. of motifs in intergenic sequences	Expected no. of motifs in intergenic sequences
26695	1667867	1494807	173060	3.76	6269	2669	5950	2392	319	277
J99	1643831	1486413	157418	3.89	6399	2630	6110	2378	289	252
H1	1563305	1436409	126896	3.74	5846	2501	5655	2298	191	203
BCM-300	1667883	1520688	147195	3.91	6523	2669	6273	2433	250	236

Table 2. Shared differentially expressed genes (DEGs), displaying GCGC methylation-dependent transcription in *H. pylori* J99 and BCM-300. Positive values for fold change (FC) indicate lower transcription in the mutants compared to the wild type strains

Gene	Description	J99 locus_tag	J99 FC	BCM-300 locus_tag	BCM-300 FC
bioD	dethiobiotin synthetase	*jhp_0025*	2.1986	*BCM_00034*	2.9978
				BCM_00035	2.9424
feoB	iron(II) transport protein	*jhp_0627*	3.8803	*BCM_00707*	4.3250
-	unknown	*jhp_0749*	3.8245	*BCM_00859*	3.1947
moeB	molybdopterin/thiamine biosynthesis activator	*jhp_0750*	4.0863	*BCM_00860*	3.6033
-	unknown	*jhp_1102*	2.4868	*BCM_01112*	2.2810
cah	alpha-carbonic anhydrase	*jhp_1112*	2.0723	*BCM_01124*	3.3563
trmU	tRNA-methyltransferase	*jhp_1254*	4.5288	*BCM_01276*	5.7005
-	unknown	*jhp_1281*	3.4690	*BCM_01305*	2.0216
-	unknown	*jhp_1253*	2.9141	*BCM_01275*	3.2789
crdR	response regulator	*jhp_1283*	2.8855	*BCM_01307*	3.2789
		jhp_1443	2.9141		

GCGC motifs within 500 bp upstream of the start codon was similar for DEGs and non-DEGs (Figure 2C). However, the 10 DEGs of strain J99 that were shared with BCM-300 showed the same overrepresentation of GCGC motifs observed in strain BCM-300 (Figure 2B, D). Furthermore, DEGs in BCM-300 displayed more motifs within their CDS than expected if GCGC motifs were distributed randomly across the whole genome, while the opposite effect occurred for the non-DEGs. The same trend was evident in J99 when we only compared the DEGs shared with BCM-300 with the rest of the genes (Supplementary Figure S6A).

In addition, we observed that 6 of the 10 shared DEGs harbored GCGC motifs within the 50 bp sequence upstream of the TSS described by Sharma and colleagues in strain 26695 (49), called here region upstream of the TSS (upTSS). Sequences within the putative promoter regions immediately upstream of the TSS are likely to exert the strongest influence on transcriptional regulation. We compared the upTSS of 26695 with J99 and BCM-300 via sequence alignment. There were 48 genes in J99 and 45 in BCM-300 with GCGC motifs within the 50 bp upstream sequence (sRNA and asRNA were excluded). In J99, 13 of the 225 DEGs contained GCGC motifs within the upTSS sequence. In BCM-300, 11 of the 29 DEGs contained motifs within the upTSS. This proportion of DEGs with motifs within the upTSS suggests that the window of 50 bp upstream of the TSS may play a role in transcription regulation. Indeed, the FC was slightly increased by motifs within the upTSS (Supplementary Figure S6B). Gene *jhp1283*, the only of the 10 shared DEGs identified by RNA-Seq that was not confirmed in qPCR assays, did not have any GCGC motif within the upTSS, suggesting that this gene might not be directly regulated by methylation.

Direct regulation of gene expression by ^{m5}C methylation

Inactivation of the M.Hpy99III MTase had different effects on the transcriptomes of the two strains tested, with far more genes affected in strain J99 versus the BCM-300 strain. We hypothesized that the loss of GCGC methylation might have both direct and indirect effects on transcription. In order to demonstrate a direct association between methylation and gene expression, we generated a set of mutants in strain J99 where site-specific mutations were introduced into selected GCGC motifs located within the CDS as well as in the upstream region of one specific gene showing strong differential regulation.

The selected gene for this approach (*jhp0832*) was downregulated in J99-mut (FC = 5.95). Its homolog in *H. pylori* strain 26695 (HP0893) was reported to be an antitoxin from a Type II Toxin–Antitoxin (TA) system (58). The cognate toxin (*jhp0831*) was also downregulated in J99-mut (FC = 3.64). The two genes belong to the same operon where the antitoxin is located upstream of the toxin. No homologous genes were found in BCM-300.

Two GCGC motifs were located within the 500 bp upstream window of the antitoxin gene and one motif was located within the coding sequence. Of the two upstream motifs, one was located within the upTSS in J99 and overlapped with the -10 box of the predicted promoter (Figure 3). Thus, owing to the high FC and this distribution of three GCGC motifs, *jhp0832* seemed to be a good candidate to dissect the role of different GCGC motifs in the transcriptional regulation of *jhp0832*.

We constructed three mutants where each of the motifs was individually changed to GAGC so that the motif could no longer be methylated (*jhp0832* mut1, *jhp0832* mut2 and *jhp0832* mut3). We also constructed two mutants (*jhp0832*

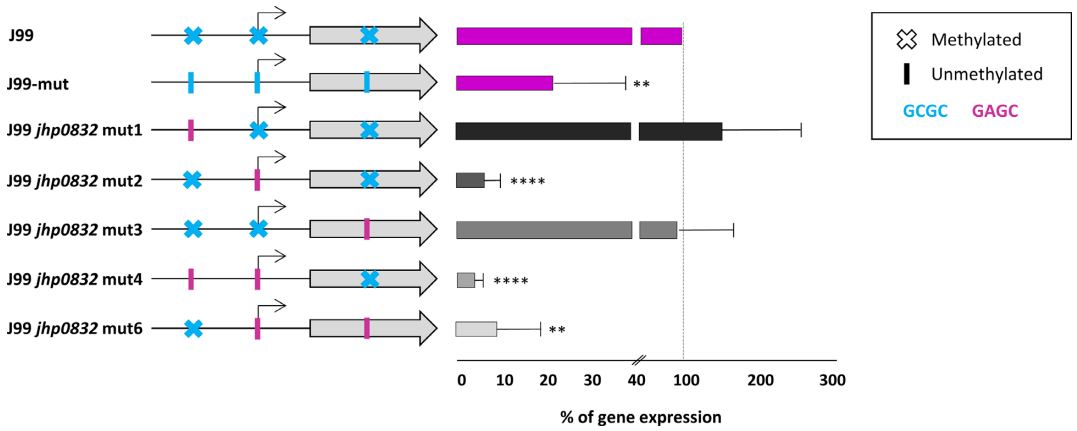


Figure 3. Quantification of transcript amounts of *jhp0832* in *H. pylori* strains J99, J99-mut and the J99 mutants with point mutations within the GCGC motifs. qPCR results are represented in the right panel, three different biological replicates were performed. Statistics: One-way ANOVA, $^{**}P < 0.01$, $^{****}P < 0.0001$, bars: SD. Legend: The *jhp0832* gene is shown as a gray arrow. The predicted promoter is represented by a black arrow. Crosses represent methylated motifs while vertical lines mean unmethylated motifs (due to site-directed mutation, or to inactivation of the MTase in strain J99-mut). The GCGC motifs appear in blue and the motifs mutated to GAGC are colored in pink.

Table 3. List of mutants carrying different point mutations modifying the GCGC motifs within or immediately upstream of *jhp0832*

Mutant name	GCGC motif mutated	Plasmid
jhp0832 mut1	1	pSUS3427
jhp0832 mut2	2	pSUS3428
jhp0832 mut3	3	pSUS3429
jhp0832 mut4	1, 2	pSUS3427, pSUS3428
jhp0832 mut6	2, 3	pSUS3428, pSUS3429

All mutants were constructed using the MuGent technique (see Materials and Methods) using the indicated plasmids and the rdxA::CAT PCR product. Thus, all the mutants were resistant to chloramphenicol.

mut4 and *jhp0832* mut6) where two out of the three GCGC motifs were mutated (Figure 3A, Table 3). We were unable to generate a triple mutant carrying combined point mutations in all three motifs, which might be due to toxic dysregulation of the toxin–antitoxin system after removal of all methylatable GCGC motifs.

Differential expression of *jhp0832* was determined by qPCR. Three of the mutants (*jhp0832* mut2, *jhp0832* mut4 and *jhp0832* mut6) displayed a strong downregulation of *jhp0832* expression, similar to J99-mut. Interestingly, these mutants shared the mutation in the GCGC motif located within the upTSS and the predicted promoter of the gene. In contrast, modification of the motifs outside of the upTSS did not consistently alter the expression of the gene (Figure 3).

Phenotypes of *H. pylori* GCGC MTase mutants: growth, viability and shape

In order to test whether the absence of ^{m5}C methylation and the associated differential transcriptomes had a role in the fitness of *H. pylori*, we determined the growth of the strains in liquid medium (Figure 4A). J99-mut had a significant growth defect compared with the J99 wild type strain. Complementation of the MTase gene restored the observed growth phenotype. Similarly, a significant reduction in growth was shown for BCM-300-mut at stationary phase that could be restored to wild-type growth by functional complementation. Although non-significant, a slight delay in growth was noted in 26695-mut and H1-mut compared to the wild type and the complemented strains.

Bacterial morphology serves to optimize biological functions and confers advantages to particular niches. *H. pylori* is a spiral-shaped bacterium that can enter a coccoid state under certain stress conditions (59). *H. pylori* J99-mut entered a coccoid state very early in liquid cultures. A substantial proportion of coccoid forms were visible between 6 and 9 h after inoculation while they are rarely found in the wild type strain at this time point (Supplementary Figure S7A). An effect of the inactivation of JHP1050 on the morphology was not observed for the other three strains 24 h post-inoculation (Supplementary Figure S7B). Complementation of J99-mut restored the wild type phenotype. We note that live/dead staining did not show a significant difference between the percentage of live vs. dead bacteria between the wild type and the mutant strains collected from 22–24 h plates. There was a slight reduction in viability in the BCM-300-mut strain, but no differences were found in the other strains (Figure 4B). As in the liquid cultures, an increased number of rounded bacteria were noticed for J99-mut (Figure 4C).

^{m5}C methylation contributes to the high mutation frequency in *H. pylori*

H. pylori lacks most of the genes involved in mismatch repair (MMR) in other bacteria which is thought to be at least partially responsible for the high mutation rate of this bacterium (42,60). Deamination of ^{m5}C to thymine (T) is responsible for the most common single nucleotide mutation (61). *H. pylori* is known to have a very high mutation rate, and ^{m5}C MTases might contribute to that by increasing the number of nucleotides susceptible to deamination. To test whether ^{m5}C methylation within GCGC motifs played a role in *H. pylori* evolution by favouring deamination, we aligned whole genomes of two *H. pylori* strains (26695

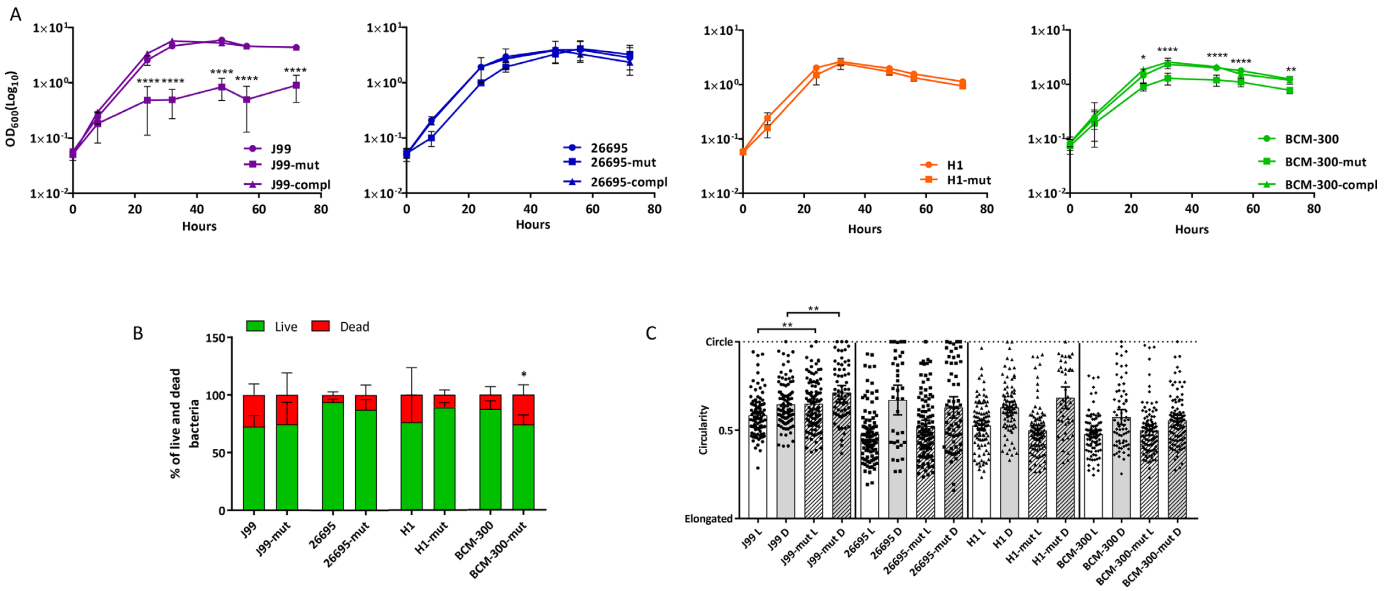


Figure 4. MTase JHP1050 inactivation causes phenotypic effects that vary between strains: growth, viability and morphology. (**A**) Growth curves for four wild type strains and mutants and for the complemented strains J99-compl, 26695-compl and BCM-300-compl were measured for 72 h. The doubling time for *H. pylori* was calculated to be 3.87 h (46). Statistics: two-way ANOVA, $*P < 0.05$, $**P < 0.01$, $****P < 0.0001$, bars: SD. (**B**) Viability of the strains was studied using epifluorescence microscopy after live/dead staining. Statistics: two-way ANOVA, $*P < 0.05$, bars: SD. (**C**) Bacterial morphology was quantitated from epifluorescence microscopy pictures using ImageJ. A value of 0 represents completely elongated bacteria, while a value of 1 means a complete circle (coccoid bacteria). Statistics: one-way ANOVA, $**P < 0.01$, bars: 95% confidence interval (CI).

and PeCan18), used as reference, against 11 other complete genome sequences (see Material and Methods for details). The results strongly support a role of ^{m5}C methylation in *H. pylori* mutagenesis, since the percentage of C→T mutations within GCGC motifs was significantly higher than the overall percentage of C→T or C→ another base transition in the genomes of all the tested strains. In addition, we performed the deamination analysis on (i) other ^{m5}C methylated cytosines within different motifs, (ii) on non-methylated cytosines within motifs containing ^{m5}Cs and (iii) non-methylated motifs containing cytosines. We observed a higher frequency of C → T mutations for the ^{m5}C within the motifs.

Therefore, the ^{m5}C methylation of the common GCGC motif in all *H. pylori* strains may contribute to the high mutation rate of *H. pylori* and its overall low GC content by favouring deamination (Supplementary Figure S4B).

Regulation of Outer Membrane Proteins (OMPs) and adherence by ^{m5}C methylation in GCGC motifs is strain-specific

OMP genes represent ∼4% of the *H. pylori* genome (62). Fourteen OMPs were found to be upregulated in J99-mut (Supplementary Table S5). Confirmation of the upregulation of OMP genes was obtained using qPCR in J99-mut (Supplementary Figure S5C, D). We detected either no regulation or weak upregulation in the other mutated strains (Supplementary Figure S5C, D), which was in agreement with the transcriptome data obtained for BCM-300. Only three of these OMPs were also slightly upregulated in BCM-300-mut, but the FC was lower than the stringent cut-off of 2 used in the transcriptome analyses. A bacterial adherence assay based on coincubation of fixed AGS cells with all four wild type strains and corresponding isogenic GCGC MTase mutants was performed to test for an adherence phenotype. Only J99-mut had a significantly higher adherence to the cells compared to the respective wild type strain, while no significant differences in adherence were determined for the rest of the strains (Figure 5C). Taken together, the increased expression of a number of OMP genes in the absence of methylation in J99 might contribute to a stronger adherence of the bacteria to the cells, while this was not observed for the other tested strains.

GCGC methylation regulates natural competence in *H. pylori*

Natural competence is a hallmark of *H. pylori*. Competence is conferred by the ComB system, an unusual type IV secretion system related to the VirB system of *Agrobacterium tumefaciens* (63). RNA-Seq results identified three *com* genes (*comB8*, *comB9* and *comEC*) that were less transcribed in J99-mut compared to the wild type strain, while the genes were not found to be differentially regulated in BCM-300. ComB9 and ComB8 are part of the outer- and inner-membrane channels of the DNA uptake system, while ComEC allows the translocation of the DNA through the inner membrane to the cytoplasm. qPCR demonstrated the downregulation of these genes in the two additional strains tested, 26695-mut and H1-mut, in comparison with their respective wild type strains (Supplementary Figures S5A and S5B).

The DNA uptake capacity of the four MTase-mutated strains in comparison to the wild types was quantitated by counting recombinant colonies carrying an antibiotic resistance cassette after standardized transformation experiments (see Materials and Methods). A significant reduction in the efficiency of transformation to chloramphenicol

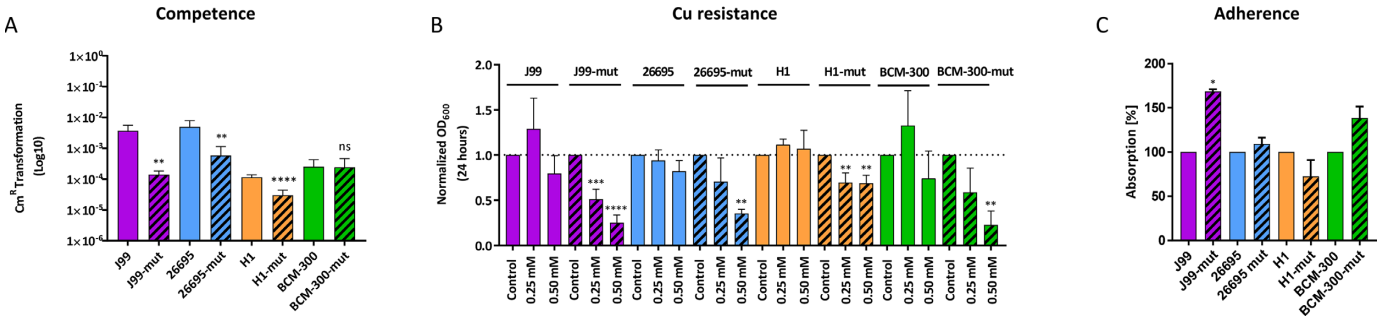


Figure 5. MTase JHP1050 inactivation causes phenotypic effects that vary between strains: natural competence, resistance to copper, and adherence to host cells. (**A**) Transformation experiments were performed with 1 μg/ml of gDNA. Statistics: Welch's unpaired *t* test, $^{**}P < 0.01$, $^{****}P < 0.0001$, bars: SD. (**B**) The growth of J99 wild type, J99-mut, BCM-300 wild type and BCM-300-mut strains was measured 24 h post-inoculation after addition of different concentrations of copper sulfate to the cultures. Data was normalized to a control culture without copper. Statistics: One-Way ANOVA, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$, bars: SD. (**C**) Adherence of *H. pylori* wild type and mutant strains to fixed AGS cells. Statistics: unpaired *t*-test, $^{*}P < 0.05$, bars: SD.

resistance was observed in the J99, 26695 and H1 mutants compared to their respective wild type strains, but no difference was apparent for BCM-300 (Figure 5A). The down-regulation of these three components of the ComB system might be sufficient to reduce the competence in three of the strains.

Loss of ^{m5}C methylation of GCGC motifs increases susceptibility to copper

Copper is an essential metal used by *H. pylori* as a cofactor in multiple processes and it has been shown, for example, to be important for colonization (64). However, an excess of heavy metals can be toxic for the bacterial cells, leading to the existence of several mechanisms to control copper homeostasis. One of the mechanisms involves the two-component system CrdR/S. In the presence of copper, the sensor kinase CrdS phosphorylates the response regulator CrdR triggering the activation of a copper resistance protein and a copper efflux complex (65).

The transcriptional regulator gene *crdR* was less expressed in both J99 and BCM-300 MTase mutants (Table 2). In both strains, one GCGC motif is located within the upTSS of the transcriptional regulator, suggesting a direct regulation via ^{m5}C methylation. To test whether the mutated strains were less resistant to copper due to the lower expression of the *crdR* gene, we compared the influence of added copper sulfate on growth in liquid culture between MTase mutants and wild type strains. The presence of copper caused a clear growth defect of the mutants when compared with the wild type strains, and with a control culture without added copper (Figure 5B). The results indicate that ^{m5}C methylation within the upTSS is required to ensure sufficient transcription of the transcriptional regulator to protect against an excess of copper.

DISCUSSION

Most previous studies of R–M systems in *Helicobacter pylori* have focussed on the striking diversity of methylation patterns and its implications. In contrast to the dozens of MTases only present in subsets of strains, *H. pylori* also possesses few enzymes that are highly conserved between strains. Here, we have explored the function of one ^{m5}C MTase (JHP1050) that is very highly conserved and that we predicted to be active in all of a globally representative collection of 459 *H. pylori* strains analyzed. The collection included isolates from the most ancestral *H. pylori* population, hpAfrica2, and the presence of the MTase in all *H. pylori* phylogeographic populations and subpopulations indicates that the gene has been part of the *H. pylori* core genome since before the Out of Africa migrations, and before the *cag* pathogenicity island was acquired (24). The cognate REase gene was detected in few strains only, almost all of which belong to African *H. pylori* populations. This indicates that the REase was lost from the genome very early in the history of this gastric pathogen. These data indicate a strong selective pressure to maintain the presence and activity of the MTase, while the REase gene either lost its function or was completely deleted. The apparent strong selection of the maintenance of this MTase in the *H. pylori* genome was in striking contrast to the cognate REase and to the vast majority of R–M systems so far identified in *H. pylori*, indicating that the MTase alone is likely to serve an important function for the bacterium. Since methylation has been shown to influence gene expression in several bacterial species, we considered a regulatory function most likely, and performed global transcriptome analysis using RNA-Seq.

The results obtained by RNA-Seq analysis of two *H. pylori* wild type strains, J99 and BCM-300, and their respective MTase mutants confirmed our hypothesis that GCGC methylation affects the transcription of multiple *H. pylori* genes, but we were surprised by the substantial differences between the two strains. While there were 225 DEGs in J99, whose transcription was significantly changed in the MTase mutant, only 29 genes showed an altered expression in BCM-300, and only 10 DEGs were shared between both strains.

To better understand the relationship between GCGC methylation and transcriptional gene regulation, we studied the correlation between the presence of GCGC motifs within coding sequences and upstream regulatory sequences and the effect of a loss of methylation on transcription. When we screened the 500 bp of sequence upstream of

the start codons of all DEGs for GCGC motifs and compared the results with those obtained for the upstream sequences of non-DEGs, we observed that DEGs frequently contained more than three GCGC motifs while the majority of the non-DEGs had 0 or 1 motifs (Figure 2). Among the DEGs, the presence of GCGC motifs within the upTSS was significantly associated with higher fold change (FC) values (Supplementary Figure S6B). Moreover, there were more DEGs with higher number of motifs within the coding sequence than expected when compared with the non-DEGs (Supplementary Figure S6A). These results are similar to reports from *Vibrio cholerae*, where a significant correlation between differential regulation and the number of motifs within the coding sequence was reported for a ^{m5}C MTase (21).

Six of the 10 DEGs shared between J99 and BCM-300 contained GCGC motifs within the upTSS. We therefore investigated the relationship between the presence of a methylatable GCGC sequence and gene transcription using site directed mutagenesis. When the GCGC motif overlapping the putative promoter of the DEG *jhp0832* was changed to a non-methylated GAGC motif, this caused a downregulation of the transcription that was similar to the effect of MTase inactivation, providing strong evidence that methylation of the GCGC motif within a promoter sequence affects gene transcription. Similar findings were previously reported for G^{m6}ACC motifs methylated by the *H. pylori* ModH5 MTase, which are involved in the control of the activity of the *flaA* promoter in strain P12 (40). We note that the introduced point mutation itself (in addition to the absence of methylation) might have an influence on the promoter activity. We thus introduced the mut2 allele into a methylase-deficient strain as a control. However, this strain grew so poorly that reliable qPCR assays could not be performed, so that this possibility cannot formally be ruled out. The exact mechanism(s) how methylated sequence motifs within promoters and most likely also within coding sequences influence gene expression in *H. pylori* is still unknown. One emerging paradigm is exemplified by the essential cell cycle regulator GcrA from *Caulobacter crescentus*, a σ70 cofactor that binds to almost all σ70 promoters, but only induces transcription of genes that harbour G^{m6}ANTC methylated sites in their promoters (66).

The 10 DEGs shared by both strains were less expressed in the absence of methylation. Thus, in contrast to eukaryotes, where CpG methylation in promoter regions leads to the silencing of genes, methylation of GCGC sites in *H. pylori* promoters enhances transcription. Many of the shared DEG belong to conserved cellular pathways (i.e. biotin synthesis, Fe(II) uptake, molybdopterin biosynthesis, bicarbonate and proton production, tRNA modification) and also include a transcriptional regulator involved in copper resistance. Based on these observations, we propose that the conserved GCGC-specific MTase directly controls the expression of those genes involved in various, partially fundamental, cellular pathways.

The inactivation of the MTase caused a substantial growth defect and accelerated conversion to coccoid cells in *H. pylori* J99 that were restored to wild type growth in a complemented strain. The three other wild type strains investigated did not show a similar growth defect when the MTase was inactivated. Other phenotypic effects induced by the MTase inactivation were observed in all or multiple strains. They included functions important for virulence, such as morphology, competence and adherence to gastric epithelial cells. The genome diversity of *H. pylori*, the distribution of motifs among the genomes and the variable methylomes due to the activity of other MTases must influence global gene expression. It was demonstrated recently that deletions of two strain-specific MTases, the ^{m5}C MTase M.HpyAVIB (67) and the ^{m4}C MTase M2.HpyAII (22) both also had regulatory effects on the *H. pylori* transcriptome. While the effects differed widely from those observed for the M.Hpy99III MTase studied here, some genes were differentially regulated by more than one MTase, suggesting that the effects of different MTases may be interlinked. Thus, the strain-specific phenotypes observed in the absence of ^{m5}C methylation in GCGC motifs are likely to reflect the complex and intrinsic diversity of *H. pylori* at the genome, methylome, and transcriptome levels. It is interesting to note that the overrepresentation of GCGC motifs is far more pronounced in coding sequences, and that *H. pylori* has a strong bias for codons overlapping the GCGC motif, such as CGC as the by far most common codon for arginine, and GCG as the second most common codon for alanine (68). The preference of *H. pylori* for these codons may be one reason why a methyltransferase with specificity for GCGC has evolved to serve such a special function.

While we clearly showed that methylation of a GCGC motif overlapping the promoter within the upTSS directly affected transcription, we currently do not understand how the presence or absence of GCGC methylation can affect so many genes in strain J99, and which mechanisms contribute to strain-variable effects. It seems likely that at least some of the massive changes observed in strain J99 are indirect effects, e.g. resulting from the downregulation of genes affecting growth. The effect of MTase inactivation in any given strain is likely to be the net outcome of interlinked direct and indirect regulatory effects that will need to be further elucidated in the future. Methylation may affect DNA topology, which has a strong influence on genome-wide gene regulation, causing secondary effects on the global transcriptome by a plethora of mechanisms. For example, modifications of DNA topology affect the binding of DnaA to the OriC2 of *H. pylori* (69). The *flaA* promoter, whose expression is governed mainly by the transcription factor σ^{28}, was shown by extensive mutagenesis to be strongly modulated in a topology-dependent manner during the growth phase (70). This also fits to the previously described methylation-dependent indirect regulation of the *flaA* promoter (40). Finally, several direct and indirect means of methylation-mediated regulatory mechanisms might not exclude each other, generating an intricate network fine-tuning gene expression, which depends on genome-wide methylation.

CONCLUSION

Global changes in ^{m5}C DNA methylation patterns in *H. pylori* affect the expression of several genes directly or indirectly, which results in both strain-independent (conserved) and strain-dependent effects. Motifs situated within pro-

moter sequences have a direct effect on transcription, while surrounding motifs might modulate the expression indirectly by, for example, altering the topology of the DNA. Furthermore, methylation of GCGC target sequences ensures adequate levels of transcription for numerous genes involved in metabolic pathways, competence and adherence to gastric epithelial cells.

DATA AVAILABILITY

RNA-Seq data was placed in the ArrayExpress database with accession number: E-MTAB-7162

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Sandra Nell for help with assembling the collection of 459 globally representative *H. pylori* genomes, and Gudrun Pfaffinger for excellent technical assistance. We also thank three anonymous reviewers for extremely helpful comments.

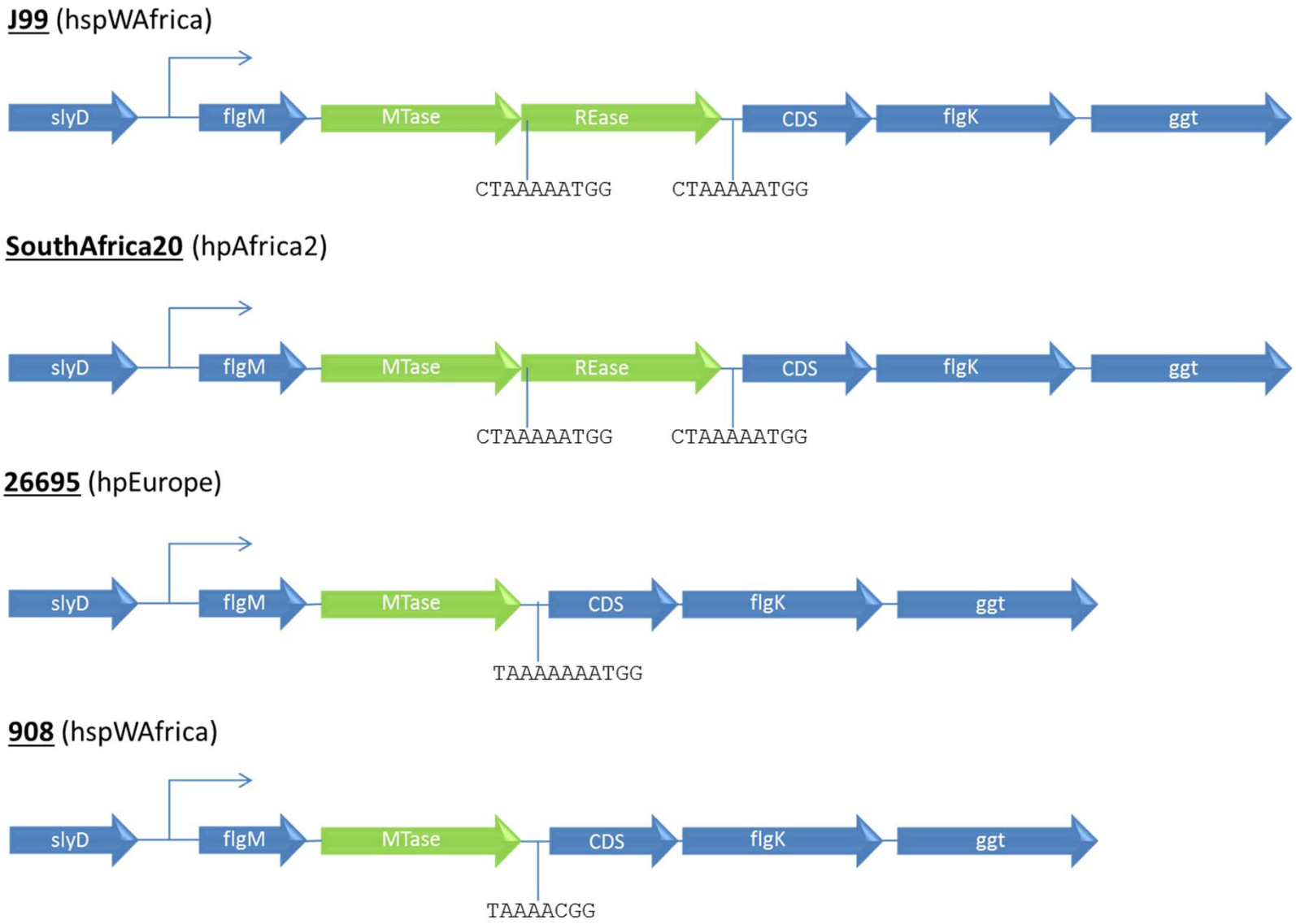
FUNDING

German Research Foundation [SFB 900/A1 and SFB 900/Z1 to S.S. and SFB 900/B6 to C.J.]. Funding for open access charge: German Research Foundation (DFG).
Conflict of interest statement. None declared.

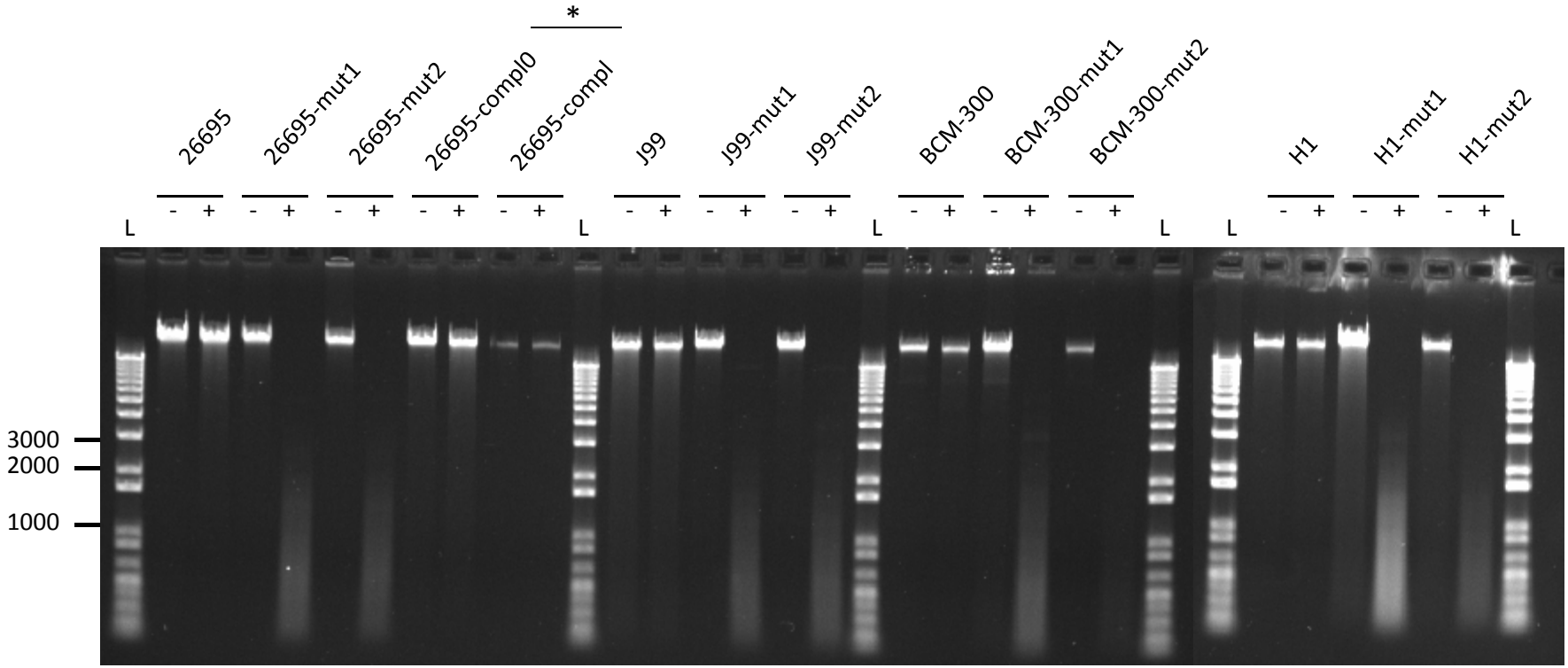
REFERENCES

1. Wilson,G.G. and Murray,N.E. (1991) Restriction and modification systems. *Annu. Rev. Genet.*, **25**, 585–627.
2. Jeltsch,A. (2002) Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *Chem. Bio. Chem.*, **3**, 274–293.
3. Roberts,R.J., Belfort,M., Bestor,T., Bhagwat,A.S., Bickle,T.A., Bitinaite,J., Blumenthal,R.M., Degtyarev,S.K., 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.
4. Wilson,G.G. (1991) Organization of restriction-modification systems. *Nucleic Acids Res.*, **19**, 2539–2566.
5. Loenen,W.A., Dryden,D.T., Raleigh,E.A., Wilson,G.G. and Murray,N.E. (2014) Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res.*, **42**, 3–19.
6. 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.
7. Sanchez-Romero,M.A., Cota,I. and Casadesus,J. (2015) DNA methylation in bacteria: from the methyl group to the methylome. *Curr. Opin. Microbiol.*, **25**, 9–16.
8. Ershova,A.S., Rusinov,I.S., Spirin,S.A., Karyagina,A.S. and Alexeevski,A.V. (2015) Role of Restriction–Modification systems in prokaryotic evolution and ecology. *Biochemistry (Mosc.)*, **80**, 1373–1386.
9. Bubendorfer,S., Krebes,J., Yang,I., Hage,E., Schulz,T.F., Bahlawane,C., Didelot,X. and Suerbaum,S. (2016) Genome-wide analysis of chromosomal import patterns after natural transformation of *Helicobacter pylori*. *Nat. Commun.*, **7**, 11995.
10. Nou,X., Skinner,B., Braaten,B., Blyn,L., Hirsch,D. and Low,D. (1993) Regulation of pyelonephritis-associated pili phasevariation in *Escherichia coli*: binding of the PapI and the Lrp regulatory proteins is controlled by DNA methylation. *Mol. Microbiol.*, **7**, 545–553.
11. Wion,D. and Casadesus,J. (2006) N6-methyladenine: an epigenetic signal for DNA-protein interactions. *Nat. Rev. Microbiol.*, **4**, 183–192.
12. Severin,P.M.D., Zou,X., Gaub,H.E. and Schulten,K. (2011) Cytosine methylation alters DNA mechanical properties. *Nucleic Acids Res.*, **39**, 8740–8751.
13. Messer,W., Bellekes,U. and Lother,H. (1985) Effect of *dam* methylation on the activity of the *E. coli* replication origin, *oriC*. *EMBO J.*, **4**, 1327–1332.
14. Kang,S., Lee,H., Han,J.S. and Hwang,D.S. (1999) Interaction of SeqA and Dam methylase on the hemimethylated origin of *Escherichia coli* chromosomal DNA replication. *J. Biol. Chem.*, **274**, 11463–11468.
15. Kozdon,J.B., Melfi,M.D., Luong,K., Clark,T.A., Boitano,M., Wang,S., Zhou,B., Gonzalez,D., Collier,J., Turner,S.W. *et al.* (2013) Global methylation state at base-pair resolution of the *Caulobacter* genome throughout the cell cycle. *PNAS*, **110**, E4658–E4667.
16. Fox,K.L., Dowideit,S.J., Erwin,A.L., Srikhanta,Y.N., Smith,A.L. and Jennings,M.P. (2007) *Haemophilus influenzae* phasevarions have evolved from type III DNA restriction systems into epigenetic regulators of gene expression. *Nucleic Acids Res.*, **35**, 5242–5252.
17. Srikhanta,Y.N., Dowideit,S.J., Edwards,J.L., Falsetta,M.L., Wu,H.J., Harrison,O.B., Fox,K.L., Seib,K.L., Maguire,T.L., Wang,A.H. *et al.* (2009) Phasevarions mediate random switching of gene expression in pathogenic *Neisseria*. *PLoS Pathog.*, **5**, e1000400.
18. Srikhanta,Y.N., Fox,K.L. and Jennings,M.P. (2010) The phasevarion: phase variation of Type III DNA methyltransferases controls coordinated switching in multiple genes. *Nat. Rev. Microbiol.*, **8**, 196–206.
19. Srikhanta,Y.N., Maguire,T.L., Stacey,K.J., Grimmond,S.M. and Jennings,M.P. (2005) The phasevarion: a genetic system controlling coordinated, random switching of expression of multiple genes. *PNAS*, **102**, 5547–5551.
20. Kahramanoglou,C., Prieto,A.I., Khedkar,S., Haase,B., Gupta,A., Benes,V., Fraser,G.M., Luscombe,N.M. and Seshasayee,A.S. (2012) Genomics of DNA cytosine methylation in *Escherichia coli* reveals its role in stationary phase transcription. *Nat. Commun.*, **3**, 886.
21. Chao,M.C., Zhu,S., Kimura,S., Davis,B.M., Schadt,E.E., Fang,G. and Waldor,M.K. (2015) A cytosine methyltransferase modulates the cell envelope stress response in the cholera pathogen. *PLoS Genet.*, **11**, e1005739.
22. Kumar,S., Karmakar,B.C., Nagarajan,D., Mukhopadhyay,A.K., Morgan,R.D. and Rao,D.N. (2018) N4-cytosine DNA methylation regulates transcription and pathogenesis in *Helicobacter pylori*. *Nucleic Acids Res.*, **46**, 3429–3445.
23. Suerbaum,S. and Michetti,P. (2002) *Helicobacter pylori* infection. *N. Engl. J. Med.*, **347**, 1175–1186.
24. Moodley,Y., Linz,B., Bond,R.P., Nieuwoudt,M., Soodyall,H., Schlebusch,C.M., Bernhoft,S., Hale,J., Suerbaum,S., Mugisha,L. *et al.* (2012) Age of the association between *Helicobacter pylori* and man. *PLoS Pathog.*, **8**, e1002693.
25. Suerbaum,S., Maynard Smith,J., Bapumia,K., Morelli,G., Smith,N.H., Kunstmann,E., Dyrek,I. and Achtman,M. (1998) Free recombination within *Helicobacter pylori*. *PNAS*, **95**, 12619–12624.
26. 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. *PNAS*, **98**, 15056–15061.
27. Didelot,X., Nell,S., Yang,I., Woltemate,S., van der,M.S. and Suerbaum,S. (2013) Genomic evolution and transmission of *Helicobacter pylori* in two South African families. *PNAS*, **110**, 13880–13885.
28. Falush,D., Wirth,T., Linz,B., Pritchard,J.K., Stephens,M., Kidd,M., Blaser,M.J., Graham,D.Y., Vacher,S., Perez-Perez,G.I. *et al.* (2003) Traces of human migrations in *Helicobacter pylori* populations. *Science*, **299**, 1582–1585.
29. Linz,B., Balloux,F., Moodley,Y., Manica,A., Liu,H., Roumagnac,P., Falush,D., Stamer,C., Prugnolle,F., van der Merwe,S.W. *et al.* (2007) An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature*, **445**, 915–918.
30. Moodley,Y., Linz,B., Yamaoka,Y., Windsor,H.M., Breurec,S., Wu,J.Y., Maady,A., Bernhoft,S., Thiberge,J.M., Phuanukoonnon,S. *et al.* (2009) The peopling of the Pacific from a bacterial perspective. *Science*, **323**, 527–530.

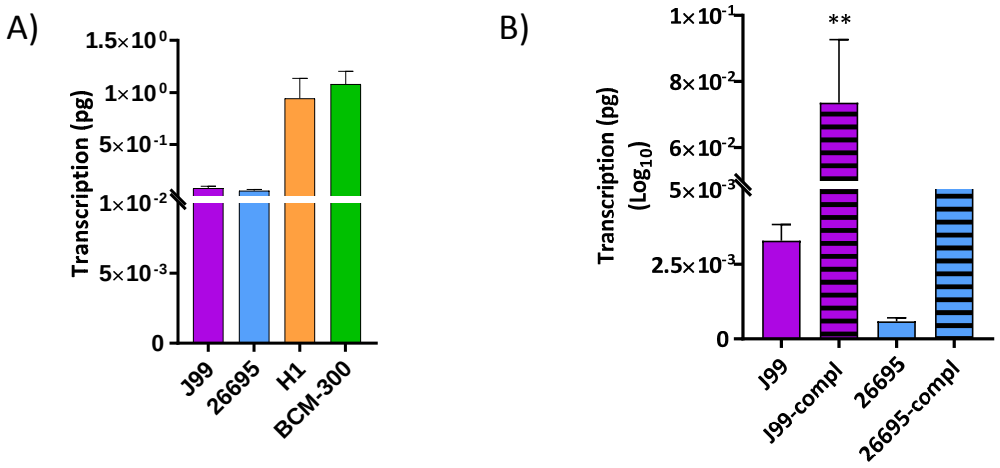
31. Vasu,K. and Nagaraja,V. (2013) Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol. Mol. Biol. Rev.*, **77**, 53–72.
32. Roberts,R.J., Carneiro,M.O. and Schatz,M.C. (2013) The advantages of SMRT sequencing. *Genome Biol.*, **14**, 405.
33. Flusberg,B.A., Webster,D.R., Lee,J.H., Travers,K.J., Olivares,E.C., Clark,T.A., Korlach,J. and Turner,S.W. (2010) Direct detection of DNA methylation during Single-Molecule, Real-Time sequencing. *Nat. Methods*, **7**, 461–465.
34. Krebes,J., Morgan,R.D., Bunk,B., Sproer,C., Luong,K., Parusel,R., Anton,B.P., Konig,C., Josenhans,C. and Overmann,J. *et al*. (2014) The complex methylome of the human gastric pathogen *Helicobacter pylori*. *Nucleic Acids Res.*, **42**, 2415–2432.
35. Nell,S., Estibariz,I., Krebes,J., Bunk,B., Graham,D.Y., Overmann,J., Song,Y., Spröer,C., Yang,I., Wex,T. *et al.* (2018) Genome and methylome variation in *Helicobacter pylori* with a *cag* pathogenicity island during early stages of human infection. *Gastroenterology*, **154**, 612–623.
36. Furuta,Y., Namba-Fukuyo,H., Shibata,T.F., Nishiyama,T., Shigenobu,S., Suzuki,Y., Sugano,S., Hasebe,M. and Kobayashi,I. (2014) Methylome diversification through changes in DNA methyltransferase sequence specificity. *PLoS Genet.*, **10**, e1004272.
37. Lee,W.C., Anton,B.P., Wang,S., Baybayan,P., Singh,S., Ashby,M., Chua,E.G., Tay,C.Y., Thirriot,F., Loke,M.F. *et al.* (2015) The complete methylome of *Helicobacter pylori* UM032. *BMC Genomics*, **16**, 424.
38. Vale,F.F., Megraud,F. and Vitor,J.M. (2009) Geographic distribution of methyltransferases of *Helicobacter pylori*: evidence of human host population isolation and migration. *BMC Microbiol.*, **9**, 193.
39. Srikhanta,Y.N., Gorrell,R.J., Steen,J.A., Gawthorne,J.A., Kwok,T., Grimmond,S.M., Robins-Browne,R.M. and Jennings,M.P. (2011) Phasevarion mediated epigenetic gene regulation in *Helicobacter pylori*. *PLoS One*, **6**, e27569.
40. Srikhanta,Y.N., Gorrell,R.J., Power,P.M., Tsyganov,K., Boitano,M., Clark,T.A., Korlach,J., Hartland,E.L., Jennings,M.P. and Kwok,T. (2017) Methylomic and phenotypic analysis of the ModH5 phasevarion of *Helicobacter pylori*. *Sci. Rep.*, **7**, 16140.
41. Xu,Q., Morgan,R.D., Roberts,R.J. and Blaser,M.J. (2000) Identification of Type II restriction and modification systems in *Helicobacter pylori* reveals their substantial diversity among strains. *PNAS*, **97**, 9671–9676.
42. 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.
43. 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.
44. Kennemann,L., Didelot,X., Aebischer,T., Kuhn,S., Drescher,B., Droege,M., Reinhardt,R., Correa,P., Meyer,T.F., Josenhans,C. *et al.* (2011) *Helicobacter pylori* genome evolution during human infection. *PNAS*, **108**, 5033–5038.
45. Moccia,C., Krebes,J., Kulick,S., Didelot,X., Kraft,C., Bahlawane,C. and Suerbaum,S. (2012) The nucleotide excision repair (NER) system of *Helicobacter pylori*: role in mutation prevention and chromosomal import patterns after natural transformation. *BMC Microbiol.*, **12**, 67.
46. Schweinitzer,T., Mizote,T., Ishikawa,N., Dudnik,A., Inatsu,S., Schreiber,S., Suerbaum,S., Aizawa,S. and Josenhans,C. (2008) Functional characterization and mutagenesis of the proposed behavioral sensor TlpD of *Helicobacter pylori*. *J. Bacteriol.*, **190**, 3244–3255.
47. Huang,S., Kang,J. and Blaser,M.J. (2006) Antimutator role of the DNA glycosylase *mutY* gene in *Helicobacter pylori*. *J. Bacteriol.*, **188**, 6224–6234.
48. Dalia,A.B., McDonough,E. and Camilli,A. (2014) Multiplex genome editing by natural transformation. *PNAS*, **111**, 8937–8942.
49. Sharma,C.M., Hoffmann,S., Darfeuille,F., Reignier,J., Findeiß,S., Sittka,A., Chabas,S., Reiche,K., Hackermüller,J., Reinhardt,R. *et al.* (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature*, **464**, 250–255.
50. Solovyev,V. and Salamov,A. (2011) Automatic annotation of microbial genomes and metagenomic sequences. In: Li,RW (ed). *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*. Nova Science Publishers, pp. 61–78.
51. Vanet,A., Marsan,L., Labigne,A. and Sagot,M.F. (2000) Inferring regulatory elements from a whole genome. an analysis of *Helicobacter pylori* σ80 family of promoter signals. *J. Mol. Biol.*, **297**, 335–353.
52. Schneider,C.A., Rasband,W.S. and Eliceiri,K.W. (2012) NIH image to ImageJ: 25 years of image analysis. *Nat. Methods*, **9**, 671–675.
53. Bönig,T., Olbermann,P., Bats,S.H., Fischer,W. and Josenhans,C. (2016) Systematic site-directed mutagenesis of the *Helicobacter pylori* CagL protein of the Cag type IV secretion system identifies novel functional domains. *Sci. Rep.*,**6**, 38101.
54. Stein,S.C., Faber,E., Bats,S.H., Murillo,T., Speidel,Y., Coombs,N. and Josenhans,C. (2017) *Helicobacter pylori* modulates host cell responses by CagT4SS-dependent translocation of an intermediate metabolite of LPS inner core heptose biosynthesis. *PLoS Pathog.*, **13**, e1006514.
55. Kearse,M., Moir,R., Wilson,A., Stones-Havas,S., Cheung,M., Sturrock,S., Buxton,S., Cooper,A., Markowitz,S., Duran,C. *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.
56. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
57. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2015) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.*, **43**, D298–D299.
58. Han,K.D., Ahn,D.H., Lee,S.A., Min,Y.H., Kwon,A.R., Ahn,H.C. and Lee,B.J. (2013) Identification of chromosomal HP0892-HP0893 Toxin-Antitoxin proteins in *Helicobacter pylori* and structural elucidation of their protein-protein interaction. *J. Biol. Chem.*, **288**, 6004–6013.
59. Azevedo,N.F., Almeida,C., Cerqueira,L., Dias,S., Keevil,C.W. and Vieira,M.J. (2007) Coccoid form of *Helicobacter pylori* as a morphological manifestation of cell adaptation to the environment. *Appl. Environ. Microbiol.*, **73**, 3423–3427.
60. Dorer,M.S., Sessler,T.H. and Salama,N.R. (2011) Recombination and DNA repair in *Helicobacter pylori*. *Annu. Rev. Microbiol.*, **65**, 329–348.
61. Hershberg,R. and Petrov,D.A. (2011) Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet.*, **6**, e1001115.
62. Dossumbekova,A., Prinz,C., Gerhard,M., Brenner,L., Backert,S., Kusters,J.G., Schmid,R.M. and Rad,R. (2006) *Helicobacter pylori* outer membrane proteins and gastric inflammation. *Gut*, **55**, 1360–1361.
63. Hofreuter,D., Odenbreit,S. and Haas,R. (2001) Natural transformation competence in *Helicobacter pylori* is mediated by the basic components of a Type IV secretion system. *Mol. Microbiol.*, **41**, 379–391.
64. Montefusco,S., Esposito,R., D'Andrea,L., Monti,M.C., Dunne,C., Dolan,B., Tosco,A., Marzullo,L. and Clyne,M. (2013) Copper promotes TFF1-mediated *Helicobacter pylori* colonization. *PloS One*, **8**, e79455.
65. Haley,K.P. and Gaddy,J.A. (2015) Metalloregulation of *Helicobacter pylori* physiology and pathogenesis. *Front. Microbiol.*, **6**, 911.
66. Haakonsen,D.L., Yuan,A.H. and Laub,M.T. (2015) The bacterial cell cycle regulator GcrA is a σ70 cofactor that drives gene expression from a subset of methylated promoters. *Genes Dev.*, **29**, 2272–2286.
67. Kumar,R., Mukhopadhyay,A.K., Ghosh,P. and Rao,D.N. (2012) Comparative transcriptomics of *H. pylori* strains AM5, SS1 and their *hpyAVIBM* deletion mutants: possible roles of cytosine methylation. *PLoS One*, **7**, e42303.
68. Lafay,B., Atherton,J.C. and Sharp,P.M. (2000) Absence of translationally selected synonymous codon usage bias in *Helicobacter pylori*. *Microbiology*, **146**, 851–860.
69. Donczew,R., Weigel,C., Lurz,R., Zakrzewska-Czerwinska,J. and Zawilak-Pawlik,A. (2012) *Helicobacter pylorioriC*—the first bipartite origin of chromosome replication in Gram-negative bacteria. *Nucleic Acids Res.*, **40**, 9647–9660.
70. Ye,F., Brauer,T., Niehus,E., Drlica,K., Josenhans,C. and Suerbaum,S. (2007) Flagellar and global gene regulation in *Helicobacter pylori* modulated by changes in DNA supercoiling. *Int. J. Med. Microbiol.*, **297**, 65–81.



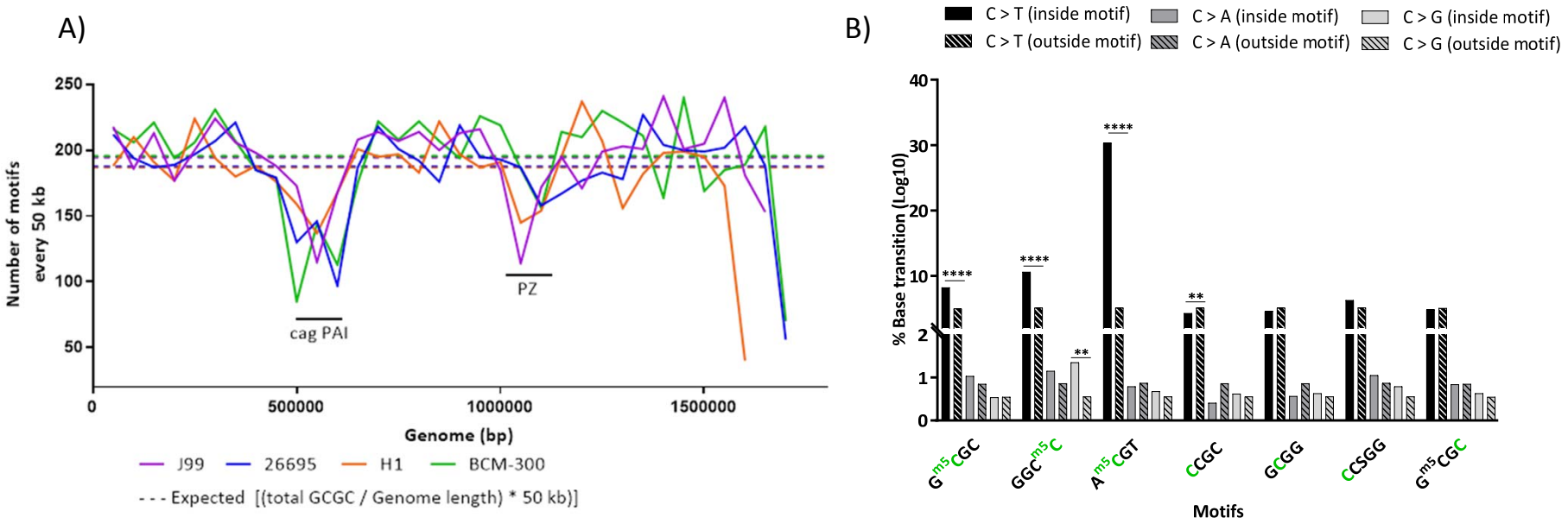
Supplementary Figure 1. Localization of the *jhp1050* MTase gene and flanking genes in the genomes of four *H. pylori* strains. The repeat sequences flanking the REase gene and homologous sequences downstream of orphan MTase genes are shown. Strain ID (bold and underlined), populations (in parentheses) and the predicted promoters (arrows) are also depicted.



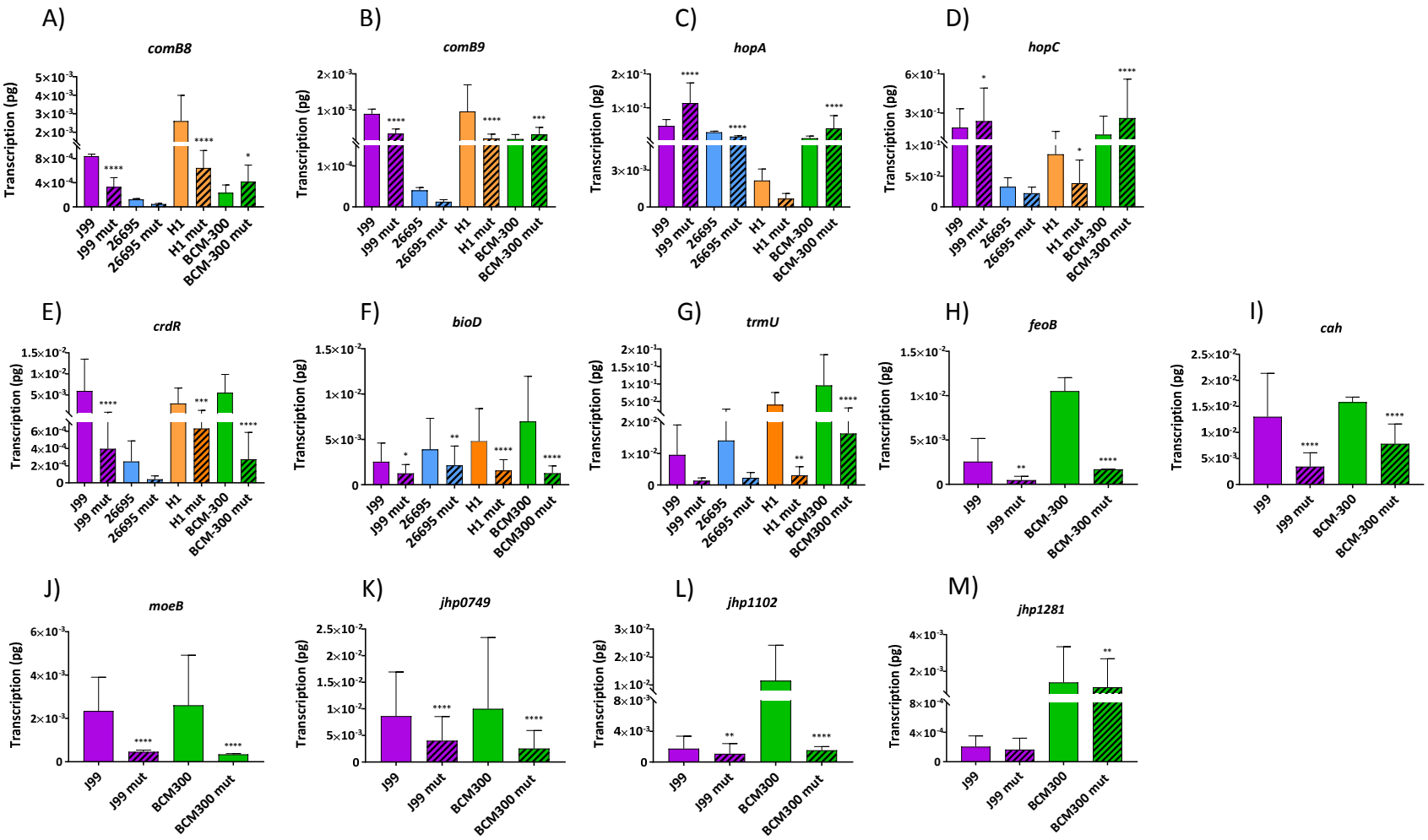
Supplementary Figure 2. Restriction analysis of GCGC methylation state using HhaI restriction enzyme. Wild type strains, two MTase-deficient mutant clones (-mut1/-mut2) and the complemented strains (26695-compl) were tested. Complete restriction of the gDNA can be observed for the mutated strains when the restriction enzyme was added. Methylation of GCGC motifs in the wild type and complemented strains protects the DNA from digestion. L: ladder (1 kb), -: HhaI not added, +: HhaI added.
* Two different versions of the complemented strains were tested (called in this figure –compl0 and -compl). The 26695-compl strain is the complemented strain in 26695-mut expressing the MTase gene under the urease promoter and was used thoughout the manuscript. The construct labelled –compl0 was not further used in the manuscript.



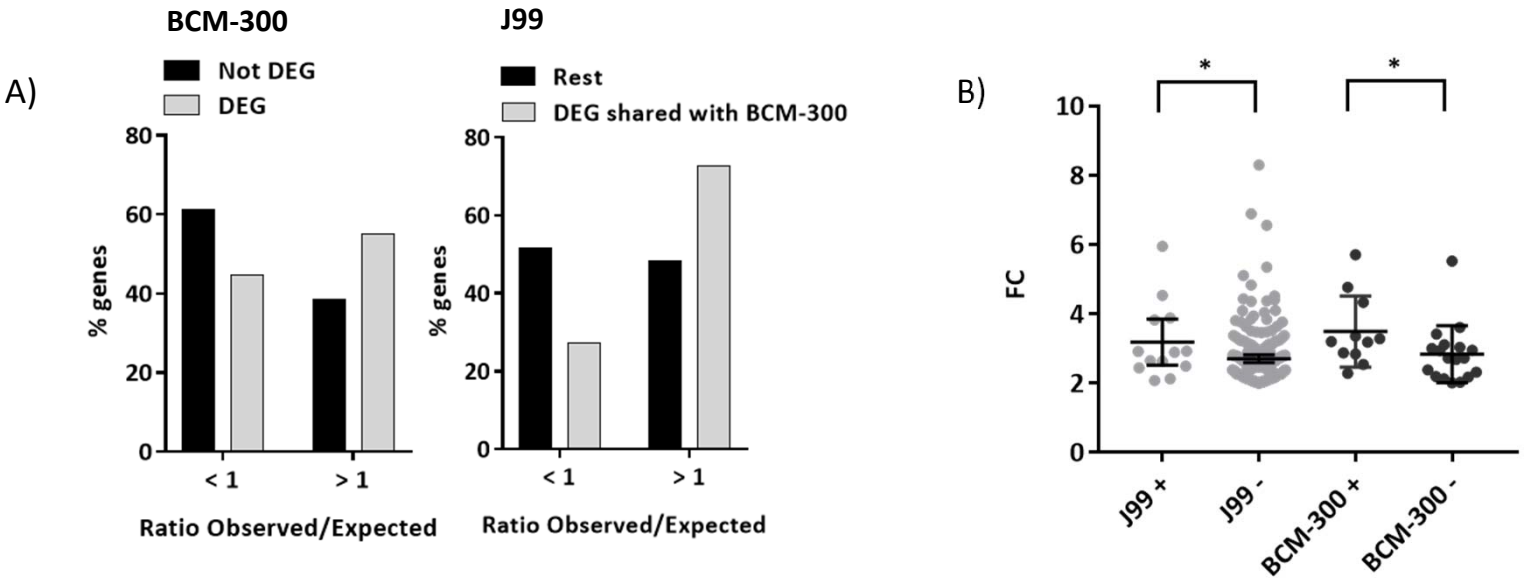
Supplementary Figure 3. qPCRs of transcript amounts of the *jhp1050* MTase gene. A) in the four wild type strains used in the study. B) in the J99 and 26695 wild type strains and respective complementation strains J99-compl and 26695-compl. The transcript amounts are given in pg per µl cDNA and were normalized against the 16S rRNA transcripts in the respective strains, ** ($p < 0.01$).



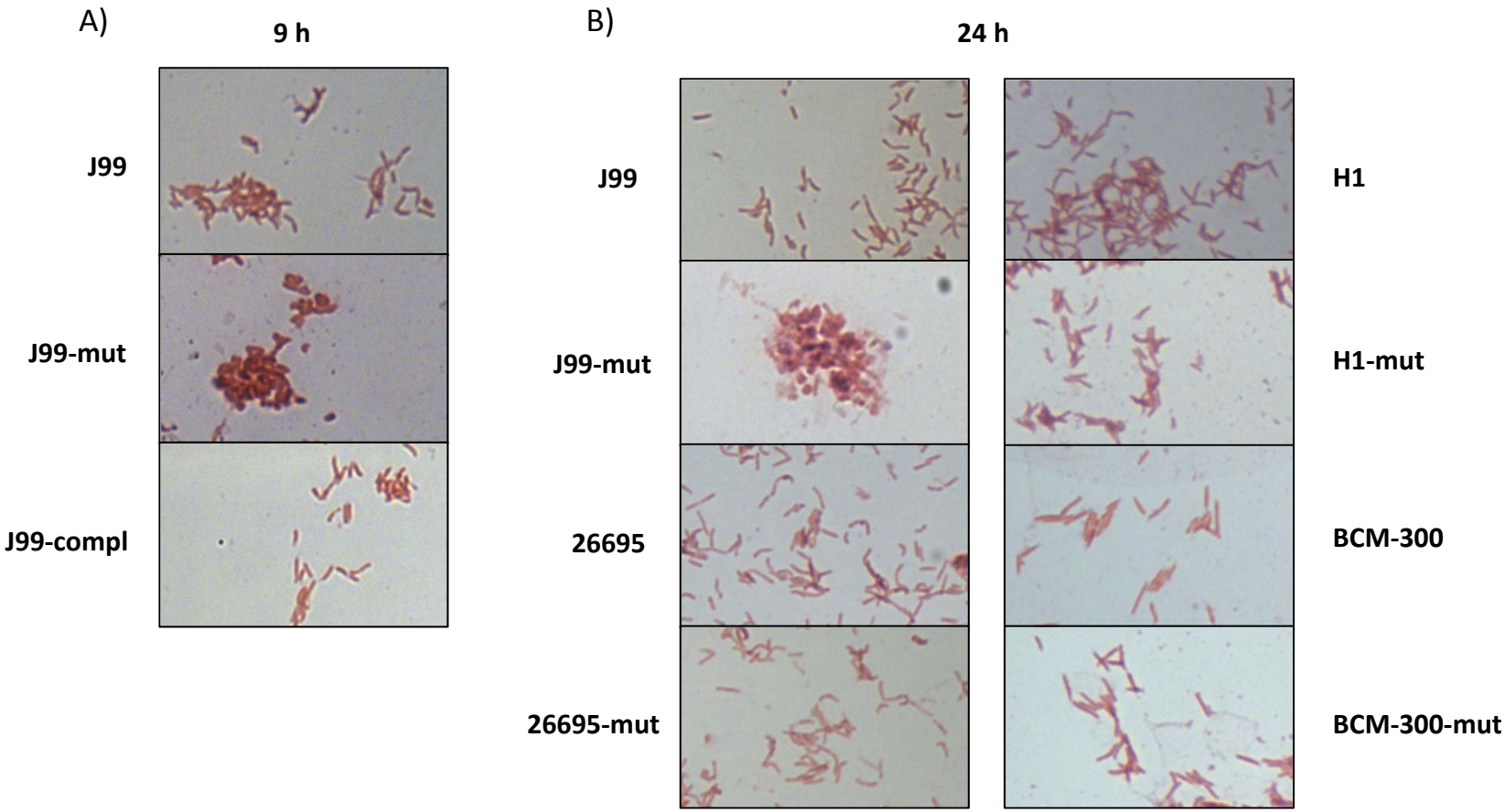
Supplementary Figure 4. Distribution of GCGC motifs in 4 *H. pylori* genomes and frequencies of single base changes for cytosines located within and outside of different methylatable motifs and at methylatable and non-methylatable positions within one motif. A) Distribution of motifs within every 50 kb of genomic sequence. The two regions with a lower density of motifs represent the *cag*PAI and the plasticity zone (PZ). B) Deamination of C to T and changes of C to other bases within several motifs targeted by ^{m5}C MTases and non-^{m5}C methylated motifs. Conversion of methylated and non-methylated Cs to another base that are analysed in each case, are colored in green. Unpaired t-test with Welsh correction, **** ($p < 0.0001$) ** ($p < 0.01$).



Supplementary Figure 5. qPCRs performed to verify RNA-Seq results Transcript amounts are given in pg per µl cDNA and were normalized against 16S rRNA transcript for each strain. Statistics: 2-way ANOVA, * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$), **** ($p < 0.0001$)



Supplementary Figure 6. Percentage of genes with more or less GCGC motifs than expected and the influence of the presence of GCGC motifs within the 50 bp sequence upstream of the TSS (upTSS) on transcription (FC). A) Ratio of observed motifs *vs* expected motifs within DEGs and non-DEGs. A value >1 means more motifs observed than expected, while a value <1 means fewer observed motifs than expected. B) Transcriptional regulation (fold change calculated from RNA-Seq data) for DEGs with GCGC motifs present within the upTSS (+) or without GCGC motifs within the upTSS (-). Test: Mann-Whitney, bars: 95% CI, * ($p < 0.05$).



Supplementary Figure 7. Gram staining of *H. pylori* strains. A) Gram staining after 9 hours of liquid culture. Note that J99-mut entered an early coccoid state when compared with J99-wt and the complemented strain. B) Staining after growth for 24 hours. Note that only J99-mut underwent a morphological change from helical shape to coccoid, while the rest of the strains maintained the helical shape.

Supplementary Table 1. List of the strains analyzed in the study.

The MTase gene (*jhp1050*) and the flanking genes were always found in the genome sequences with the exception of the REase gene (*jhp1049*), which was present only in 61 strains.

NCBI Accession numbers, Strains ID, Strain Population, presence (yes) or absence (no) of the genes; and the origin of the sequences (draft, WGS = whole genome sequence) are displayed on the table.

#	Accession No.	Strain ID	Population	jhp1046	jhp1047	jhp1048	jhp1049	jhp1050	jhp1051	jhp1052	Sequence
1	CBOR01	SA253C	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
2	CBOJ01	SA253A	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
3	AVNI01	SouthAfrica50	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
4	CBMV01	SA29A	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
5	CBQN01	SA160A	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
6	CBQF01	SA160C	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
7	CBPG01	SA172C	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
8	CBPB01	SA47C	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
9	CBOT01	SA47A	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
10	CBNN01	SA40A	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
11	CBQH01	SA169A	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
12	CBPI01	SA34A	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
13	CBOK01	SA169C	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
14	CBQI01	SA36C	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
15	CBPS01	SA174A	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
16	CBPO01	SA175A	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
17	CBOI01	SA175C	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
18	CBQC01	SA251A	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
19	CBPA01	SA166A	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
20	CBOZ01	SA251C	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
21	CBOX01	SA144C	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
22	CBNJ01	SA144A	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
23	CBMW01	SA303C	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
24	CBOH01	SA155C	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
25	CBNZ01	SA155A	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
26	CBPD01	SA233C	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
27	CBMX01	SA233A	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
28	CBPW01	SA194A	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
29	CBNK01	SA194C	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
30	CBPN01	SA40C	hpAfrica2	yes	yes	yes	no	yes	yes	yes	draft
31	CBNY01	SA34C	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	draft
32	NC_022130.1	SouthAfrica20	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	WGS
33	NC_017361.1	SouthAfrica7	hpAfrica2	yes	yes	yes	**yes**	yes	yes	yes	WGS
34	AONK01	UM018	hpAsia2	yes	yes	yes	no	yes	yes	yes	draft
35	AJFA02	NAB47	hpAsia2	yes	yes	yes	no	yes	yes	yes	draft
36	AONL01	UM054	hpAsia2	yes	yes	yes	no	yes	yes	yes	draft
37	AONJ01	NAK7	hpAsia2	yes	yes	yes	no	yes	yes	yes	draft
38	AUSN01	UM067	hpAsia2	yes	yes	yes	no	yes	yes	yes	draft
39	CBQE01	SA222A	hpAsia2	yes	yes	yes	no	yes	yes	yes	draft
40	CBON01	SA222C	hpAsia2	yes	yes	yes	no	yes	yes	yes	draft
41	NC_017376.1	SNT49	hpAsia2	yes	yes	yes	no	yes	yes	yes	WGS
42	NZ_CP006820.1	oki102	hpAsia2	yes	yes	yes	no	yes	yes	yes	WGS
43	NZ_CP006827.1	oki898	hpAsia2	yes	yes	yes	no	yes	yes	yes	WGS
44	NZ_CP006824.1	oki422	hpAsia2	yes	yes	yes	no	yes	yes	yes	WGS
45	NZ_CP006821.1	oki112	hpAsia2	yes	yes	yes	no	yes	yes	yes	WGS
46	NC_017372.1	India7	hpAsia2	yes	yes	yes	no	yes	yes	yes	WGS
47	NZ_CP007605.1	BM012B	hpAsia2	yes	yes	yes	no	yes	yes	yes	WGS
48	NC_022911.1	BM012S	hpAsia2	yes	yes	yes	no	yes	yes	yes	WGS
49	NC_022886.1	BM012A	hpAsia2	yes	yes	yes	no	yes	yes	yes	WGS
50	AMOR01	R030b	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
51	AKPC01	Hp H-11	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
52	AOTW01	Hp A-11	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
53	AKPU01	Hp P-30	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
54	AFAO01	B45	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
55	AKOW01	Hp A-27	hpEurope	yes	yes	yes	**yes**	yes	yes	yes	draft
56	AKOV01	Hp A-26	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
57	AKPQ01	Hp P-16	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
58	AKQJ01	Hp P-15b	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
59	AKPX01	Hp P-74	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
60	AKPR01	Hp P-23	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
61	AKPP01	Hp P-15	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
62	AKOH01	Hp H-27	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
63	AKNW01	NQ4044	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
64	AKNS01	NQ4200	hpEurope	yes	yes	yes	no	yes	yes	yes	draft

65	ABSY01	B128	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
66	CBQM01	SA164C	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
67	CBQK01	SA164A	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
68	CBQJ01	SA165C	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
69	CBQD01	SA165A	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
70	CBPC01	SA171C	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
71	CBOY01	SA172A	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
72	CBOA01	SA171A	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
73	AKNV01	NQ4053	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
74	AKNR01	NQ4216	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
75	CBOF01	SA302A	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
76	CBNA01	SA302C	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
77	CADN01	NQ4191	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
78	CADM01	NQ1671	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
79	CADL01	NQ367	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
80	CADF01	NQ1712	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
81	CADE01	NQ315	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
82	ASYT01	PZ5026	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
83	AKNZ01	NQ4110	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
84	CADK01	NQ4060	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
85	CADJ01	NQ1707	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
86	CADI01	NQ392	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
87	AMOX01	R055a	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
88	AKNT01	NQ4228	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
89	AKNX01	NQ4076	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
90	AWNG01	X47-2AL	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
91	AOTV01	UMB_G1	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
92	AKOQ01	Hp H-45	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
93	AMOV01	R038b	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
94	AMYU01	A45	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
95	CBKZ01	HPARG8G	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
96	AONO01	UM045	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
97	AMOT01	R036d	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
98	AKOT01	Hp A-14	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
99	AJGJ02	NAD1	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
100	CBQA01	SA221A	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
101	CBPT01	SA221C	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
102	AWUL01	CG-IMSS-2012	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
103	AKNY01	NQ4161	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
104	AMOS01	R32b	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
105	AMOY01	R056a	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
106	AMOW01	R046Wa	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
107	AMOU01	R037c	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
108	AMOQ01	R018c	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
109	AKOI01	Hp H-28	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
110	AKOC01	Hp A-9	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
111	AKNU01	NQ4099	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
112	CBOO01	SA37A	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
113	AYHQ01	E48	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
114	AIHW01	P79	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
115	AKOO01	Hp H-43	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
116	CAHX01	N6	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
117	AIHX01	NCTC 11637	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
118	AKPA01	Hp H-9	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
119	CBPK01	SA213C	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
120	CBOU01	SA173A	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
121	CBOM01	SA213A	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
122	CBNG01	SA173C	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
123	JSXX01	173/00	hpEurope	yes	yes	yes	**yes**	yes	yes	yes	draft
124	CCMT01	H3016	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
125	CBOP01	SA37C	hpEurope	yes	yes	yes	no	yes	yes	yes	draft
126	CADG01	NQ352	hpEurope	yes	yes	yes	**yes**	yes	yes	yes	draft
127	NC_014555.1	PeCan4	hpEurope	yes	yes	yes	no	yes	yes	yes	WGS
128	NC_012973.1	B38	hpEurope	yes	yes	yes	no	yes	yes	yes	WGS
129	NC_008086.1	HPAG1	hpEurope	yes	yes	yes	no	yes	yes	yes	WGS
130	NZ_AP014523.1	NY40	hpEurope	yes	yes	yes	no	yes	yes	yes	WGS
131	NZ_CP007604.1	BM013A	hpEurope	yes	yes	yes	no	yes	yes	yes	WGS
132	NZ_CP007606.1	BM013B	hpEurope	yes	yes	yes	no	yes	yes	yes	WGS
133	NC_014256.1	B8	hpEurope	yes	yes	yes	no	yes	yes	yes	WGS
134	NC_017063.1	ELS37	hpEurope	yes	yes	yes	no	yes	yes	yes	WGS
135	NC_017733.1	HUP-B14	hpEurope	yes	yes	yes	no	yes	yes	yes	WGS

136	NC_014560.1	SJM180	hpEurope	yes	yes	yes	no	yes	yes	yes	WGS
137	NZ_CP007603.1	J166	hpEurope	yes	yes	yes	no	yes	yes	yes	WGS
138	NC_011333.1	G27	hpEurope	yes	yes	yes	no	yes	yes	yes	WGS
139	NC_017362.1	Lithuania75	hpEurope	yes	yes	yes	no	yes	yes	yes	WGS
140	NC_000915.1	26695	hpEurope	yes	yes	yes	no	yes	yes	yes	WGS
141	NC_021217.3	UM037	hpEurope	yes	yes	yes	no	yes	yes	yes	WGS
142	NC_011498.1	P12	hpEurope	yes	yes	yes	**yes**	yes	yes	yes	WGS
143	not published	CAM-105A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
144	not published	CAM-117A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
145	not published	CAM-21A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
146	not published	CAM-20A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
147	not published	CAM-9A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
148	not published	CAM-60A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
149	not published	CAM-58A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
150	not published	CAM-40A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
151	not published	CAM-34A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
152	not published	CAM-104A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
153	not published	CAM-77C1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
154	not published	CAM-5A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
155	not published	CAM-120A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
156	not published	CAM-109A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
157	not published	CAM-94A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
158	not published	CAM-188A2	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
159	not published	CAM-48A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
160	not published	CAM-110A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
161	not published	CAM-118A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
162	not published	CAM-189C1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
163	not published	CAM-189A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
164	not published	CAM-153A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
165	not published	CAM-14A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
166	not published	CAM-133A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
167	not published	CAM-30A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
168	not published	CAM-132A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
169	not published	CAM-107A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
170	not published	CAM-25A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
171	not published	CAM-125A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
172	not published	CAM-106A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
173	not published	CAM-96C1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
174	not published	CAM-96A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
175	not published	CAM-37A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
176	not published	CAM-147A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
177	not published	CAM-121A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
178	not published	CAM-144A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
179	not published	CAM-141A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
180	not published	CAM-8A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
181	not published	CAM-67C1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
182	not published	CAM-67A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
183	not published	CAM-50A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
184	not published	CAM-39A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
185	not published	CAM-35A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
186	not published	CAM-33A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
187	not published	CAM-55A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
188	not published	CAM-7A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
189	not published	CAM-77A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
190	not published	CAM-69C1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
191	not published	CAM-69A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
192	not published	CAM-31A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
193	not published	CAM-165C1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
194	not published	CAM-165A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
195	not published	CAM-146A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
196	not published	CAM-140A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
197	not published	CAM-135A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
198	not published	CAM-124A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
199	not published	CAM-116A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
200	not published	CAM-108A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
201	not published	CAM-94C1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
202	not published	CAM-38A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
203	not published	CAM-119A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
204	not published	CAM-22A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
205	not published	CAM-24A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
206	not published	CAM-23A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft

207	not published	CAM-190C1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
208	not published	CAM-190A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
209	not published	CAM-193C1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
210	not published	CAM-193A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
211	not published	CAM-149A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
212	not published	CAM-126A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
213	not published	CAM-41A1	hpNEAfrica	yes	yes	yes	no	yes	yes	yes	draft
214	ALWV01	Sahul64	hpSahul	yes	yes	yes	no	yes	yes	yes	draft
215	not published	PNGhigh12A	hpSahul	yes	yes	yes	no	yes	yes	yes	draft
216	not published	HUI1769	hspAmerind	yes	yes	yes	no	yes	yes	yes	draft
217	NC_019560.1	Aklavik117	hspAmerind	yes	yes	yes	**yes**	yes	yes	yes	WGS
218	NC_017359.1	Sat464	hspAmerind	yes	yes	yes	no	yes	yes	yes	WGS
219	NC_010698.2	Shi470	hspAmerind	yes	yes	yes	no	yes	yes	yes	WGS
220	NC_019563.1	Aklavik86	hspAmerind	yes	yes	yes	no	yes	yes	yes	WGS
221	NC_017740.1	Shi169	hspAmerind	yes	yes	yes	no	yes	yes	yes	WGS
222	NC_017739.1	Shi417	hspAmerind	yes	yes	yes	no	yes	yes	yes	WGS
223	NC_017741.1	Shi112	hspAmerind	yes	yes	yes	no	yes	yes	yes	WGS
224	NC_017358.1	Cuz20	hspAmerind	yes	yes	yes	no	yes	yes	yes	WGS
225	NC_017378.1	Puno120	hspAmerind	yes	yes	yes	no	yes	yes	yes	WGS
226	NC_017355.1	v225d	hspAmerind	yes	yes	yes	no	yes	yes	yes	WGS
227	AUSK01	UM023	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
228	AKHV02	GC26	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
229	JAAA01	HLJ039	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
230	AONM01	UM007	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
231	ANIO01	D33	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
232	AKNJ01	CPY1124	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
233	AUSL01	UM038	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
234	AOTT01	CPY1662	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
235	AKNM01	CPY3281	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
236	AKNK01	CPY1313	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
237	ALKB01	HLJHP271	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
238	CADD01	8A3	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
239	AUSM01	UM065	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
240	AKNL01	CPY1962	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
241	AKHQ02	FD568	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
242	CADC01	BCS100 H1	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
243	AKNN01	CPY6081	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
244	JDVU01	wls-5-5	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
245	AKNQ01	CPY6311	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
246	AKNP01	CPY6271	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
247	ALKC01	HLJHP253	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
248	AKHO02	FD506	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
249	AONN01	UM034	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
250	AUSR01	UM111	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
251	AUSP01	UM085	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
252	AKNO01	CPY6261	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
253	ALKA01	HLJHP256	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
254	AKHR02	FD577	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
255	AUSQ01	UM077	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
256	ABSX01	98-10	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
257	ALJI01	HLJHP193	hspEAsia	yes	yes	yes	no	yes	yes	yes	draft
258	NC_020508.1	OK113	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
259	NC_017367.1	F57	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
260	NC_017368.1	F16	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
261	NZ_CP010013.1	Hp238	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
262	NC_017360.1	35A	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
263	NC_017926.1	XZ274	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
264	AP014711.1	ML2	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
265	NC_017375.1	83	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
266	AP014712.1	ML3	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
267	NC_021882.2	UM298	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
268	NC_021216.3	UM299	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
269	NC_021215.3	UM032	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
270	NC_017366.1	F32	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
271	AP014710.1	ML1	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
272	NC_017382.1	51	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
273	NC_017379.1	Puno135	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
274	NZ_CP006822.1	oki128	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
275	NC_017354.1	52	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
276	NC_020509.1	OK310	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
277	NC_021218.3	UM066	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS

278	NZ_CP006826.1	oki828	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
279	NZ_CP006825.1	oki673	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
280	NZ_CP006823.1	oki154	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
281	NC_017365.1	F30	hspEAsia	yes	yes	yes	no	yes	yes	yes	WGS
282	not published	inma52	hspMaori	yes	yes	yes	**yes**	yes	yes	yes	draft
283	CBOD01	SA46C	hspSAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
284	CBNV01	SA46A	hspSAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
285	CBQG01	SA300A	hspSAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
286	CBPY01	SA158C	hspSAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
287	CBPR01	SA45C	hspSAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
288	CBPM01	SA31C	hspSAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
289	CBPL01	SA300C	hspSAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
290	CBOW01	SA163C	hspSAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
291	CBOQ01	SA210C	hspSAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
292	CBOL01	SA158A	hspSAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
293	CBOG01	SA163A	hspSAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
294	CBNW01	SA210A	hspSAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
295	CBNU01	SA35C	hspSAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
296	CBNE01	SA35A	hspSAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
297	CBNB01	SA45A	hspSAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
298	CBQB01	SA30C	hspSAfrica	yes	yes	yes	no	yes	yes	yes	draft
299	CBPQ01	SA30A	hspSAfrica	yes	yes	yes	no	yes	yes	yes	draft
300	CBPJ01	SA157C	hspSAfrica	yes	yes	yes	no	yes	yes	yes	draft
301	CBNF01	SA157A	hspSAfrica	yes	yes	yes	no	yes	yes	yes	draft
302	CBNH01	SA226A	hspSAfrica	yes	yes	yes	no	yes	yes	yes	draft
303	CBPV01	SA156C	hspSAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
304	CBNR01	SA156A	hspSAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
305	CBQL01	SA170C	hspSAfrica	yes	yes	yes	no	yes	yes	yes	draft
306	CBMZ01	SA170A	hspSAfrica	yes	yes	yes	no	yes	yes	yes	draft
307	CBPZ01	SA146A	hspSAfrica	yes	yes	yes	no	yes	yes	yes	draft
308	CBPF01	SA146C	hspSAfrica	yes	yes	yes	no	yes	yes	yes	draft
309	AKQN01	Hp M1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
310	AKOG01	Hp H-24	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
311	APDL01	GAM245Ai	hspWAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
312	AKPO01	Hp P-13	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
313	AKPN01	Hp P-11	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
314	APDA01	GAM114Ai	hspWAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
315	AKOJ01	Hp H-29	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
316	AKPI01	Hp P-1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
317	APED01	GAM80Ai	hspWAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
318	APDM01	GAM246Ai	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
319	APDF01	GAM120Ai	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
320	AKQF01	Hp H-5b	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
321	AKPV01	Hp P-41	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
322	AKOP01	Hp H-44	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
323	APDD01	GAM118Bi	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
324	APDZ01	GAM268Bii	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
325	APDT01	GAM260ASi	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
326	APDS01	GAM254Ai	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
327	AKPF01	Hp H-21	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
328	APDX01	GAM264Ai	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
329	AKOL01	Hp H-36	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
330	AKOZ01	Hp H-6	hspWAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
331	AKOE01	Hp A-20	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
332	APDW01	GAM263BFi	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
333	AKON01	Hp H-42	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
334	CADH01	NQ1701	hspWAfrica	yes	yes	yes	**yes**	yes	yes	yes	draft
335	AOTU01	CCHI 33	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
336	AKPW01	Hp P-62	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
337	AKPK01	Hp P-3	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
338	AKOM01	Hp H-41	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
339	APER01	HP250ASi	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
340	APEP01	HP250AFiii	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
341	APEM01	GAMchJs136i	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
342	APDQ01	GAM252Bi	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
343	APDP01	GAM250T	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
344	AKOF01	Hp H-16	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
345	AKOA01	Hp A-4	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
346	ANFP01	GAM100Ai	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
347	CBPX01	SA220C	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
348	CBND01	SA220A	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft

349	APDE01	GAM119Bi	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
350	AKPS01	Hp P-25	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
351	AKPM01	Hp P-8	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
352	AKPJ01	Hp P-2	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
353	AKOD01	Hp A-17	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
354	AKOB01	Hp A-5	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
355	CBPH01	SA301A	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
356	CBOS01	SA162A	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
357	CBNX01	SA227A	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
358	CBNT01	SA301C	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
359	CBNQ01	SA162C	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
360	CBNL01	SA227C	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
361	APFC01	HP260Bi	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
362	APEG01	GAM93Bi	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
363	APDU01	GAM260Bi	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
364	AKPD01	Hp H-18	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
365	AKOS01	Hp A-8	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
366	APDH01	GAM210Bi	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
367	APDG01	GAM121Aii	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
368	AOTX01	Hp H-1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
369	AKPT01	Hp P-26	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
370	AKOY01	Hp H-4	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
371	CBPE01	SA161A	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
372	CBOE01	SA214A	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
373	CBNS01	SA214C	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
374	CBNO01	SA215C	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
375	CBNM01	SA161C	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
376	CBNC01	SA216C	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
377	AKOK01	Hp H-30	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
378	AWER01	GAM117Ai	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
379	APEA01	GAM270ASi	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
380	AKPL01	Hp P-4	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
381	AKPH01	Hp H-34	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
382	AKPE01	Hp H-19	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
383	APCY01	GAM105Ai	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
384	AKOU01	Hp A-16	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
385	APDY01	GAM265BSii	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
386	AKPB01	Hp H-10	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
387	APDB01	GAM115Ai	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
388	APCX01	GAM103Bi	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
389	AKQM01	Hp P-28b	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
390	AKOX01	Hp H-3	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
391	APEF01	GAM83T	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
392	APDN01	GAM249T	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
393	APEC01	GAM71Ai	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
394	AKPG01	Hp H-23	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
395	APEI01	GAMchJs106B	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
396	APEB01	GAM42Ai	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
397	APCZ01	GAM112Ai	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
398	APDK01	GAM244Ai	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
399	APDJ01	GAM239Bi	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
400	AKOR01	Hp A-6	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
401	APEH01	GAM96Ai	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
402	APEK01	GAMchJs117Ai	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
403	APEJ01	GAMchJs114i	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
404	APDC01	GAM201Ai	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
405	APDI01	GAM231Ai	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
406	APFB01	HP260BFii	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
407	APDV01	GAM260BSi	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
408	APEL01	GAMchJs124i	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
409	APCW01	GAM101Biv	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
410	CBPU01	SA168A	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
411	CBPP01	SA224C	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
412	CBOV01	SA252C	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
413	CBOC01	SA168C	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
414	CBOB01	SA224A	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
415	CBNI01	SA29C	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
416	CBNP01	SA252A	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
417	not published	CAM-3A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
418	not published	CAM-129A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
419	not published	CAM-122A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft

420	not published	CAM-152A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
421	not published	CAM-148A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
422	not published	CAM-16A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
423	not published	CAM-11A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
424	not published	CAM-44A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
425	not published	CAM-32A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
426	not published	CAM-201C1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
427	not published	CAM-195C1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
428	not published	CAM-195A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
429	not published	CAM-75C24	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
430	not published	CAM-75A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
431	not published	CAM-123A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
432	not published	CAM-130A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
433	not published	CAM-6A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
434	not published	CAM-46A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
435	not published	CAM-57C1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
436	not published	CAM-47A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
437	not published	CAM-173C1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
438	not published	CAM-173A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
439	not published	CAM-10A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
440	not published	CAM-143A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
441	not published	CAM-131A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
442	not published	CAM-127A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
443	not published	CAM-4A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
444	not published	CAM-19A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
445	not published	CAM-43A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
446	not published	CAM-45A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
447	not published	CAM-205C1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
448	not published	CAM-201A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
449	not published	CAM-128A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
450	not published	CAM-36A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
451	not published	CAM-204C1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
452	not published	CAM-204A1	hspWAfrica	yes	yes	yes	no	yes	yes	yes	draft
453	NZ_LT837687.1	BCM-300	hspWAfrica	yes	yes	yes	no	yes	yes	yes	WGS
454	NC_000921.1	J99	hspWAfrica	yes	yes	yes	**yes**	yes	yes	yes	WGS
455	NC_017381.1	2018	hspWAfrica	yes	yes	yes	no	yes	yes	yes	WGS
456	NC_017374.1	2017	hspWAfrica	yes	yes	yes	no	yes	yes	yes	WGS
457	NC_017357.1	908	hspWAfrica	yes	yes	yes	no	yes	yes	yes	WGS
458	NC_017371.1	Gambia94/24	hspWAfrica	yes	yes	yes	no	yes	yes	yes	WGS
459	NC_017742.1	PeCan18	hspWAfrica	yes	yes	yes	no	yes	yes	yes	WGS

Supplementary Table 2. REase gene sequences analysed to predict their activity.
Predicted full length REase was determined as sequences without any stop codon within the CDS.
J99 REase sequence used as reference sequence.
Accession numbers, Strains ID, Strain Population, predicted full length protein (+: yes, -: no)
and the aminoacid length of the protein are on the table.

#	Accession No.	Population	Strain ID	Predicted full length protein	Aminoacids
1	NC_000921.1	hspWAfrica	J99	+	250
2	APDL01	hspWAfrica	GAM245Ai	+	250
3	APED01	hspWAfrica	GAM80Ai	+	250
4	APDA01	hspWAfrica	GAM114Ai	-	179
5	CADH01	hspWAfrica	NQ1701	-	124
6	AKOZ01	hspWAfrica	Hp H-6	+	250
7	NC_022130.1	hpAfrica2	SouthAfrica20	+	250
8	CBPS01	hpAfrica2	SA174A	+	250
9	CBOK01	hpAfrica2	SA169C	+	250
10	CBMV01	hpAfrica2	SA29A	+	250
11	CBPO01	hpAfrica2	SA175A	-	18
12	CBOI01	hpAfrica2	SA175C	-	18
13	CBOR01	hpAfrica2	SA253C	+	250
14	CBOJ01	hpAfrica2	SA253A	+	250
15	AVNI01	hpAfrica2	SouthAfrica50	+	250
16	CBOH01	hpAfrica2	SA155C	-	8
17	CBNZ01	hpAfrica2	SA155A	-	8
18	CBPI01	hpAfrica2	SA34A	-	10
19	CBNK01	hpAfrica2	SA194C	-	10
20	CBPB01	hpAfrica2	SA47C	-	10
21	CBOT01	hpAfrica2	SA47A	-	10
22	CBNN01	hpAfrica2	SA40A	-	10
23	CBPW01	hpAfrica2	SA194A	-	10
24	CBQI01	hpAfrica2	SA36C	-	10
25	CBQC01	hpAfrica2	SA251A	-	10
26	CBPD01	hpAfrica2	SA233C	-	10
27	CBOX01	hpAfrica2	SA144C	-	10
28	CBNJ01	hpAfrica2	SA144A	-	10
29	CBMX01	hpAfrica2	SA233A	-	10
30	CBPG01	hpAfrica2	SA172C	-	10
31	CBQN01	hpAfrica2	SA160A	-	10
32	CBQF01	hpAfrica2	SA160C	-	10
33	NC_017361.1	hpAfrica2	SouthAfrica7	-	8
34	NC_011498.1	hpEurope	P12	+	250
35	CADG01	hpEurope	NQ352	+	250
36	AKOW01	hpEurope	Hp A-27	-	80
37	JSXX01	hpEurope	173/00	+	250
38	NC_019560.1	hspAmerind	Aklavik117	-	125
39	not published	hspMaori	inma52	-	219
40	CBOD01	hspSAfrica	SA46C	-	229
41	CBNV01	hspSAfrica	SA46A	-	229
42	CBNU01	hspSAfrica	SA35C	-	229

43	CBNE01	hspSAfrica	SA35A	-	229
44	CBPV01	hspSAfrica	SA156C	-	208
45	CBNR01	hspSAfrica	SA156A	-	208
46	CBPR01	hspSAfrica	SA45C	-	229
47	CBNB01	hspSAfrica	SA45A	-	229
48	CBQG01	hspSAfrica	SA300A	-	229
49	CBPY01	hspSAfrica	SA158C	-	229
50	CBPM01	hspSAfrica	SA31C	-	229
51	CBPL01	hspSAfrica	SA300C	-	229
52	CBOW01	hspSAfrica	SA163C	-	229
53	CBOQ01	hspSAfrica	SA210C	-	229
54	CBOL01	hspSAfrica	SA158A	-	229
55	CBOG01	hspSAfrica	SA163A	-	229
56	CBNW01	hspSAfrica	SA210A	-	229
57	-	-	Helicobacter cetorum MIT 99-5656	+	250

Supplementary Table 3. Sequences flanking the REase gene and downstream of the MTase gene.
Repeat sequences flanking the REase gene in the REase-carrying strains (superior table)
and downstream of the MTase gene in the non-carrying strains (table below)
Accession numbers, Strains ID, Strain Population, and flanking sequences are on the table

#	Accession No.	Population	Strain ID	Flanking sequence 1	Flanking sequence 2
1	NC_017361.1	hpAfrica2	SouthAfrica7	CAAAATGG	CTAAAAATGG
2	NC_022130.1	hpAfrica2	SouthAfrica20	CTAAAAATGG	CTAAAAATGG
3	NC_000921.1	hspWAfrica	J99	CTAAAAATGG	CTAAAAATGG
4	APDL01	hspWAfrica	GAM245Ai	CTAAAAATGG	CTAAAAACGG
5	CBNE01	hspSAfrica	SA35A	CTAAAAATGG	CTAAAAATGG
6	CBNV01	hspSAfrica	SA46A	CTAAAAATGG	CTAAAAATGG
7	NC_011498.1	hpEurope	P12	CTAAAAATGG	CTAAAATGG
8	CADG01	hpEurope	NQ352	CTAAAAATGG	CTAAAAACGG
9	NC_019560.1	hspAmerind	Akklavik117	CAAAAACTGG	CTAAAAATGG
10	-	-	H. cetorum MIT 99-5656	CTAAAAATGG	CTAAAATGG
11	-	-	H. cetorum MIT 99-5665	CTAAAAATGG	CTAAAATGG

#	Accession No.	Population	Strain ID	Downstream sequence
1	NC_000915.1	hpEurope	26695	TAAAAAAATGG
2	NC_017362.1	hpEurope	Lithuania75	CTAAAAATGG
3	NC_017381.1	hspWAfrica	2018	ATAAAACGG
4	NC_017357.1	hspWAfrica	908	ATAAAACGG
5	CBQB01	hspSAfrica	SA30C	ATAAAACGG
6	CBPJ01	hspSAfrica	SA157C	CTAAAAACGG
7	-	hpNEAfrica	CAM-105A1	CTAAAAATGG
8	-	hpNEAfrica	CAM-117A1	CTAAAAATGG
9	NC_017376.1	hpAsia2	SNT49	CTAAAAATGG
10	NZ_CP006820.1	hpAsia2	oki102	CTAAAAATGG
11	ALWV01	hpSahul	Sahul64	ATAAAATGG
12	-	hpSahul	PNGhigh12A	ATAAAATGG
13	NC_020508.1	hspEAsia	OK113	CTAAAAATGG
14	NC_017367.1	hspEAsia	F57	ATAAAATGG
15	NC_017359.1	hspAmerind	Sat464	ATAAAATGG
16	-	-	H. acinonychis str. Sheeba	ATAAAATGG

Supplementary Table 4. Shared methylated motifs between the four strains tested in the study.
Specificity, type of methylation and presence (+) or absence (-) of methylation.

#	Specificity	26695	J99	H1	BCM-300	Type of methylation
1	**GCGC**	**+**	**+**	**+**	**+**	**^{m5}C**
2	GATC	+	+	+	+	^{m6}A
3	CATG	+	+	+	+	^{m6}A
4	CCTC / GAGG	+	- / +	? /+	? /+	^{m5}C / ^{m6}A

Supplementary Table 5. RNA-Seq results.

List of DEGs (FC > 2, p-adj < 0.01) in J99 (left) and in BCM-300 (right).

The result is based on the upregulation or downregulation of the genes in the mutated strains compared to the wild type.

DEGs shared between both strains are highlited in blue.

J99

#	locus_tag	FC	Regulation mutant	Gene Name
1	jhp_0842	8.296	up	-
2	jhp_1208	6.552	up	-
3	jhp_0623	5.101	up	-
4	jhp_0538	4.833	up	oorB
5	jhp_0424	4.512	up	horE
6	jhp_0548	4.427	up	flaA
7	jhp_0585	4.402	up	putative 3-hydroxyacid dehydrogenase
8	jhp_0206	4.375	up	putative aminotransferase
9	jhp_0043	4.356	up	type II DNA MTase, M.Hpy99V
10	jhp_0207	4.096	up	-
11	jhp_0537	3.941	up	oorA
12	jhp_0051	3.853	up	-
13	jhp_0539	3.764	up	oorC
14	jhp_0656	3.502	up	-
15	jhp_0536	3.454	up	oorD
16	jhp_0632	3.367	up	Predicted N-methylhydantoinase
17	jhp_1321	3.329	up	Histidine and glutamine-rich metal-binding protein
18	jhp_0022	3.306	up	gltA
19	jhp_0008	3.249	up	groEL
20	jhp_0101	3.222	up	dnaK
21	jhp_0212	3.204	up	hopM
22	jhp_0633	3.195	up	hyuA
23	jhp_0576	3.134	up	hyaC
24	jhp_0843	3.132	up	hook assembly protein
25	jhp_1261	3.120	up	hopN
26	jhp_0763	3.025	up	trxA
27	jhp_0263	2.974	up	gppA
28	jhp_0622	2.966	up	-
29	jhp_0334	2.963	up	kgtP
30	jhp_1432	2.934	up	horL
31	jhp_0586	2.900	up	frxA
32	jhp_0849	2.885	up	hopB
33	jhp_0786	2.876	up	hsdM 2 (Type I R-M system)
34	jhp_0804	2.863	up	flgE
35	jhp_0495	2.841	up	cagA
36	jhp_0660	2.835	up	dcuA
37	jhp_0554	2.823	up	hefC
38	jhp_0541	2.816	up	-
39	jhp_0968	2.803	up	cfa
40	jhp_0612	2.790	up	-
41	jhp_0262	2.789	up	Ferredoxin
42	jhp_1209	2.731	up	-
43	jhp_0661	2.719	up	ansB
44	jhp_1093	2.717	up	pgi
45	jhp_0584	2.688	up	-
46	jhp_0214	2.651	up	hopA
47	jhp_0850	2.636	up	hopG
48	jhp_0715	2.636	up	-
49	jhp_0009	2.633	up	groES
50	jhp_1183	2.630	up	nuoC
51	jhp_0764	2.623	up	trxB 1
52	jhp_0610	2.619	up	hemN 1
53	jhp_0768	2.586	up	guaB
54	jhp_0228	2.573	up	napA
55	jhp_0717	2.548	up	-
56	jhp_0575	2.541	up	hyaB
57	jhp_0098	2.495	up	metB
58	jhp_0631	2.480	up	-
59	jhp_1182	2.479	up	nuoB
60	jhp_0469	2.468	up	cag1
61	jhp_1320	2.461	up	hpn, histidine-rich metal binding polypeptide
62	jhp_0848	2.460	up	hopC
63	jhp_0689	2.431	up	fliD
64	jhp_0727	2.407	up	hmcT, cadA
65	jhp_0079	2.397	up	-
66	jhp_0296	2.393	up	-
67	jhp_1038	2.393	up	porB
68	jhp_1352	2.391	up	-
69	jhp_0716	2.389	up	acnB
70	jhp_1106	2.379	up	nupC
71	jhp_0615	2.372	up	aspB
72	jhp_0833	2.348	up	babA
73	jhp_1164	2.334	up	babB

BCM-300

#	locus_tag	FC	Regulation mutant	Gene Name
1	BCM_00032	3.183	down	icdA
2	BCM_00033	3.027	down	-
3	BCM_00034	2.998	down	bioD
4	BCM_00035	2.942	down	bioD
5	BCM_00073	2.724	down	-
6	BCM_00172	2.006	down	-
7	BCM_00173	2.188	down	prtC
8	BCM_00513	4.766	down	mraY
9	BCM_00514	3.105	down	murD
10	BCM_00589	2.123	down	-
11	BCM_00590	2.383	down	gtp1
12	BCM_00591	2.536	down	pepA
13	BCM_00707	4.325	down	feoB
14	BCM_00850	2.835	down	lex2B
15	BCM_00859	3.195	down	-
16	BCM_00860	3.603	down	moeB
17	BCM_00938	5.520	down	-
18	BCM_01111	2.312	down	gluP, fucP
19	BCM_01112	2.281	down	-
20	BCM_01124	3.356	down	cah
21	BCM_01188	2.681	down	holA
22	BCM_01189	2.174	down	vacB
23	BCM_01191	2.867	down	-
24	BCM_01192	2.724	down	oppB
25	BCM_01275	2.938	down	-
26	BCM_01276	5.701	down	trmU
27	BCM_01305	2.022	down	-
28	BCM_01306	3.412	down	-
29	BCM_01307	3.279	down	crdR

74	jhp_0387	2.333	up	pepQ
75	jhp_0073	2.328	up	horA
76	jhp_0382	2.291	up	-
77	jhp_0574	2.282	up	hyaA
78	jhp_1022	2.257	up	putative OMP
79	jhp_0573	2.246	up	-
80	jhp_1094	2.245	up	hofH
81	jhp_0710	2.243	up	-
82	jhp_0126	2.242	up	-
83	jhp_0376	2.232	up	-
84	jhp_0997	2.227	up	-
85	jhp_0102	2.202	up	grpE
86	jhp_0570	2.173	up	dapD
87	jhp_0077	2.167	up	rplM
88	jhp_0888	2.165	up	rdxA
89	jhp_0839	2.161	up	-
90	jhp_1158	2.154	up	carA
91	jhp_1348	2.149	up	-
92	jhp_0348	2.139	up	nixA
93	jhp_1260	2.138	up	tonB 2
94	jhp_0775	2.134	up	-
95	jhp_0553	2.133	up	hefB
96	jhp_0264	2.133	up	waaC
97	jhp_0125	2.126	up	-
98	jhp_1103	2.119	up	hopQ
99	jhp_0074	2.110	up	-
100	jhp_0031	2.109	up	-
101	jhp_1105	2.100	up	deoB
102	jhp_0249	2.098	up	clpB
103	jhp_1349	2.086	up	lpp20
104	jhp_0690	2.067	up	fliS
105	jhp_0099	2.061	up	cysK
106	jhp_0193	2.059	up	mpr
107	jhp_1181	2.058	up	nuoA
108	jhp_1194	2.055	up	nuoN
109	jhp_0684	2.050	up	-
110	jhp_0171	2.037	up	glyA
111	jhp_0844	2.030	up	flgE_2
112	jhp_0589	2.028	up	-
113	jhp_1159	2.024	up	amiF
114	jhp_1047	2.014	up	flgK
115	jhp_0388	2.009	up	folK
116	jhp_1368	2.001	down	kdtB
117	jhp_1278	2.004	down	ubiA
118	jhp_0455	2.007	down	-
119	jhp_1270	2.013	down	-
120	jhp_1420	2.014	down	glmS
121	jhp_1436	2.025	down	-
122	jhp_1410	2.026	down	res 2
123	jhp_0665	2.027	down	-
124	jhp_0186	2.030	down	rpmF
125	jhp_0365	2.053	down	-
126	jhp_0766	2.061	down	-
127	jhp_0751	2.062	down	motA
128	jhp_0812	2.067	down	-
129	jhp_0092	2.072	down	-
130	jhp_1112	2.072	down	cah
131	jhp_0244	2.079	down	Type II MTase, M.Hpy99VII
132	jhp_0482	2.087	down	cagS
133	jhp_1414	2.104	down	-
134	jhp_1355	2.106	down	secreted protein
135	jhp_0191	2.113	down	-
136	jhp_0853	2.121	down	carB
137	jhp_0647	2.122	down	-
138	jhp_0330	2.125	down	-
139	jhp_1165	2.134	down	rpsR
140	jhp_0547	2.169	down	spaB
141	jhp_0885	2.170	down	-
142	jhp_0027	2.172	down	-
143	jhp_1415	2.173	down	exoA
144	jhp_1408	2.185	down	-
145	jhp_0025	2.199	down	bioD
146	jhp_1322	2.216	down	ksgA
147	jhp_1140	2.227	down	purD
148	jhp_1412	2.231	down	recG
149	jhp_0929	2.235	down	-
150	jhp_1493	2.251	down	-
151	jhp_1013	2.254	down	-
152	jhp_1048	2.254	down	-
153	jhp_0468	2.256	down	-
154	jhp_1462	2.257	down	-

155	jhp_0966	2.257	down	-
156	jhp_0412	2.263	down	pyrD
157	jhp_0045	2.283	down	Type II Mtase, M.Hpy99II
158	jhp_0203	2.284	down	-
159	jhp_0299	2.304	down	-
160	jhp_1419	2.351	down	-
161	jhp_1077	2.351	down	-
162	jhp_1422	2.354	down	Type I R-M system S subunit, S.Hpy99XV
163	jhp_1021	2.361	down	-
164	jhp_1279	2.380	down	comEC
165	jhp_0035	2.387	down	comB2
166	jhp_1056	2.389	down	-
167	jhp_1303	2.394	down	-
168	jhp_0241	2.396	down	-
169	jhp_0744	2.398	down	acpS
170	jhp_0935	2.414	down	-
171	jhp_1076	2.415	down	-
172	jhp_0164	2.417	down	Type IV REase, Hpy99McrBP
173	jhp_0444	2.436	down	putative paralog of HpaA
174	jhp_0311	2.451	down	lpxK
175	jhp_0995	2.459	down	-
176	jhp_0115	2.465	down	rpmI
177	jhp_1309	2.469	down	-
178	jhp_1478	2.475	down	-
179	jhp_1102	2.487	down	-
180	jhp_1069	2.541	down	fmt
181	jhp_0064	2.549	down	ureF
182	jhp_0046	2.556	down	Type II REase, Hpy99II
183	jhp_1310	2.566	down	miaA
184	jhp_0818	2.568	down	cysS
185	jhp_0331	2.570	down	-
186	jhp_0116	2.583	down	rplT
187	jhp_0385	2.588	down	-
188	jhp_1318	2.602	down	-
189	jhp_0160	2.647	down	-
190	jhp_0930	2.650	down	-
191	jhp_1134	2.694	down	-
192	jhp_1465	2.710	down	fliE
193	jhp_1012	2.733	down	Type II MTase, M.Hpy99XVIII
194	jhp_0796	2.769	down	-
195	jhp_0880	2.772	down	-
196	jhp_0496	2.776	down	murI
197	jhp_0243	2.783	down	xseA
198	jhp_0693	2.797	down	-
199	jhp_0013	2.850	down	-
200	jhp_1283	2.885	down	crdR
201	jhp_1443	2.914	down	crdR
202	jhp_1179	2.919	down	-
203	jhp_0034	3.046	down	comB1
204	jhp_0971	3.085	down	paralog of HpaA
205	jhp_0940	3.103	down	-
206	jhp_0933	3.134	down	-
207	jhp_0970	3.219	down	-
208	jhp_0441	3.226	down	-
209	jhp_0941	3.283	down	xerCD
210	jhp_0310	3.373	down	flaG
211	jhp_1253	3.467	down	-
212	jhp_1281	3.469	down	-
213	jhp_1477	3.518	down	-
214	jhp_0014	3.628	down	-
215	jhp_0831	3.641	down	toxin of a Type II Toxin-Antitoxin (TA) system
216	jhp_0852	3.765	down	-
217	jhp_0934	3.805	down	-
218	jhp_0749	3.825	down	-
219	jhp_0627	3.880	down	feoB
220	jhp_0165	4.036	down	-
221	jhp_0750	4.086	down	moeB
222	jhp_1254	4.529	down	trmU
223	jhp_0785	5.345	down	Type I R-M system S subunit, S.Hpy99XVI
224	jhp_0832	5.949	down	antitoxin of a Type II Toxin-Antitoxin (TA) system
225	jhp_0951	6.885	down	xerCD

Supplementary Table 6. Plasmids and mutants used and created in the present study.

Ab = antibiotic, amp = ampicillin, aphA3 = kanamycin, CAT = cholamphenicol

#	Plasmid code	Description	Ab cassette
1	pSUS3401	pUC19_BCM300_GCGC-M_aphA3	*amp* , *aphA3*
2	pSUS3402	pUC19_26695_GCGC-M_aphA3	*amp* , *aphA3*
3	pSUS3403	pUC19_J99_GCGC-M_aphA3	*amp* , *aphA3*
4	pSUS3404	pUC19_H1_GCGC-M_aphA3	*amp* , *aphA3*
5	pSUS3406	pUC19_J99_GCGC_RM_aphA3	*amp* , *aphA3*
6	pSUS3411	pADC_ 26695_GCGCcomp_P+	*amp* , *aphA3* , CAT
7	pSUS3413	pADC_ 26695_GCGCcomp_P-	*amp* , *aphA3* , CAT
8	pSUS3426	pUC19_jhp0832	*amp*
9	pSUS3427	pUC19_jhp0832_mut1	*amp*
10	pSUS3428	pUC19_jhp0832_mut2	*amp*
11	pSUS3429	pUC19_jhp0832_mut3	*amp*
12	pUC19	-	*amp*
13	pADC/CAT	pUC19 derivative	*amp* , CAT

#	Description	Plasmid / PCR product for transformation	Ab cassette
1	H. pylori 26695 hp1121::aphA3	pSUS3402	*aphA3*
2	H. pylori BCM-300 GCGC_MTase::aphA3	pSUS3401	*aphA3*
3	H. pylori BCS100-H1 GCGC_MTase::aphA3	pSUS3404	*aphA3*
4	H. pylori J99 jhp1050-49::aphA3	pSUS3406	*aphA3*
5	H. pylori 26695 hp1121::aphA3 complemented (P+)	pSUS3411	CAT, *aphA3*
6	H. pylori 26695 hp1121::aphA3 complemented (P-)	pSUS3413	CAT, *aphA3*
7	H. pylori J99 hp1121::aphA3 complemented (P+)	pSUS3411	CAT, *aphA3*
8	H. pylori J99 hp1121::aphA3 complemented (P-)	pSUS3413	CAT, *aphA3*
9	H. pylori jhp0832 mut 1 rdxA::CAT	pSUS3427 + rdxA::CAT (PCR product)	CAT
10	H. pylori jhp0832 mut 2 rdxA::CAT	pSUS3428 + rdxA::CAT (PCR product)	CAT
11	H. pylori jhp0832 mut 3 rdxA::CAT	pSUS3429 + rdxA::CAT (PCR product)	CAT
12	H. pylori jhp0832 mut 4 rdxA::CAT	pSUS3427 + pSUS3428 + rdxA::CAT (PCR product)	CAT
13	H. pylori jhp0832 mut 6 rdxA::CAT	pSUS3428 + pSUS3429 + rdxA::CAT (PCR product)	CAT

Supplementary Table 7. Oligonucleotide sequences used in the present study.

Lower case letters in the nucleotide sequence indicate the restriction site or the overlapping sequences for overlap PCR.

In the restriction site column, minus means that a restriction site was not added to the primers

#	Primer name	Nucleotide sequence (5' - 3')	Restriction site	Direction	Purpose
1	GCGC_M_PstI_for	atactgcagATTTTAACTTTTATGGATTTTTG	PstI	for	MTase / R-M system inactivation
2	GCGC_M_OL1_rev_A	cggatcggacagtgcTTAACATTTTCAAGCAAGAAAC	-	rev	MTase / R-M system inactivation
3	GCGC_M_OL1_rev_B	cggatcggacagtgcTTAACGTTTTCAAGCAAGAAAC	-	rev	MTase / R-M system inactivation
4	GCGC_M_OL2_for_A	gctcggtagccttgcACTTTTAATATTATTATCAAAGCC	-	for	MTase / R-M system inactivation
5	GCGC_M_OL2_for_B	gctcggtagccttgcACTTTTAAAACCATTATCAAAGCC	-	for	MTase / R-M system inactivation
6	GCGC_M_BamHI_rev	ataggatccCAATCACATTCACGCTCA	BamHI	rev	MTase / R-M system inactivation
7	GCGC_R_EcoRI_for	atagaattcATAATCATACAAGTTCTTTTGG	EcoRI	for	MTase / R-M system inactivation
8	GCGC_R_OL1_rev	cggatcggacagtgcTATATCTTGTCTTGTCGGCT	-	rev	MTase / R-M system inactivation
9	GCGC_R_OL2_for	gctcggtagccttgcTGCTTGCATATGAATTTACC	-	for	MTase / R-M system inactivation
10	GCGC_R_XhoI_rev	atactcgagTCCTGTAGGGTCAATTTTAA	XhoI	rev	MTase / R-M system inactivation
11	jhp0832_PstI_for	atactgcagACTCCCTTTTTTAACCCTC	PstI	for	Site-directed mutagenesis
12	jhp0832_BamHI_rev	ataggatccACTACGGCAAGCTAAACTA	BamHI	rev	Site-directed mutagenesis
13	jhp0832_mut_For1_NEW	ATCTCTTAAAAAAGAGCTTCTTTAGGGGG	-	for	Site-directed mutagenesis
14	jhp0832_mut_Rev1_NEW	TACCGCTTGAATTAACGCCAAGTTTCC	-	rev	Site-directed mutagenesis
15	jhp0832_mut_For2_NEW	TCTATCGCTTGAGCTATAATGAATCG	-	for	Site-directed mutagenesis
16	jhp0832_mut_Rev2_NEW	GCGGTATTAAAAAGGCTAGAAAATT	-	rev	Site-directed mutagenesis
17	jhp0832_mut_For3_NEW	TCCAATTTAAAGAGCTCCATTTAACC	-	for	Site-directed mutagenesis
18	jhp0832_mut_Rev3_NEW	GTGCTTTTTTAATCCTTTGCTTTTGA	-	rev	Site-directed mutagenesis
19	jhp0832_check1	ATCCAAACTCAAAAGCAAAG	-	for	qPCR
20	jhp0832_check2	TCTTAGTGTCCTTAGCCCT	-	rev	qPCR
21	RTqPCR_comB8_1	AACCGAACACCATTTCGTG	-	for	semiqRT-PCR and qPCR
22	RTqPCR_comB8_2	ATGAGCGAACGAGCCAAC	-	rev	semiqRT-PCR and qPCR
23	RTqPCR_comB9_1	ATGCAGGATTTAAACGCCAT	-	for	semiqRT-PCR and qPCR
24	RTqPCR_comB9_2	ATCGCTAATGGGTTTTGAAA	-	rev	semiqRT-PCR and qPCR
25	RTqPCR_comEC_1	ATGATATAGGGGTTTTGGTG	-	for	semiqRT-PCR and qPCR
26	RTqPCR_comEC_2	AGCAATTTAGGGAGTAACGC	-	rev	semiqRT-PCR and qPCR
27	RTqPCR_horA_1	AGGTATGGCATTAGCCTT	-	for	semiqRT-PCR and qPCR
28	RTqPCR_horA_2	ATCGGGGTGTTATTGTTAG	-	rev	semiqRT-PCR and qPCR
29	RTqPCR_horE_1	AGGCTTTGCCAATAAATGGT	-	for	semiqRT-PCR and qPCR
30	RTqPCR_horE_2	ACCTAGAGCGAATTTATCC	-	rev	semiqRT-PCR and qPCR
31	RTqPCR_horL_1	ATGGATGGTAATGGCGT	-	for	semiqRT-PCR and qPCR
32	RTqPCR_horL_2	ATTGGCATGAGCGTAGTCA	-	rev	semiqRT-PCR and qPCR
33	RTqPCR_hofH_1	ACTCACCTATAAACCTCATC	-	for	semiqRT-PCR and qPCR
34	RTqPCR_hofH_2	ATCAGAGCGAAAGCCTGT	-	rev	semiqRT-PCR and qPCR
35	RTqPCR_hopA_1	ATCAAGTCTCAAGCGTTAT	-	for	semiqRT-PCR and qPCR
36	RTqPCR_hopA_2	AGAATATTGGGTTTCGTTGA	-	rev	semiqRT-PCR and qPCR
37	RTqPCR_hopB_1	TGAGCGCTAAGGAATTCACT	-	for	semiqRT-PCR and qPCR
38	RTqPCR_hopB_2	AGTGTTTGTGGTGTTGTTTA	-	rev	semiqRT-PCR and qPCR
39	RTqPCR_hopC_1	ACAACAACACCGGAGGCA	-	for	semiqRT-PCR and qPCR
40	RTqPCR_hopC_2	AAAGAGCGTTCGCAGACAC	-	rev	semiqRT-PCR and qPCR
41	RTqPCR_hopG_1	TGGAGTTCTTTTGGGAGAG	-	for	semiqRT-PCR and qPCR
42	RTqPCR_hopG_2	ATGCCAGCCATAATTGATGA	-	rev	semiqRT-PCR and qPCR
43	RTqPCR_hopQ_1	TGCTCCTTTAAATAGCAAAG	-	for	semiqRT-PCR and qPCR
44	RTqPCR_hopQ_2	ATCAACATGGGGCAATAATC	-	rev	semiqRT-PCR and qPCR
45	RTqPCR_hopN/M_1	ACCCTAATTCCCCAGAAGGT	-	for	semiqRT-PCR and qPCR
46	RTqPCR_hopN/M_2	ATACCTTAACCCCCAATTCCT	-	rev	semiqRT-PCR and qPCR
47	16S_RT1	TTACTAGCGATTCCAGCTTC	-	for	semiqRT-PCR and qPCR
48	16S_RT2	TGAGATGTTGGGTTAAGTCC	-	rev	semiqRT-PCR and qPCR

MIQE Checklist for qPCR Assays:

1. Sample

H. pylori strains were grown in liquid media containing:

- Brain Heart Infusion (BHI) (32 g/L)+ yeast extract (2.5 g/L) - 90%
- Horse Serum (heat-inactivated, autoclaved) - 10%
- Amphotericin B (4 mg/ml) - Dilution of 1:1000
- antibiotic cocktail (Vancomycin 10 mg/L, Polymyxin 3,2 mg/L, Trimethoprim 5 g/L)

Bacteria were grown under microarobic conditions generated in airtight jars 90 (Oxoid, Wesel, Germany) with Anaerocult C gas producing bags (Merck, Darmstadt, Germany), incubated with shaking (37°C, 140 rpm, microaerobic conditions) and grown to an OD_{600} ~ 0.6.

For RNA extraction, 5 ml of the cultures were pelleted (4°C, 6000 x g, 3 min). Pellets were snap-frozen in liquid nitrogen and stored at -80 °C.

2. Nucleic Acids

Bacterial pellets were disrupted with a FastPrep® FP120 Cell Disrupter (Thermo Savant) using Lysing Matrix B 2 ml tubes containing 0.1 mm silica beads (MP Biomedicals, Eschwege, Germany).

RNA was isolated using the RNeasy kit (Qiagen, Hilden, Germany) and on-column DNase digestion with DNase I. A second DNase treatment was carried out according to the manufacturer's instructions using the Ambion TURBO DNA-free™ Kit (Ambion, Kaufungen, Germany).

Isolated RNA was checked for the absence of DNA contamination by PCR reaction using primers for the housekeeping gene *efp* (listed below).

Concentrations of the RNA samples were measured with the NanoDrop 2000 spectrophotometer (Peqlab Biotechnologies).

RNA quality was given as RINe number, measured with an Agilent 4200 Tape Station system using RNA Screen Tapes (Agilent, Waldbronn, Germany). A RIN number > 8 was taken as acceptable quality.

3. Reverse transcription

One µg of total RNA (consistent total amount over all samples) was transcribed into cDNA in a total volume of 40 µl using the SuperScript™ III Reverse Transcriptase (Thermo Fisher Scientific, Darmstadt, Germany).

The protocol was as follows:

Step 1: RNA-Primer Mix

Component	Volume
Total RNA	1 µg
Random Primers (Invitrogen) 3 µg/µl	2 µl (1:60)

A. dest (RNase-free, QIAGEN)	Ad 27 µl

Incubation: 5 min at 65°C, 1 min on ice

Step 2

Setting up the cDNA Synthesis-Mix:

Component	Vol. 1x [µl]
5x First-Strand Buffer ($MgCl_2$), Invitrogen	8
DTT (0,1 M)	2
dNTP (10 mM)	1
RNaseOUT (40 U/µl)	1
SuperScript III RT (200 U/µl)	1

Addition of the 13 µl mix to the RNA-Primer Mix (27 µl)
Incubation: 5 min at room temperature (25°C), 90 min at 50°C
Inactivation: 15 min at 70°C

Step 3

cDNA synthesis was checked via PCR using primers for housekeeping genes.

cDNA was stored at -20 °C.

4. Target

Genomes of the four strains used in this study are available at the NCBI database and they have been published previously in:

- Krebes, J., Morgan, R.D., Bunk, B., Sproer, C., Luong, K., Parusel, R., Anton, B.P., Konig, C., Josenhans, C., Overmann, J. et al. (2014) The complex methylome of the human gastric pathogen Helicobacter pylori. Nucleic Acids Res., 42, 2415-2432.
- Nell, S., Estibariz, I., Krebes, J., Bunk, B., Graham, D.Y., Overmann, J., Song, Y., Spröer, C., Yang, I., Wex, T. et al. (2018) Genome and methylome variation in Helicobacter pylori with a cag Pathogenicity Island during early stages of human infection. Gastroenterology, 154, 612-623.
- 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.
- 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.
- Kennemann, L., Didelot, X., Aebischer, T., Kuhn, S., Drescher, B., Droege, M., Reinhardt, R., Correa, P., Meyer, T.F., Josenhans, C. et al. (2011) Helicobacter pylori genome evolution during human infection. PNAS, 108, 5033-5038.

Gene accession numbers of genes used for qPCR in the two reference strains J99 and 26695 can be easily found in the http://genolist.pasteur.fr/PyloriGene/:

Gene	Locus tag (J99)	Locus tag (26695)
comB8	*jhp_0034*	*hp_0038*
comB9	*jhp_0035*	*hp_0039*
hopA	*jhp_0214*	*hp_0229*
hopC	*jhp_0848*	*hp_0912*
crdR	*jhp_1283, jhp_1443*	*hp_1365*
bioD	*jhp_0025*	*hp_0029*
trmU	*jhp_1254*	*hp_1335*
feoB	*jhp_0627*	*hp_0687*
cah	*jhp_1112*	*hp_1186*
moeB	*jhp_0692*	*hp_0755*
-	*jhp_0749*	*hp_0813*
-	*jhp_1102*	*hp_1102*
-	*jhp_1281*	*hp_1363*

All the RNA-Seq data was placed in the ArrayExpress database with accession number E-MTAB-7162

5. Primers

Primers were designed manually and produced by Metabion International AG (Germany). All the annealing temperatures were between 50°C and 55°C.

Primers are listed here:

Primer name	Nucleotide sequence (5' - 3')	Amplicon Length (bp)	Target gene	Annealing temperature
jhp0832_check1	ATCCAAACTCAAAAGCAAAG	112	*jhp_0832*	50°C
jhp0832_check2	TCTTAGTGTCCTTAGCCCT			
qPCR_crdR_for	AGGATAACGCTACCTTAAAAA	109	*crdR*	51°C
qPCR_crdR_rev	TCATTGAAAAACCTTTTAATGC			
qPCR_bioD_for	ACATGCGCTAGGCTATTAGC	148	*bioD*	54°C
qPCR_bioD_rev	AGCGATCTAAAAGGCGGTTA			
qPCR_moeB_for	AGGGTGTATCCAAGCGAGC	127	*moeB*	54°C
qPCR_moeB_rev	AGGGTTTTTGGGTGCTTGAA			
qPCR_jhp1102_for	AGAAGTTGGGTCATGCGAAG	128	*jhp_1102*	55°C
qPCR_jhp1102_rev	ACAAGCGTAGCCTTATGGGT			
qPCR_trmU_for	TTTAGCCTTGAATGCGATGC	132	*trmU*	53°C
qPCR_trmU_rev	ACCACGCCCTCTTTTTCCA			
qPCR_jhp1281_for	ATCGCAAGGGATTTTTCTCA	117	*jhp_1281*	52°C
qPCR_jhp1281_rev	TAAAGCCACGCTCCCTAAA			

qPCR_jhp1253_for	ATTTGAGCGGCGTTAAGAATT	133	*jhp_1253*	53°C
qPCR_jhp1253_rev	AAAGCGCTCATCAAATCCAAT			
qPCR_jhp0749_for	AAGGCGATTTTAATCACGCA	116	*jhp_0749*	52°C
qPCR_jhp0749_rev	TCTAGCATGAACACATCGTC			
RTqPCR_comB8_1	AACCGAACACCATTTCGTG	108	*comB8*	51°C
RTqPCR_comB8_2	ATGAGCGAACGAGCCAAC			
RTqPCR_comB9_1	ATGCAGGATTTAAACGCCAT	159	*comB9*	50°C
RTqPCR_comB9_2	ATCGCTAATGGGTTTTGAAA			
16S_RT1	TTACTAGCGATTCCAGCTTC	272	16S	55°C
16S_RT2	TGAGATGTTGGGTTAAGTCC			
HPefpF02	GGGCTTGAAAATTGAATTGGGCGG	501	*efp*	53°C
HPefpR01	GTATTGACTTTAATGATCTCACCC			

6. PCR Cycling

qPCR was performed with gene specific primers (listed above) and SYBR Green Master Mix (Qiagen, Hilden, Germany).

Standards for all target genes were prepared in ultrapure H_2O at different concentrations (2 pg, 0.2 pg, 0.02 pg, 2 fg, 0.2 fg (concentrations given per µl)).

Reactions were prepared as follows:

SYBR Green Mastermix	10 µl
Primer for	1 µl
Primer rev	1 µl
H_20	7.5 µl
cDNA	0.5 µl
Final volume	20 µl

The qPCR protocol run for all primers was:

1: 95.0°C for 15:00 min
2: 95.0°C for 0:15 min
3: X for 0:30 min (X determine the temperature of each primers combination)
4: 72.0°C for 0:30 min
Plate Read
5: GOTO 2, 39 more times (40 cycles in total)
6: Melt Curve 60.0°C to 95.0°C: Increment 0.5°C 0:05

7. Data analysis

Reactions were run in a BioRad CFX96 system. Threshold, Cq and Sq values were automatically determined by the instrument. The efficiency calculated by the instrument software was in average higher than 94 % (SD = 13.84).

Standard curves and samples were run as technical triplicates. Cq values were automatically determined by the program CFX using the standards included in every qPCR run. NTCs were predominantly negative (<0.02 pg) and performed with distilled water as negative control (standard curve given automatically by the software).

For quantitative comparisons, samples were normalized to an internal 16S rRNA control qPCR. The controls were run as technical triplicates. The Sq means of the technical triplicates for all runs were normalized to the wild-type strain values (which were set to 1, for each strain and gene separately), in order to create the correction factors. Then, the mean values of the target genes were corrected using the corresponding calculated correction factors for each strain.

6. DISCUSSION

Rapid genetic diversification is a characteristic feature of *H. pylori*. The bacteria are normally acquired during early childhood and, while all infected individuals develop histological gastritis, most do not develop further symptoms. *H. pylori* infection is commonly only detected later in life when the bacteria have established a chronic infection. By the time of writing this thesis, most of the genomic studies have been accomplished using isolates obtained at different intervals of time in chronically infected patients. These studies allowed the calculation of mutation rates and recombination frequencies, and the analysis of the distribution of genetic changes. Nonetheless, very little was known about the genome evolution during the early stages of *H. pylori* infection. *H. pylori* harbors a large number of genes encoding R-M systems and every *H. pylori* strain carries a unique set of R-M systems leading to variable methylomes. Nowadays, the methylome evolution and the functional role of the R-M systems in *H. pylori* are not well understood. In the course of this thesis, I worked on both topics. I studied the genome and methylome evolution of *H. pylori* during early-phase *H. pylori* infection, and I characterized the role of a highly conserved MTase in epigenetic regulation of *H. pylori*. The thesis resulted in two published manuscripts and yet unpublished data. Each manuscript contains detailed discussions of the subjects; therefore, this section aims at integrating these novel findings into the previous knowledge of *H. pylori*'s biology.

6.1. Genome and methylome evolution of *H. pylori* during acute infection

6.1.1. The mutation rates during early-stage of infection are in agreement with the mutation rates in chronic infections

In manuscript I, we analyzed whole-genome sequences of isolates from 12 human volunteers who were given a prophylactic vaccine candidate or placebo and subsequently challenged with a fully virulent *H. pylori* strain (BCM-300). Afterwards, the volunteers were treated with antibiotics to eradicate the infection (Malfertheiner, Selgrad et al. 2018). *In vivo* genome evolution of *H. pylori* and the calculation of the mutation rates from sequential and paired isolates from chronic infections have been studied extensively (Morelli, Didelot et al. 2010, Kennemann, Didelot et al. 2011, Didelot, Nell et al. 2013). The estimated mutation rates during long-term infections were found to be higher than for most of the bacteria analyzed so far. Our whole-genome comparisons of the re-isolates revealed that mutation rates during early-stage (i.e. the first 10-12 weeks) infection were in agreement with those calculated for strains from chronically infected individuals. We did not observe any recombination event, which is plausible, because the volunteers were *H. pylori*-negative until they were challenged

with the BCM-300 strain. In a previous study from our group, two isolates from volunteers who participated in a vaccination trial and were challenged with a *cag*PAI-negative strain called BCS 100 (Aebischer, Bumann et al. 2008) were sequenced using 454 sequencing technology (Kennemann, Didelot et al. 2011). In an ongoing study (Estibariz, Suerbaum *et al*., unpublished data) we sequenced antrum and corpus isolates that were harvested 10 weeks post-infection from volunteers who participated in the same vaccination trial (Aebischer, Bumann et al. 2008). We calculated the average mutation rate as 4.50×10^{-6} mutations per site per year, which was also in agreement with previous estimates of the mutation rate during chronic infection. In contrast to our results, in the only available other study investigating the genetic evolution of *H. pylori* during acute infection, Linz and colleagues (Linz, Windsor et al. 2014) reported a mutation rate 140-fold higher than our estimates, and a high recombination rate. Frequent exchange of DNA usually occurs during mixed infections (Falush, Kraft et al. 2001). Thus, it is likely that in the study conducted by Linz and colleagues there was an ongoing mixed infection. In this particular study, two initially *H. pylori* positive human volunteers were re-infected with *H. pylori* after having received eradication therapy with antibiotics. The success of the eradication therapy was monitored only by UBTs. In our samples, although some UBTs were negative for *H. pylori* infection, we were able to culture bacteria from gastric biopsies. Therefore, in the context of challenge trials, negative UBTs do not exclude the presence of low levels of *H. pylori* infection. It is possible that in the study conducted by Linz *et al*. (Linz, Windsor et al. 2014), the infection was not fully eradicated by the antibiotic treatment and, therefore, the strains analyzed could be part of the preexisting infection. Our results showed no evidence of recombination in the absence of a second unrelated *H. pylori* strain.

In conclusion, the challenge of *H. pylori*-negative human volunteers with a reference *H. pylori* strain in the context of a carefully monitored clinical study, and the subsequent analysis of *H. pylori* isolates directly evolved from the challenge strain permits an accurate investigation of mutation rates during early-stage infection (Kennemann, Didelot et al. 2011, Nell, Estibariz et al. 2018). The major limitations in these studies (manuscript I and unpublished data) are the small number of isolates per individual that we obtained. Moreover, this type of infection studies do not allow the investigation of recombination events due to the absence of mixed infections.

6.1.2. Variation of OMP-related genes and virulence factors during acute infection

Recombination and mutations during chronic infection generate allelic variation in *H. pylori*, which is thought to be important in the adaptation to selective pressures encountered in novel stomach niches after transmission to a new host (Suerbaum and Josenhans 2007). In manuscript I, we observed that several isolates displayed sequence changes within OMP-encoding genes. The same phenomenon was previously reported for chronic infection studies. Several isolates carried mutations within genes of

the *hop/hof/hor* families or switched the activity of the adhesins *sabA* and *sabB* due to phase-variation and intra-chromosomal recombination events in these two adhesins. These observations were in agreement with the results of genome analyses of *H. pylori* isolates from chronically infected individuals that demonstrated that OMP-related genes show a significantly higher tendency to have genetic changes during *in vivo* colonization (Kennemann, Didelot et al. 2011, Krebes, Didelot et al. 2014). In an ongoing study (Estibariz, Suerbaum *et al*., unpublished data); we observed that several isolates contained mutations in *babA*. It was previously shown that there is strong selective pressure affecting the major adhesin BabA during the colonization of humans (Colbeck, Hansen et al. 2006, Nell, Kennemann et al. 2014), Rhesus monkeys (Solnick, Hansen et al. 2004) and rodents (Styer, Hansen et al. 2010).

Modifications in OMP-related genes seem to occur in many bacteria to establish the infection. For example, genetic diversification of adhesins and OMP-related genes was observed in *Salmonella enterica* serovar Typhimurium (Yue, Han et al. 2015) or in *Burkholderia dolosa* (Lieberman, Flett et al. 2014). Quick diversification of OMP-associated genes might be important for *H. pylori* in the adaptation to new stomach niches, or to new host individuals. A recent study from our group has provided support for this hypothesis: the study analyzed gastric biopsies from three stomach regions of 16 *H. pylori*-infected individuals. The results showed that there is an association between gene polymorphisms affecting motility, chemotaxis and OMPs, and the adaptation to different stomach parts (antral and oxyntic mucosa) (Ailloud, Didelot et al. 2019).

In manuscript I, we also observed changes in two major virulence factors of *H. pylori*, the *cag*PAI and VacA. Three of the 12 re-isolates lost *cag*PAI function due to frameshift mutations in *cagY*, or the insertion of a mobile element in *cagE*. In addition, two isolates containing non-synonymous SNPs in *cagA* and *cagW* showed a reduction in IL-8 induction and another isolate carried additional *cagA* copies. Thus, at least in the context of the strain BCM-300, we observed a selection against *cag*PAI function in multiple individuals. It was reported before that the *cag*PAI could be partially or completely lost during chronic infections (Bjorkholm, Lundin et al. 2001, Kraft, Stack et al. 2006, Ailloud, Didelot et al. 2019). Modifications in the *cag*PAI and its ability to induce IL-8 may have an impact on adaptation to new hosts by modulating the inflammatory response of the gastric mucosa.

The ability to produce VacA was abrogated in three isolates from the vaccine group due to stop codons in the gene sequence. Although diverse allelic variants of *vacA* displaying different toxicities and the loss of *vacA* activity have been already identified *in vivo* (Falush, Kraft et al. 2001, Aviles-Jimenez, Letley et al. 2004), VacA inactivation in these three isolates from the vaccination group possibly occurred as a response to vaccine-induced selective pressure. Interestingly, VacA has been reported to be an important factor in *H. pylori* colonization of animal models, although it was not completely required

for colonization (Wirth, Beins et al. 1998, Salama, Otto et al. 2001, Winter, Letley et al. 2014). Thus, inactivation of VacA activity could affect *H. pylori* colonization, but it might also be a way to evade the immune-induced response caused by the vaccine.

The study reported in manuscript I, is one of the very first studies (Kennemann, Didelot et al. 2011, Linz, Windsor et al. 2014) investigating genome adaptation of *H. pylori* during early-stage infection. We showed that genetic changes affecting OMP-related genes and virulence factors occurred early in *H. pylori* infection, potentially contributing to the rapid adaptation of this pathogen to a novel gastric niche. Thus, early modulation of adhesion and virulence factor activity might maintain a balance between the pathogenicity of the bacteria and the immune response of the host, favoring *H. pylori* to establish a chronic infection.

6.1.3. Vaccine-induced modulation of virulence factors

The rapid increase of bacterial resistance to antibiotics is a major concern. There is a need for developing novel approaches to stop bacterial infections. Vaccines against *H. pylori* could end the transmission and produce a decrease in antibiotic resistance since fewer antimicrobials would be prescribed. For example, a vaccine against *Streptococcus pneumoniae* was shown to be effective in immunocompetent patients and reduced the rate of pneumococcal infections (Shapiro, Berg et al. 1991, Daniels, Rogers et al. 2016).

In manuscript I, *H. pylori* isolates were obtained from human volunteers that were given a prophylactic vaccine candidate or placebo. The vaccine was composed of three recombinant *H. pylori* antigens (VacA, CagA, and NAP, a neutrophil-activating protein) (Malfertheiner, Selgrad et al. 2018). Although well tolerated and capable of inducing an immune response in the volunteers, the vaccine was ineffective against *H. pylori* infection. Our whole-genome analysis showed that three of the seven isolates from the vaccine group displayed premature stop codons in *vacA*. In contrast, the analysis of whole genomes of isolates from another vaccination and *H. pylori* challenge study (Aebischer, Bumann et al. 2008) did not show modifications within these major virulence factors (Estibariz, Suerbaum *et al*., unpublished data). The vaccine used in this study was a *Salmonella* Typhi Ty21a strain expressing *H. pylori* urease. Thus, despite the small cohort available to us the use of VacA in the prophylactic vaccine (manuscript I) likely led to a selection of isolates with a disrupted *vacA* gene to avoid the immune response.

In both vaccine studies (manuscript I and Estibariz, Suerbaum *et al*., unpublished data), we showed that OMP-related genes and LPS were prone to modifications, suggesting that these genetic changes might help *H. pylori* escape the immune system. Immune system avoidance due to vaccine-related selective pressure has been reported in other microorganisms. For example, despite the effectivity of the Hib vaccine preventing *Haemophilus influenzae* infections, the wide usage of the vaccine caused

the emergence of capsular polysaccharide variants, one of the main virulence factors of this bacterium. Modifications within this capsular polysaccharide resulted in non-typeable pathogenic *H. influenzae* serotypes (Agrawal and Murphy 2011).

The development of a successful vaccine is challenging due to the great genetic diversity between *H. pylori* strains, which are able to adapt quickly to novel niches. Although a vaccine would be very beneficial to stop the transmission, especially in countries where the infection rate is very high with a higher incidence of gastric malignancies, it seems that there are no current efforts by the pharmaceutical industry to develop novel vaccines against *H. pylori*. A recent review by Sutton & Boag summarizes the status of the *H. pylori* vaccine development programs from the past years (Sutton and Boag 2018). Most of the studies stopped after preclinical or phase I stages with the exception of one vaccine candidate that reached phase III. The recombinant vaccine was given to children and followed-up to 1 and 3 years. The authors observed a reduced *H. pylori* infection rate (Zeng, Mao et al. 2015, Sutton and Boag 2018). Despite this reduction, the results should be confirmed after longer periods.

In conclusion, the rapid genetic evolution of *H. pylori* likely contributes to the ability of the bacteria to escape from the action of the vaccines tested so far. However, the exact mechanisms leading to vaccine failure have not been elucidated. Thus, to develop a successful vaccine candidate, more efforts are needed to understand how *H. pylori* genetic variability contributes to the avoidance and modulation of the immune system.

6.2. The role of methylation in *H. pylori*

6.2.1. Discovery of novel R-M systems and methylomes

R-M systems are almost ubiquitous in bacteria. Indeed, DNA methylation was found in 93% of bacterial and archaeal genomes (Blow, Clark et al. 2016). In manuscript I, we obtained the complete methylomes of the challenge strain BCM-300 and all the re-isolates analyzed in the study. We discovered 15 methylated motifs and of those, 11 were assigned to already known MTase genes. Inactivation of candidate genes and subsequent SMRT® sequencing of the mutants allowed the discovery of three novel R-M systems. In addition, we studied the methylome of the BCS 100 challenge strain and the re-isolates (Estibariz, Suerbaum *et al*., unpublished data). Here, we identified 24 methylated motifs corresponding to 22 active Type II and Type III R-M systems, and of those, three were novel R-M systems not characterized before. Our findings about the methylomes of BCM-300 and BCS 100 strains were in agreement with previous studies in *H. pylori*, where we observed that *H. pylori* strains harbor a large number of R-M systems. The number of R-M systems found in bacteria usually shows a positive correlation with the genome size. *Helicobacter* and *Campylobacter* species are an exception to this, because they contain a surprisingly large amount of R-M systems despite their small genomes (Vasu

and Nagaraja 2013). Thus, while most bacteria harbour two to six R-M systems (Lluch-Senar, Luong et al. 2013, Vasu and Nagaraja 2013, Fischer, Römling et al. 2019), *H. pylori* strains can hold up to 30 R-M systems (Vasu and Nagaraja 2013).

Since the development of SMRT® sequencing technology, we and other researchers have been able to characterize the methylomes of many different *H. pylori* strains (Krebes, Morgan et al. 2014, Lee, Anton et al. 2015, Roberts, Vincze et al. 2015, Lamichhane, Chua et al. 2019) demonstrating how complex the methylome of *H. pylori* is. Our comparison of the methylomes of J99, 26695, BCM-300 and BCS 100 showed that only three motifs (GCGC, GATC and CATG) were shared between the four strains. Previously, Vale *et al*. analyzed 221 *H. pylori* gDNA samples for their susceptibility to cleavage by 29 REases (Vale, Megraud et al. 2009). The authors observed very few MTases common to the majority of the strains and only one motif (GCGC) present in all. The great diversity of R-M systems and orphan MTases in *H. pylori* suggests they must have another role a part of self-DNA protection.

6.2.2. Phase-variable MTases are responsible for changes in the methylome of *H. pylori* isolates

H. pylori methylome studies have uncovered that every strain shows a distinct methylation pattern, but as far as we know, there is no information about methylome evolution *in vivo* in this gastric pathogen. In manuscript I, we dissected for the first time methylome modifications during early-stage human *H. pylori* infection and we observed that differences were due to phase-variable MTase genes with homopolymeric tracts. Similar results were obtained for the isolates of human volunteers challenged with the BCS 100 strain (Estibariz, Suerbaum *et al*., unpublished data). In both studies, we found two different phase-variable R-M systems. In manuscript I, we identified one phase-variable R-M system gene with two homopolymeric tracts. Alterations in the number of nucleotides within the first tract affected the activity (ON/OFF) of the R-M system. Changes in the second repetitive sequence led to the methylation of different motifs. The R-M system (Hpy300XI) was homologous to HpyAXVI (*H. pylori* 26695) and Hpy99XIV (*H. pylori* J99-R3) (Krebes, Morgan et al. 2014). This frameshift-mediated sequence specificity switch seems to be a particularity of *H. pylori*. At least to my knowledge, by the time writing this thesis, it has not been discovered in other bacteria.

Phase-variable R-M systems have been described in many other bacteria. Srikhanta and colleagues described what is known as “phasevarion” (Srikhanta, Maguire et al. 2005). Phasevarions are groups of genes whose expression is jointly affected by a reversible switch of phase-variable MTase genes. The authors studied the role in the phenotype and transcriptome of a phase-variable Type III MTase of *H. influenzae*. Several genes were differentially expressed between the wild type and a mutant strain with an inactive MTase gene, which affected many phenotypic traits.

Every MTase gene prone to phase-variation could be a potential regulator of the expression of several genes. Phasevarions are found in multiple bacterial pathogens and contribute to their virulence and adaptation. In *M. catarrhalis*, different phase-variable MTases have been proposed to modulate the expression of genes involved in colonization and defence against host immunity and proposed an association with otitis media (Blakeway, Power et al. 2014). The reversible switching of ModA2 in *H. influenzae* resulted in different abilities to fight oxidative stress and resist neutrophil-mediated killing (Brockman, Branstool et al. 2017). In pathogenic *Neisseria* species, phasevarions have been shown to affect the expression of genes coding for virulence factors and vaccine antigens (Seib, Jen et al. 2017). Likewise, phasevarions have been studied *in vitro* in *H. pylori*. The ModH5 MTase activity was found to control the expression of the outer membrane gene *hopG*. Motility was also affected by the regulation of the expression of the *flaA* gene encoding the Flagellin A in the strain P12 (Srikhanta, Gorrell et al. 2011, Srikhanta, Gorrell et al. 2017).

Based on these previous studies, it is possible to speculate that phasevarions might influence *in vivo* adaptation of *H. pylori* to novel niches. So far, we did not investigate the phasevarions associated to the phase-variable MTases discovered in manuscript I and in the ongoing project, but this will be part of future investigations.

6.2.3. A very highly conserved MTase found in all *H. pylori* strains

When comparing the *H. pylori* methylomes available, we observed that one methylated motif (GCGC) described by Vale and colleagues (Vale, Megraud et al. 2009) was present in all *H. pylori* methylome studies (Krebes, Morgan et al. 2014, Lee, Anton et al. 2015, Nell, Estibariz et al. 2018). Therefore, in manuscript II we aimed to characterize the highly conserved MTase (JHP1050) found to be present in all *H. pylori* methylomes analyzed so far.

We discovered that the MTase gene (*jhp1050*) was present in the 459 *H. pylori* genomes analyzed in the study and, based on the nucleotide sequence, were predicted to be active in all strains. In contrast, the REase gene (*jhp1049*) was present only in 61 strains and predicted to be functional in 15 of these. A 10 bp repeat sequence was identified flanking the REase gene. The same sequence was observed downstream of the MTase gene and 48 bp upstream of *jhp1048* in REase-negative strains. The sequence contained a homopolymeric region with a variable number of adenines. Recombination between the repeat sequences flanking the REase gene might have been responsible for the gene excision. The fact that the strains carrying the REase had mostly an African ancestry and that the MTase was found to be present and active in all the strains, suggested that the R-M system is an ancient system acquired early in the history of *H. pylori* and that the REase was lost likely before the modern humans left Africa. The phylogenetic trees of the MTase and REase gene sequences performed in

manuscript II clustered the strains into geographical populations according to the MLST genes. MLST of seven housekeeping genes has been extensively used to genotype *H. pylori* into geographical populations (Achtman, Azuma et al. 1999, Nell, Eibach et al. 2013, Secka, Moodley et al. 2014). Our phylogenetic analysis of the *jhp1050* MTase gene demonstrated its very high conservation in *H. pylori*, a rare exception among genes belonging to R-M systems.

Methylated GCGC motifs are not exclusive to *H. pylori*. The search of the motif in the REBASE database (Roberts, Vincze et al. 2015) exhibited many other species whose methylomes contain this motif or similar variants (i.e. RGCGCY, WGCGCD, TGCGCA) including *Campylobacters* and other *Helicobacter* species. Furthermore, in eukaryotes, methylation usually occurs in CpG dinucleotides and plays a pivotal role in cell differentiation and gene silencing (Moore, Le et al. 2013, Jang, Shin et al. 2017).

In this way, the high conservation of cytosine methylation within GC sequences might reflect its importance in the biology of both prokaryotes and higher organisms.

6.2.4. ^{m5}C-Methylation regulates gene expression and the phenotype of *H. pylori*

Although R-M systems have been described as "primitive immune systems" in bacteria, the role of methylation in modulating gene expression in prokaryotes is extensively recognized. ^{m6}A-Methylation is the most common type of modification in bacterial genomes, and consequently, its implication in epigenetic regulation has been a focus for many years (Messer, Bellekes et al. 1985, Kang, Lee et al. 1999, Kozdon, Melfi et al. 2013). In manuscript II, we dissected the role of ^{m5}C-methylation of GCGC motifs in the transcriptome of two different *H. pylori* strains and their isogenic MTase mutants. So far, very few studies attempted to understand the role of cytosine methylation in the biology of *H. pylori* or any other bacterium. The role of ^{m5}C-methylation in the transcriptome of *H. pylori* was investigated using microarrays. The deletion of one ^{m5}C-MTase gene in two *H. pylori* strains led to significant differences in their transcriptomes (Kumar, Mukhopadhyay et al. 2012). Later, the regulatory role of ^{m4}C-methylation in transcription in one *H. pylori* strain was shown using more quantitative techniques like RNA-seq (Kumar, Karmakar et al. 2018). Our transcriptome results in manuscript II showed that the role of ^{m5}C-methylation differed significantly between strains. While the absence of methylation modified the expression of 225 genes in strain J99, only 29 genes were differentially regulated when GCGC methylation was inactivated in strain BCM-300. Of those, 10 were shared between the two strains. We observed that differential expression of OMP-related genes and competence genes had an impact on cell adhesion and natural DNA uptake, respectively. Furthermore, the downregulation of the response regulator *crdR* in the absence of methylation had a direct impact on the resistance to copper. Interestingly, the strain J99 lacking GCGC methylation entered a coccoid state leading to a growth defect in liquid cultures. Thus, changes in the transcriptomes resulted in strain-specific and

conserved phenotypes. The genetic diversity between *H. pylori* strains and the distribution of G^{m5}CGC motifs among the genomes could contribute to the differences observed between the strains. Besides, the activity of other strain-specific MTases could also have an impact on gene expression.

We observed that differential gene expression was associated with the presence of three or more motifs within coding sequences or with motifs within regulatory regions. It was described before in *V. cholerae*, that the number of motifs within coding sequences was correlated with changes in gene expression (Chao, Zhu et al. 2015). Therefore, to my knowledge, our study is one of the first exploring the link between the amount of motifs within coding sequences and regulatory regions with changes in gene expression.

We showed experimentally that one methylated motif overlapping the promoter sequence had a direct impact on gene expression. Using site-directed mutagenesis targeted methylatable GCGC motifs were modified to non-methylatable GAGC motifs. We observed a direct impact on gene expression of the selected gene when the GCGC motif overlapping the putative promoter was modified, as it happened with the J99 mutant strain lacking the MTase gene. In contrast, the modification of a single motif upstream of the putative promoter or within the coding sequence did not affect the transcription of the target gene. Thus, we were able to confirm that the methylation of target motifs within regulatory regions can have a direct impact on gene expression. Similarly, the expression of FlaA was modulated by direct methylation of target sites within the *flaA* promoter (Srikhanta, Gorrell et al. 2017). DNA methylation of promoter sequences has been shown to regulate the progression of the cell cycle in *C. crescentus* (Gonzalez, Kozdon et al. 2014, Lina, Alzahrani et al. 2014) and DNA replication in *E. coli* (Messer, Bellekes et al. 1985, Kang, Lee et al. 1999) by competition with DNA binding proteins. In manuscript II we did not investigate whether the exact same mechanism is behind this direct regulation of gene expression. Furthermore, the great amount of DEGs and the fact that very few genes contain motifs within regulatory regions, points to an indirect regulation of the transcriptome by methylation. Indirect regulation might be caused by changes in DNA topology when the global methylation pattern is modified. Thus, a global effect in the transcriptomes might be caused by both, direct and indirect methylation-dependent regulation.

The highly conserved ^{m5}C-MTase methylating GCGC motifs in *H. pylori* may act as a safeguard to maintain suitable transcript levels of many genes, in order to ensure adequate activity of several biological functions. Indeed, we showed that the inactivation of GCGC methylation caused multiple strain-specific phenotypes, including deleterious phenotypes such as a general loss of natural competence, reduced resistance to copper, growth defects, changes in morphology and reduced adhesion to cells.

6.3. Lessons learned and future directions

In the current PhD thesis, I investigated the molecular mechanisms driving genomic and epigenomic evolution of *H. pylori*. To do so, we compared whole genomes and methylomes from isolates collected from early-stage human infections, and characterized the role of a highly conserved ^{m5}C-MTase in the transcriptome of this gastric pathogen. Despite recent advances made by us and many other scientists, many question are still unresolved.

In manuscript I, we showed that modifications in the methylation patterns were due to changes in the activity of phase-variable MTases. However, we did not investigate the effect that the switch in the activity of the MTases could cause in the transcriptome. As mentioned above, intensive research has shown that methylation plays a key role in gene regulation (Kahramanoglou, Prieto et al. 2012, Gonzalez, Kozdon et al. 2014, Chao, Zhu et al. 2015, Kumar, Karmakar et al. 2018). Thus, it would be interesting to investigate the role of the phase-variable MTases during the early-stage of the infection, to understand how phasevarions could have an impact on adaptation. In Figure 1, a schematic representation of how phase-variable MTases could act in early-stage infection is shown. Host specific selective pressures would select for *H. pylori* isolates with ON or OFF alleles. Changes in the methylation pattern would change the expression of several genes, as we showed in manuscript II. These alterations of gene expression would trigger changes in protein translation and could enhance *H. pylori*‘s adaptation to new stomachs. Our transcriptome analysis showed that the absence of methylation in GCGC motifs located in regulatory regions led to a lower gene expression, whether this is the case for other motifs is not known. However, in the figure, this case is represented.

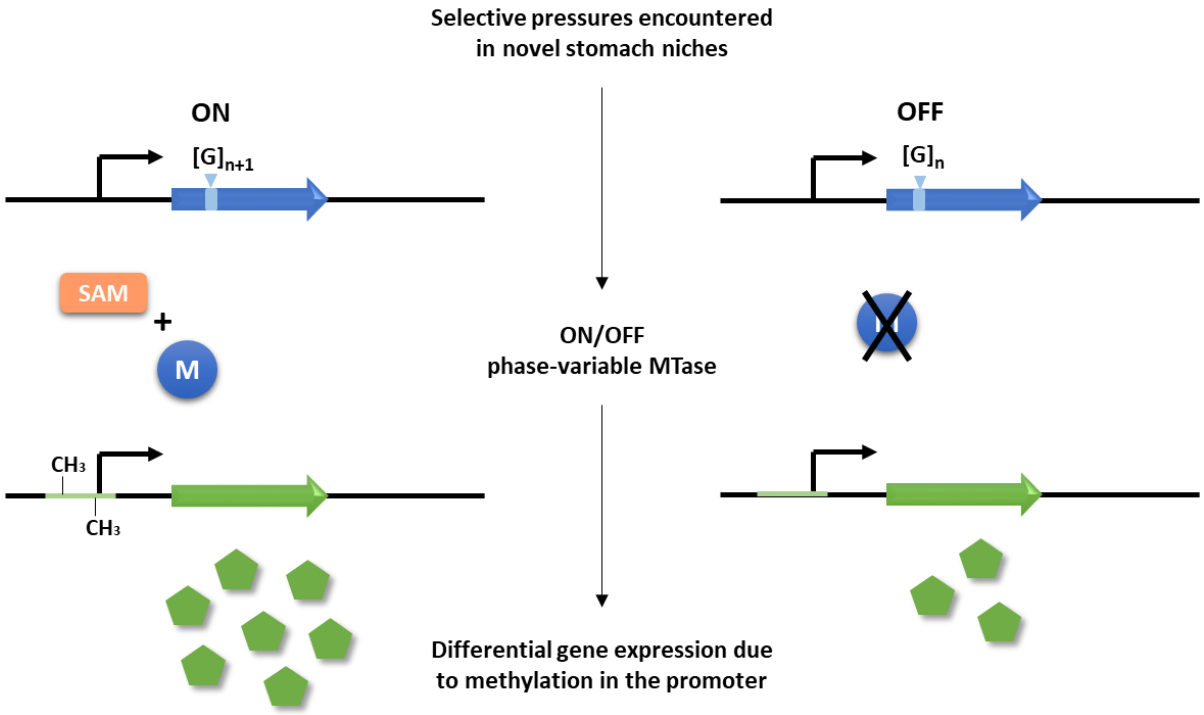


Figure 1. Schematic representation of phase-variable MTases and their influence on gene expression during acute infection. When the MTase gene (dark blue arrow) is in frame (homopolymeric tract is shown in light blue), the MTase enzyme (blue circle) methylates the motif (light green line) overlapping the promoter sequence (black arrow). Methylation of motifs overlapping the promoter would lead to a higher expression of the target gene (green arrow). SAM refers to the donor S-adenyl methionine.

The WHO introduced *H. pylori* in the priority pathogen list for research and development of new antibiotics. The rapid increase in bacterial resistance requires new solutions to eradicate and prevent infections. The genetic variability of *H. pylori* and the high mutation rate we observed during early-stage infection could partially explain why the vaccine candidates tested so far were not successful in preventing the infection. Genomic modifications, such as the ones affecting OMP-related genes, might favour immune evasion. Thus, further research in *H. pylori*'s immune avoidance would be needed in order to develop successful vaccines or alternatives to prevent *H. pylori* infection.

The majority of the SNPs observed during early-stage infection occurred in genes with an OMP-related role, indicating that active modification of the bacterial surface is important for the adaptation to novel niches. However, there were also several other uncharacterized genes carrying non-synonymous SNPs. Some of these genes could be implicated in cell adhesion or related with the bacterial surface. The central roles of some adhesins such as BabA or SabA have been deeply studied (Mahdavi, Sonden et al. 2002, Aspholm-Hurtig, Dailide et al. 2004, Yamaoka 2008, Nell, Kennemann et al. 2014, Hage, Howard et al. 2015) but the main receptors of the majority of putative adhesins remains unidentified. Recently, it was described that one major adhesin, HopQ, binds human CEACAMS and that the

interaction is important for CagA translocation into human epithelial cells (Javaheri, Kruse et al. 2016, Königer, Holsten et al. 2016). Understanding the function of OMP-related genes that are still not characterized would introduce new knowledge into *H. pylori*'s biology and the interaction with the human gastric epithelial cells. Furthermore, some genes with unknown function might have other roles. They could be involved in metabolism, motility or even methylation. Therefore, the great amount of uncharacterized genes reflects that more research is needed to understand how *H. pylori* adapts to new stomachs.

Along the current thesis, the methylomes of bacterial isolates during acute infection have been described. In addition, we discovered that the deletion of a conserved ^{m5}C-MTase methylating GCGC motifs triggered several transcriptome modifications. The results suggested that this enzyme might contribute to maintain adequate transcript levels of many biological functions. Nevertheless, the exact mechanism of how methylation modulates gene expression is not fully understood. Although we linked direct regulation of gene expression to motifs within regulatory regions, we still do not fully comprehend the reason behind it. Besides, global regulation by methylation must also occur due to indirect effects caused by changes in the methylome. It was discovered that methylation modifies the curvature of the DNA (Diekmann 1987, Severin, Zou et al. 2011), and this phenomenon could have an impact on DNA-binding proteins. To my knowledge, the effect of methylation on the topology of the whole genome has not been researched. Protein-DNA interaction assays, such as chromatin immunoprecipitation assays with deep sequencing (ChIP-seq) (Schmidt, Wilson et al. 2009, Myers, Park et al. 2015), could answer the question whether the affinity of already known DNA-binding proteins is modified in the absence of methylation.

Hitherto, we have studied a highly conserved MTase in *H. pylori*. It is well established that R-M systems can act as selfish mobile-genetic elements (Kobayashi 2001), which explains partially why this pathogen harbors such a high number of genes coding for R-M systems. In contrast, the effect on the transcriptome of a newly acquired MTase has not been elucidated. Thus, the *in vitro* introduction of an enzyme into an *H. pylori* strain could give us a first notion of how the acquisition of a novel R-M system modifies the fitness of the bacterium.

7. REFERENCES

Achtman, M., T. Azuma, D. E. Berg, Y. Ito, G. Morelli, Z. J. Pan, S. Suerbaum, S. A. Thompson, A. van der Ende and L. J. van Doorn (1999). "Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions." Mol Microbiol 32(3): 459-470.

Aebischer, T., D. Bumann, H. J. Epple, W. Metzger, T. Schneider, G. Cherepnev, A. K. Walduck, D. Kunkel, V. Moos, C. Loddenkemper, I. Jiadze, M. Panasyuk, M. Stolte, D. Y. Graham, M. Zeitz and T. F. Meyer (2008). "Correlation of T cell response and bacterial clearance in human volunteers challenged with *Helicobacter pylori* revealed by randomised controlled vaccination with Ty21a-based Salmonella vaccines." Gut 57(8): 1065-1072.

Aebischer, T., A. Schmitt, A. K. Walduck and T. F. Meyer (2005). "*Helicobacter pylori* vaccine development: Facing the challenge." Int J Med Microbiol 295(5): 343–353.

Agrawal, A. and T. F. Murphy (2011). "*Haemophilus influenzae* infections in the *H. influenzae* type b conjugate vaccine era." J Clin Microbiol 49(11): 3728-3732.

Ailloud, F., X. Didelot, S. Woltemate, G. Pfaffinger, J. Overmann, R. C. Bader, C. Schulz, P. Malfertheiner and S. Suerbaum (2019). "Within-host evolution of *Helicobacter pylori* shaped by niche-specific adaptation, intragastric migrations and selective sweeps." Nat Commun 10(1): 2273.

Alm, R. A., J. Bina, B. M. Andrews, P. Doig, R. E. Hancock and T. J. Trust (2000). "Comparative genomics of *Helicobacter pylori*: Analysis of the outer membrane protein families." Infect Immun 68(7): 4155-4168.

Alm, R. A., L. S. Ling, D. T. Moir, B. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. deJonge, G. Carmel, P. J. Tummino, A. Caruso, M. Uria-Nickelsen, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis and T. J. Trust (1999). "Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*." Nature 397(6715): 176-180.

Andersen-Nissen, E., K. D. Smith, K. L. Strobe, S. L. R. Barrett, B. T. Cookson, S. M. Logan and A. Aderem (2005). "Evasion of Toll-like receptor 5 by flagellated bacteria." Proc Natl Acad Sci USA 102(26): 9247-9252.

Arber, W. and D. Dussoix (1962). "Host specificity of DNA produced by *Escherichia coli*. I. Host controlled modification of bacteriophage lambda." J Mol Biol 5: 18-36.

Ardui, S., A. Ameur, J. R. Vermeesch and M. S. Hestand (2018). "Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics." Nucleic Acids Res 46(5): 2159–2168.

Aspholm-Hurtig, M., G. Dailide, M. Lahmann, A. Kalia, D. Ilver, N. Roche, S. Vikstrom, R. Sjostrom, S. Linden, A. Backstrom, C. Lundberg, A. Arnqvist, J. Mahdavi, U. J. Nilsson, B. Velapatino, R. H. Gilman, M. Gerhard, T. Alarcon, M. Lopez-Brea, T. Nakazawa, J. G. Fox, P. Correa, M. G. Dominguez-Bello, G. I. Perez-Perez, M. J. Blaser, S. Normark, I. Carlstedt, S. Oscarson, S. Teneberg, D. E. Berg and T. Boren (2004). "Functional adaptation of BabA, the *H. pylori* ABO blood group antigen binding adhesin." Science 305(5683): 519-522.

Atherton, J. C., P. Cao, R. M. Peek, Jr., M. K. Tummuru, M. J. Blaser and T. L. Cover (1995). "Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. Association of specific *vacA* types with cytotoxin production and peptic ulceration." J Biol Chem 270(30): 17771-17777.

Aviles-Jimenez, F., D. P. Letley, G. Gonzalez-Valencia, N. Salama, J. Torres and J. C. Atherton (2004). "Evolution of the *Helicobacter pylori* vacuolating cytotoxin in a human stomach." J Bacteriol 186(15): 5182–5185.

Baldari, C. T., A. Lanzavecchia and J. L. Telford (2005). "Immune subversion by *Helicobacter pylori*." Trends Immunol 26(4): 199-207.

Baltrus, D. A., M. J. Blaser and K. Guillemin (2009). "*Helicobacter pylori* genome plasticity." Genome Dyn 6: 75-90.

Banerjee, S., A. Medina-Fatimi, R. Nichols, D. Tendler, M. Michetti, J. Simon, C. P. Kelly, T. P. Monath and P. Michetti (2002). "Safety and efficacy of low dose *Escherichia coli* enterotoxin adjuvant for urease based oral immunisation against *Helicobacter pylori* in healthy volunteers." Gut 51(5): 634-640.

Bauerfeind, P., R. Garner, B. E. Dunn and H. L. T. Mobley (1997). "Synthesis and activity of *Helicobacter pylori* urease and catalase at low pH." Gut 40(1): 25-30.

Behrens, W., T. Schweinitzer, J. Bal, M. Dorsch, A. Bleich, F. Kops, B. Brenneke, X. Didelot, S. Suerbaum and C. Josenhans (2013). "Role of energy sensor TlpD of *Helicobacter pylori* in gerbil colonization and genome analyses after adaptation in the gerbil." Infect Immun 81(10): 3534-3551.

Berdis, A. J., I. Lee, J. K. Coward, C. Stephens, R. Wright, L. Shapiro and S. J. Benkovic (1998). "A cell cycle-regulated adenine DNA methyltransferase from *Caulobacter crescentus* processively methylates GANTC sites on hemimethylated DNA." Proc Natl Acad Sci USA 95(6): 2874–2879.

Bertani, G. and J. J. Weigle (1953). "Host controlled variation in bacterial viruses." J Bacteriol 65(2): 113-121.

Bickle, T. A. (2004). "Restricting restriction." Mol Microbiol 51(1): 3-5.

Bjorkholm, B., A. Lundin, A. Sillén, K. Guillemin, N. Salama, C. Rubio, J. I. Gordon, P. Falk and L. Engstrand (2001). "Comparison of genetic divergence and fitness between two subclones of *Helicobacter pylori*." Infect Immun 69(12): 7832-7838.

Björkholm, B., M. Sjölund, P. G. Falk, O. G. Berg, L. Engstrand and D. I. Andersson (2001). "Mutation frequency and biological cost of antibiotic resistance in *Helicobacter pylori*." Proc Natl Acad Sci USA 98(25): 14607-14612.

Blakeway, L. V., P. M. Power, F. E.-C. Jen, S. R. Worboys, M. Boitano, T. A. Clark, J. Korlach, L. O. Bakaletz, M. P. Jennings, I. R. Peak and K. L. Seib (2014). "ModM DNA methyltransferase methylome analysis reveals a potential role for *Moraxella catarrhalis* phasevarions in otitis media." FASEB J 28(12): 5197-5207.

Blaser, M. J., G. I. Perez-Perez, H. Kleanthous, T. L. Cover, R. M. Peek, P. H. Chyou, G. N. Stemmermann and A. Nomura (1995). "Infection with *Helicobacter pylori* strains possessing *cagA* is associated with an increased risk of developing adenocarcinoma of the stomach." Cancer Res 55(10): 2111-2115.

Blow, M. J., T. A. Clark, C. G. Daum, A. M. Deutschbauer, A. Fomenkov, R. Fries, J. Froula, D. D. Kang, R. R. Malmstrom, R. D. Morgan, J. Posfai, K. Singh, A. Visel, K. Wetmore, Z. Zhao, E. M. Rubin, J. Korlach, L. A. Pennacchio and R. J. Roberts (2016). "The epigenomic landscape of prokaryotes." PLoS Genet 12(2): e1005854.

Bogan, J. A. and C. E. Helmstetter (1997). "DNA sequestration and transcription in the *oriC* region of *Escherichia coli*." Mol Microbiol 26(5): 889-896.

Borén, T., P. Falk, K. A. Roth, G. Larson and S. Normark (1993). "Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens." Science 262(5141): 1892-1895.

Bray, F., J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre and A. Jemal (2018). "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." CA Cancer J Cli 68(6): 394–424.

Breurec, S., B. Guillard, S. Hem, S. Brisse, F. B. Dieye, M. Huerre, C. Oung, J. Raymond, T. S. Tan, J. M. Thiberge, S. Vong, D. Monchy and B. Linz (2011). "Evolutionary history of *Helicobacter pylori* sequences reflect past human migrations in Southeast Asia." PLoS One 6(7): e22058.

Brockman, K. L., M. T. Branstool, J. M. Atack, F. Robledo-Avila, S. Partida-Sanchez, M. P. Jennings and L. O. Bakaletz (2017). "The ModA2 phasevarion of nontypeable *Haemophilus influenzae* regulates resistance to oxidative stress and killing by human neutrophils." Sci Rep 7(1): 3161.

Bubendorfer, S., J. Krebes, I. Yang, E. Hage, T. F. Schulz, C. Bahlawane, X. Didelot and S. Suerbaum (2016). "Genome-wide analysis of chromosomal import patterns after natural transformation of *Helicobacter pylori*." Nat Commun 7: 11995.

Buenrostro, J. D., B. Wu, H. Y. Chang and W. J. Greenleaf (2015). "ATAC-seq: A method for assaying chromatin accessibility genome-wide." Curr Protoc Mol Biol 109: 21.29.21–21.29.29.

Bugaytsova, J. A., O. Bjornham, Y. A. Chernov, P. Gideonsson, S. Henriksson, M. Mendez, R. Sjostrom, J. Mahdavi, A. Shevtsova, D. Ilver, K. Moonens, M. P. Quintana-Hayashi, R. Moskalenko, C. Aisenbrey, G. Bylund, A. Schmidt, A. Aberg, K. Brannstrom, V. Koniger, S. Vikstrom, L. Rakhimova, A. Hofer, J. Ogren, H. Liu, M. D. Goldman, J. M. Whitmire, J. Aden, J. Younson, C. G. Kelly, R. H. Gilman, A. Chowdhury, A. K. Mukhopadhyay, G. B. Nair, K. S. Papadakos, B. Martinez-Gonzalez, D. N. Sgouras, L. Engstrand, M. Unemo, D. Danielsson, S. Suerbaum, S. Oscarson, L. A. Morozova-Roche, A. Olofsson, G. Grobner, J. Holgersson, A. Esberg, N. Stromberg, M. Landstrom, A. M. Eldridge, B. A. Chromy, L. M. Hansen, J. V. Solnick, S. K. Linden, R. Haas, A. Dubois, D. S. Merrell, S. Schedin, H. Remaut, A. Arnqvist, D. E. Berg and T. Boren (2017). "*Helicobacter pylori* adapts to chronic infection and gastric disease via pH-responsive BabA-mediated adherence." Cell Host Microbe 21(3): 376-389.

Bury-Moné, S., N. O. Kaakoush, C. Asencio, F. Mégraud, M. Thibonnier, H. de Reuse and G. L. Mendz (2006). "Is *Helicobacter pylori* a true microaerophile?" Helicobacter 11(4): 296–303.

Censini, S., C. Lange, Z. Xiang, J. E. Crabtree, P. Ghiara, M. Borodovsky, R. Rappuoli and A. Covacci (1996). "*cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors." Proc Natl Acad Sci USA 93(25): 14648-14653.

Chao, M. C., S. Zhu, S. Kimura, B. M. Davis, E. E. Schadt, G. Fang and M. K. Waldor (2015). "A cytosine methyltransferase modulates the cell envelope stress response in the cholera pathogen." PLoS Genet 11(12): e1005739.

Colbeck, J. C., L. M. Hansen, J. M. Fong and J. V. Solnick (2006). "Genotypic profile of the outer membrane proteins BabA and BabB in clinical isolates of *Helicobacter pylori*." Infect Immun 74(7): 4375-4378.

Cover, T. L. (2016). "*Helicobacter pylori* diversity and gastric cancer risk." MBio 7(1): e01869-01815.

Cover, T. L. and S. R. Blanke (2005). "*Helicobacter pylori* VacA, a paradigm for toxin multifunctionality." Nat Rev Microbiol 3(4): 320-332.

Croxen, M. A., G. Sisson, R. Melano and P. S. Hoffman (2006). "The *Helicobacter pylori* chemotaxis receptor TlpB (HP0103) is required for pH taxis and for colonization of the gastric mucosa." J Bacteriol 188(7): 2656-2665

Czinn, S. J. and T. Blanchard (2011). "Vaccinating against *Helicobacter pylori* infection." Nat Rev Gastroenterol Hepatol 8(3): 133-140.

Daniels, C. C., P. D. Rogers and C. M. Shelton (2016). "A review of pneumococcal vaccines: current polysaccharide vaccine recommendations and future protein antigens." J Pediatr Pharmacol Ther 21(1): 27-35.

Danna, K. and D. Nathans (1971). "Specific cleavage of simian virus 40 DNA by restriction endonuclease of *Hemophilus influenzae*." Proc Natl Acad Sci USA 68(12): 2913-2917.

de Bernard, M., B. Arico, E. Papini, R. Rizzuto, G. Grandi, R. Rappuoli and C. Montecucco (1997). "*Helicobacter pylori* toxin VacA induces vacuole formation by acting in the cell cytosol." Mol Microbiol 26(4): 665-674.

de Reuse, H., D. Vinella and C. Cavazza (2013). "Common themes and unique proteins for the uptake and trafficking of nickel, a metal essential for the virulence of *Helicobacter pylori*." Front Cell Infect Microbiol 3:94.

Didelot, X., S. Nell, I. Yang, S. Woltemate, S. van der Merwe and S. Suerbaum (2013). "Genomic evolution and transmission of *Helicobacter pylori* in two South African families." Proc Natl Acad Sci USA 110(34): 13880-13885.

Diekmann, S. (1987). "DNA methylation can enhance or induce DNA curvature." EMBO J 6(13): 4213-4217.

Djekic, A. and A. Müller (2016). "The immunomodulator VacA promotes immune tolerance and persistent *Helicobacter pylori* infection through its activities on T-cells and antigen-presenting cells." Toxins (Basel) 8(6): 187.

Donahue, J. P., D. A. Israel, V. J. Torres, A. S. Necheva and G. G. Miller (2002). "Inactivation of a *Helicobacter pylori* DNA methyltransferase alters *dnaK* operon expression following host-cell adherence." FEMS Microbiol Lett 208(2): 295-301.

Dorer, M. S., T. H. Sessler and N. R. Salama (2011). "Recombination and DNA repair in *Helicobacter pylori*." Annu Rev Microbiol 65: 329-348.

Dossumbekova, A., C. Prinz, M. Gerhard, L. Brenner, S. Backert, J. G. Kusters, R. M. Schmid and R. Rad (2006). "*Helicobacter pylori* outer membrane proteins and gastric inflammation." Gut 55(9): 1360–1361.

Dryden, D. T. F., N. E. Murray and D. N. Rao (2001). "Nucleoside triphosphate-dependent restriction enzymes." Nucleic Acids Res 29(18): 3728-3741.

Dunne, C., B. Dolan and M. Clyne (2014). "Factors that mediate colonization of the human stomach by *Helicobacter pylori*." World J Gastroenterol 20(19): 5610–5624.

Dussoix, D. and W. Arber (1962). "Host specificity of DNA produced by *Escherichia coli* II. Control over acceptance of DNA from infecting phage lambda." J Mol Biol 5: 37-49.

Eaton, K. A., C. L. Brooks, D. R. Morgan and S. Krakowka (1991). "Essential role of urease in pathogenesis of gastritis induced by *Helicobacter pylori* in gnotobiotic piglets." Infect Immun 59(7): 2470-2475

Eusebi, L. H., R. M. Zagari and F. Bazzoli (2014). "Epidemiology of *Helicobacter pylori* infection." Helicobacter 19(Suppl. 1): 1-5.

Falush, D., C. Kraft, N. S. Taylor, P. Correa, J. G. Fox, M. Achtman and S. Suerbaum (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(26): 15056-15061.

Falush, D., T. Wirth, B. Linz, J. K. Pritchard, M. Stephens, M. Kidd, M. J. Blaser, D. Y. Graham, S. Vacher, G. I. Perez-Perez, Y. Yamaoka, F. Megraud, K. Otto, U. Reichard, E. Katzowitsch, X. Wang, M. Achtman and S. Suerbaum (2003). "Traces of human migrations in *Helicobacter pylori* populations." Science 299(5612): 1582-1585.

Figueiredo, C., J. C. Machado, P. Pharoah, R. Seruca, S. Sousa, R. Carvalho, A. F. Capelinha, W. Quint, C. Caldas, L. J. van Doorn, F. Carneiro and M. Sobrinho-Simoes (2002). "*Helicobacter pylori* and interleukin 1 genotyping: an opportunity to identify high-risk individuals for gastric carcinoma." J Natl Cancer Inst 94(22): 1680-1687.

Fischer, S., U. Römling and B. Tümmler (2019). "A unique methylation pattern by a type I HsdM methyltransferase prepares for DpnI rare cutting sites in the *Pseudomonas aeruginosa* PAO1 genome." FEMS Microbiol lett 366(5): fnz053.

Fischer, W. and R. Haas (2004). "The RecA protein of *Helicobacter pylori* requires a posttranslational modification for full activity." J Bacteriol 186(3): 777-784.

Fischer, W., J. Puls, R. Buhrdorf, B. Gebert, S. Odenbreit and R. Haas (2001). "Systematic mutagenesis of the *Helicobacter pylori* cag pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8." Mol Microbiol 42(5): 1337-1348.

Foegeding, N. J., R. R. Caston, M. S. McClain, M. D. Ohi and T. L. Cover (2016). "An overview of *Helicobacter pylori* VacA toxin biology." Toxins (Basel) 8(6): 173.

Fox, K. L., S. J. Dowideit, A. L. Erwin, Y. N. Srikhanta, A. L. Smith and M. P. Jennings (2007). "*Haemophilus influenzae* phasevarions have evolved from type III DNA restriction systems into epigenetic regulators of gene expression." Nucleic Acids Res 35(15): 5242-5252.

Garcia-Ortiz, M. V., S. Marsin, M. E. Arana, D. Gasparutto, R. Guerois, T. A. Kunkel and J. P. Radicella (2011). "Unexpected role for *Helicobacter pylori* DNA polymerase I as a source of genetic variability." PLoS Genet 7(6): e1002152.

Gebert, B., W. Fischer, E. Weiss, R. Hoffmann and R. Haas (2003). "*Helicobacter pylori* vacuolating cytotoxin inhibits T lymphocyte activation." Science 301(5636): 1099-1102.

Geis, G., H. Leying, S. Suerbaum, U. Mai and W. Opferkuch (1989). "Ultrastructure and chemical analysis of *Campylobacter pylori* flagella." J Clin Microbiol 27(3): 436-441.

Gewirtz, A. T., Y. Yu, U. S. Krishna, D. A. Israel, S. L. Lyons and R. M. Peek, Jr. (2004). "*Helicobacter pylori* flagellin evades toll-like receptor 5-mediated innate immunity." J Infect Dis 189(10): 1914-1920.

Gogarten, J. P., W. F. Doolittle and J. G. Lawrence (2002). "Prokaryotic evolution in light of gene transfer." Mol Biol Evol 19(12): 2226-2238.

Gonzalez, D., J. B. Kozdon, H. H. McAdams, L. Shapiro and J. Collier (2014). "The functions of DNA methylation by CcrM in *Caulobacter crescentus*: a global approach." Nucleic Acids Res 42(6): 3720-3735.

Graham, D. Y., A. R. Opekun, M. S. Osato, H. M. T. El-Zimaity, C. K. Lee, Y. Yamaoka, W. A. Qureshi, M. Cadoz and T. P. Monath (2004). "Challenge model for *Helicobacter pylori* infection in human volunteers." Gut 53(9): 1235-1243.

Gressmann, H., B. Linz, R. Ghai, K. P. Pleissner, R. Schlapbach, Y. Yamaoka, C. Kraft, S. Suerbaum, T. F. Meyer and M. Achtman (2005). "Gain and loss of multiple genes during the evolution of *Helicobacter pylori*." PLoS Genet 1(4): e43.

Hage, N., T. Howard, C. Phillips, C. Brassington, R. Overman, J. Debreczeni, P. Gellert, S. Stolnik, G. S. Winkler and F. H. Falcone (2015). "Structural basis of Lewis b antigen binding by the *Helicobacter pylori* adhesin BabA." Sci Adv 1(7): e1500315.

Hansen, L. M., P. Gideonsson, D. R. Canfield, T. Borén and J. V. Solnick (2017). "Dynamic expression of the BabA adhesin and its BabB paralog during *Helicobacter pylori* infection in Rhesus macaques." Infect Immun 85(6): e00094-17

Hatakeyama, M. (2004). "Oncogenic mechanisms of the *Helicobacter pylori* CagA protein." Nat Rev Cancer 4(9): 688-694.

Higashi, H., K. Yokoyama, Y. Fujii, S. Ren, H. Yuasa, I. Saadat, N. Murata-Kamiya, T. Azuma and M. Hatakeyama (2005). "EPIYA motif is a membrane-targeting signal of *Helicobacter pylori* virulence factor CagA in mammalian cells." J Biol Chem 280(24): 23130-23137.

Hofreuter, D., S. Odenbreit and R. Haas (2001). "Natural transformation competence in *Helicobacter pylori* is mediated by the basic components of a Type IV Secretion System." Mol Microbiol 41(2): 379-391.

Hooi, J. K. Y., W. Y. Lai, W. K. Ng, M. M. Y. Suen, F. E. Underwood, D. Tanyingoh, P. Malfertheiner, D. Y. Graham, V. W. S. Wong, J. C. Y. Wu, F. K. L. Chan, J. J. Y. Sung, G. G. Kaplan and S. C. Ng (2017). "Global prevalence of *Helicobacter pylori* infection: systematic review and meta-analysis." Gastroenterology 153(2): 420-429.

IARC (1994). "Schistosomes, liver flukes and *Helicobacter pylori*. IARC working group on the evaluation of carcinogenic risks to humans. Lyon, 7-14 June 1994." IARC Monogr Eval Carcinog Risks Hum 61: 1-241.

Jang, H. S., W. J. Shin, J. E. Lee and J. T. Do (2017). "CpG and non-CpG methylation in epigenetic gene regulation and brain function." Genes (Basel) 8(6): 148.

Javaheri, A., T. Kruse, K. Moonens, R. Mejías-Luque, A. Debraekeleer, C. I. Asche, N. Tegtmeyer, B. Kalali, N. C. Bach, S. A. Sieber, D. J. Hill, V. Königer, C. R. Hauck, R. Moskalenko, R. Haas, D. H. Busch, E. Klaile, H. Slevogt, A. Schmidt, S. Backert, H. Remaut, B. B. Singer and M. Gerhard (2016). "*Helicobacter pylori* adhesin HopQ engages in a virulence-enhancing interaction with human CEACAMs." Nat Microbiol 2: 16189.

Jones, K. R., Y. M. Joo, S. Jang, Y. J. Yoo, H. S. Lee, I. S. Chung, C. H. Olsen, J. M. Whitmire, D. S. Merrell and J. H. Cha (2009). "Polymorphism in the CagA EPIYA motif impacts development of gastric cancer." J Clin Microbiol 47(4): 959-968.

Josenhans, C., K. A. Eaton, T. Thevenot and S. Suerbaum (2000). "Switching of flagellar motility in *Helicobacter pylori* by reversible length variation of a short homopolymeric sequence repeat in *fliP*, a gene encoding a basal body protein." Infect Immun 68(8): 4598-4603.

Josenhans, C. and S. Suerbaum (2002). "The role of motility as a virulence factor in bacteria." Int J Med Microbiol 291(8): 605-614.

Kahramanoglou, C., A. I. Prieto, S. Khedkar, B. Haase, A. Gupta, V. Benes, G. M. Fraser, N. M. Luscombe and A. S. Seshasayee (2012). "Genomics of DNA cytosine methylation in *Escherichia coli* reveals its role in stationary phase transcription." Nat Commun 3: 886.

Kang, J. and M. J. Blaser (2006). "Bacterial populations as perfect gases: Genomic integrity and diversification tensions in *Helicobacter pylori*." Nat Rev Microbiol 4(11): 826-836.

Kang, S., H. Lee, J. S. Han and D. S. Hwang (1999). "Interaction of SeqA and Dam methylase on the hemimethylated origin of *Escherichia coli* chromosomal DNA replication." J Biol Chem 274(17): 11463-11468.

Kao, C. Y., B. S. Sheu and J. J. Wu (2016). "*Helicobacter pylori* infection: An overview of bacterial virulence factors and pathogenesis." Biomed J 39(1): 14-23.

Kennaway, C. K., A. Obarska-Kosinska, J. H. White, I. Tuszynska, L. P. Cooper, J. M. Bujnicki, J. Trinick and D. T. F. Dryden (2009). "The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein." Nucleic Acids Res 37(3): 762-770.

Kennemann, L., X. Didelot, T. Aebischer, S. Kuhn, B. Drescher, M. Droege, R. Reinhardt, P. Correa, T. F. Meyer, C. Josenhans, D. Falush and S. Suerbaum (2011). "*Helicobacter pylori* genome evolution during human infection." Proc Natl Acad Sci USA 108(12): 5033-5038.

Kobayashi, I. (2001). "Behavior of Restriction-Modification systems as selfish mobile elements and their impact on genome evolution." Nucleic Acids Res 29(18): 3742-3756.

Kong, H., L. F. Lin, N. Porter, S. Stickel, D. Byrd, J. Posfai and R. J. Roberts (2000). "Functional analysis of putative restriction-modification system genes in the *Helicobacter pylori* J99 genome." Nucleic Acids Res 28(17): 3216-3223.

Königer, V., L. Holsten, U. Harrison, B. Busch, E. Loell, Q. Zhao, D. A. Bonsor, A. Roth, A. Kengmo-Tchoupa, S. I. Smith, S. Mueller, E. J. Sundberg, W. Zimmermann, W. Fischer, C. R. Hauck and R. Haas (2016). "*Helicobacter pylori* exploits human CEACAMs via HopQ for adherence and translocation of CagA." Nat Microbiol 2: 16188.

Kozdon, J. B., M. D. Melfi, K. Luong, T. A. Clark, M. Boitano, S. Wang, B. Zhou, D. Gonzalez, J. Collier, S. W. Turner, J. Korlach, L. Shapiro and H. H. McAdams (2013). "Global methylation state at base-pair resolution of the *Caulobacter* genome throughout the cell cycle." Proc Natl Acad Sci USA 110(48): E4658-E4667.

Kraft, C., A. Stack, C. Josenhans, E. Niehus, G. Dietrich, P. Correa, J. G. Fox, D. Falush and S. Suerbaum (2006). "Genomic changes during chronic *Helicobacter pylori* infection." J Bacteriol 188(1): 249-254.

Kraft, C. and S. Suerbaum (2005). "Mutation and recombination in *Helicobacter pylori*: mechanisms and role in generating strain diversity." Int J Med Microbiol 295(5): 299-305.

Krebes, J., X. Didelot, L. Kennemann and S. Suerbaum (2014). "Bidirectional genomic exchange between *Helicobacter pylori* strains from a family in Coventry, United Kingdom." Int J Med Microbiol 304(8): 1135-1146.

Krebes, J., R. D. Morgan, B. Bunk, C. Sproer, K. Luong, R. Parusel, B. P. Anton, C. Konig, C. Josenhans, J. Overmann, R. J. Roberts, J. Korlach and S. Suerbaum (2014). "The complex methylome of the human gastric pathogen *Helicobacter pylori*." Nucleic Acids Res 42(4): 2415-2432.

Krulwich, T. A., G. Sachs and E. Padan (2011). "Molecular aspects of bacterial pH sensing and homeostasis." Nat Rev Microbiol 9(5): 330-343.

Kumar, R., A. K. Mukhopadhyay, P. Ghosh and D. N. Rao (2012). "Comparative transcriptomics of *H. pylori* strains AM5, SS1 and their *hpyAVIBM* deletion mutants: possible roles of cytosine methylation." PLoS One 7(8): e42303.

Kumar, S., B. C. Karmakar, D. Nagarajan, A. K. Mukhopadhyay, R. D. Morgan and D. N. Rao (2018). "N4-cytosine DNA methylation regulates transcription and pathogenesis in *Helicobacter pylori*." Nucleic Acids Res 46(7): 3429-3445.

Kusters, J. G., A. H. van Vliet and E. J. Kuipers (2006). "Pathogenesis of *Helicobacter pylori* infection." Clin Microbiol Rev 19(3): 449-490.

Lamb, A. and L. F. Chen (2013). "Role of the *Helicobacter pylori*-induced inflammatory response in the development of gastric cancer." J Cell Biochem 114(3): 491-497.

Lamichhane, B., Chua, E.G., M. J. Wise, C. Laming, B. J. Marshall and C. Y. Tay (2019). "The complete genome and methylome of *Helicobacter pylori* hpNEAfrica strain HP14039." Gut Pathog 11: 7.

Lee, S. K., A. Stack, E. Katzowitsch, S. I. Aizawa, S. Suerbaum and J. C. (2003). "*Helicobacter pylori* flagellins have very low intrinsic activity to stimulate human gastric epithelial cells via TLR5." Microbes Infect 5(15): 1345-1356.

Lee, W. C., B. P. Anton, S. Wang, P. Baybayan, S. Singh, M. Ashby, E. G. Chua, C. Y. Tay, F. Thirriot, M. F. Loke, K. L. Goh, B. J. Marshall, R. J. Roberts and J. Vadivelu (2015). "The complete methylome of *Helicobacter pylori* UM032." BMC Genomics 16(1): 424.

Lertsethtakarn, P., K. M. Ottemann and D. R. Hendrixson (2011). "Motility and chemotaxis in *Campylobacter* and *Helicobacter*." Annu Rev Microbiol 65: 389-410.

Li, Q., J. Liu, Y. Gong and Y. Yuan (2017). "Association of CagA EPIYA-D or EPIYA-C phosphorylation sites with peptic ulcer and gastric cancer risks." Medicine (Baltimore) 96(17): e6620.

Lieberman, T. D., K. B. Flett, I. Yelin, T. R. Martin, A. J. McAdam, G. P. Priebe and R. Kishony (2014). "Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures." Nat Genet 46(1): 82-87.

Lina, T. T., S. Alzahrani, J. Gonzalez, I. V. Pinchuk, E. J. Beswick and V. E. Reyes (2014). "Immune evasion strategies used by *Helicobacter pylori*." World J Gastroenterol 20(36): 12753–12766.

Linz, B., F. Balloux, Y. Moodley, A. Manica, H. Liu, P. Roumagnac, D. Falush, C. Stamer, F. Prugnolle, S. W. van der Merwe, Y. Yamaoka, D. Y. Graham, E. Perez-Trallero, T. Wadstrom, S. Suerbaum and M. Achtman (2007). "An African origin for the intimate association between humans and *Helicobacter pylori*." Nature 445(7130): 915-918.

Linz, B., H. M. Windsor, J. J. McGraw, L. M. Hansen, J. P. Gajewski, L. P. Tomsho, C. M. Hake, J. V. Solnick, S. C. Schuster and B. J. Marshall (2014). "A mutation burst during the acute phase of *Helicobacter pylori* infection in humans and Rhesus macaques." Nat Commun 5: 4165.

Lluch-Senar, M., K. Luong, V. Lloréns-Rico, J. Delgado, G. Fang, K. Spittle, T. A. Clark, E. Schadt, S. W. Turner, J. Korlach and L. Serrano (2013). "Comprehensive methylome characterization of *Mycoplasma genitalium* and *Mycoplasma pneumoniae* at single-base resolution." PLoS Genet 9(1): e1003191.

Luria, S. E. and M. L. Human (1952). "A nonhereditary, host-induced variation of bacterial viruses." J Bacteriol 64(4): 557-569.

Mahdavi, J., B. Sonden, M. Hurtig, F. O. Olfat, L. Forsberg, N. Roche, J. Angstrom, T. Larsson, S. Teneberg, K. A. Karlsson, S. Altraja, T. Wadstrom, D. Kersulyte, D. E. Berg, A. Dubois, C. Petersson, K. E. Magnusson, T. Norberg, F. Lindh, B. B. Lundskog, A. Arnqvist, L. Hammarstrom and T. Boren (2002). "*Helicobacter pylori* SabA adhesin in persistent infection and chronic inflammation." Science 297(5581): 573-578.

Malfertheiner, P., F. Megraud, C. A. O'Morain, J. P. Gisbert, E. J. Kuipers, A. T. Axon, F. Bazzoli, A. Gasbarrini, J. Atherton, D. Y. Graham, R. Hunt, P. Moayyedi, T. Rokkas, M. Rugge, M. Selgrad, S. Suerbaum, K. Sugano and E. M. El-Omar (2017). "Management of *Helicobacter pylori* infection-the Maastricht V/Florence consensus report." Gut 66(1): 6-30.

Malfertheiner, P., M. Selgrad, T. Wex, B. Romi, E. Borgogni, F. Spensieri, L. Zedda, P. Ruggiero, L. Pancotto, S. Censini, E. Palla, N. Kanesa-Thasan, B. Scharschmidt, R. Rappuoli, D. Y. Graham, F. Schiavetti and G. Del Giudice (2018). "Efficacy, immunogenicity, and safety of a parenteral vaccine against *Helicobacter pylori* in healthy volunteers challenged with a Cag-positive strain: a randomised, placebo-controlled phase 1/2 study." Lancet Gastroenterol Hepatol 3(10): 698-707.

Marshall, B. J., J. A. Armstrong, D. B. McGechie and R. J. Clancy (1985). "Attempt to fulfil Koch's postulates for pyloric *Campylobacter*." Med J Aust 142(8): 436-439.

Marshall, B. J. and J. R. Warren (1984). "Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration." Lancet 1(8390): 1311-1315.

Messer, W., U. Bellekes and H. Lother (1985). "Effect of *dam* methylation on the activity of the *E. coli* replication origin, *oriC*." EMBO J 4(5): 1327-1332.

Michetti, P., C. Kreiss, K. L. Kotloff, N. Porta, J. L. Blanco, D. Bachmann, M. Herranz, P. F. Saldinger, I. Corthésy-Theulaz, G. Losonsky, R. Nichols, J. Simon, M. Stolte, S. Ackerman, T. P. Monath and A. L. Blum (1999). "Oral immunization with urease and *Escherichia coli* heat-labile enterotoxin is safe and immunogenic in *Helicobacter pylori*-infected adults." Gastroenterology 116(4): 804-812.

Mobley, H. L. T. (2001). Urease. *Helicobacter pylori*: physiology and genetics. H. L. T. Mobley, G. L. Mendz and S. L. Hazell. Washington (DC), ASM Press.

Mogensen, T. H. (2009). "Pathogen recognition and inflammatory signaling in innate immune defenses." Clin Microbiol Rev 22(2): 240-273.

Montecucco, C. and M. de Bernard (2003). "Immunosuppressive and proinflammatory activities of the VacA toxin of *Helicobacter pylori*." J Exp Med 198(12): 1767–1771.

Moodley, Y., B. Linz, R. P. Bond, M. Nieuwoudt, H. Soodyall, C. M. Schlebusch, S. Bernhoft, J. Hale, S. Suerbaum, L. Mugisha, S. W. van der Merwe and M. Achtman (2012). "Age of the association between *Helicobacter pylori* and man." PLoS Pathog 8(5): e1002693.

Moodley, Y., B. Linz, Y. Yamaoka, H. M. Windsor, S. Breurec, J. Y. Wu, A. Maady, S. Bernhoft, J. M. Thiberge, S. Phuanukoonnon, G. Jobb, P. Siba, D. Y. Graham, B. J. Marshall and M. Achtman (2009). "The peopling of the Pacific from a bacterial perspective." Science 323(5913): 527-530.

Moore, L. D., T. Le and G. Fan (2013). "DNA methylation and its basic function." Neuropsychopharmacology 38(1): 23-38.

Morelli, G., X. Didelot, B. Kusecek, S. Schwarz, C. Bahlawane, D. Falush, S. Suerbaum and M. Achtman (2010). "Microevolution of *Helicobacter pylori* during prolonged infection of single hosts and within families." PLoS Genet 6(7): e1001036.

Morgan, R. D., Y. A. Luyten, S. A. Johnson, E. M. Clough, T. A. Clark and R. J. Roberts (2016). "Novel ^{m4}C modification in type I restriction-modification systems." Nucleic Acids Res 44(19): 9413–9425.

Moss, S. F., L. Moise, D. S. Lee, W. Kim, S. Zhang, J. Lee, A. B. Rogers, W. Martin and A. S. De Groot (2011). "HelicoVax: Epitope-based therapeutic *Helicobacter pylori* vaccination in a mouse model." Vaccine 29(11): 2085-2091.

Murray, N. E. (2000). "Type I Restriction systems: Sophisticated molecular machines (a legacy of Bertani and Weigle)." Microbiol Mol Biol Rev 64(2): 412-434.

Myers, K. S., D. Park, M., N. A. Beauchene and P. J. Kiley (2015). "Defining bacterial regulons using ChIP-seq." Methods 86: 80-88.

Nell, S., D. Eibach, V. Montano, A. Maady, A. Nkwescheu, J. Siri, W. F. Elamin, D. Falush, B. Linz, M. Achtman, Y. Moodley and S. Suerbaum (2013). "Recent acquisition of *Helicobacter pylori* by Baka pygmies." PLoS Genet 9(9): e1003775.

Nell, S., I. Estibariz, J. Krebes, B. Bunk, D. Y. Graham, J. Overmann, Y. Song, C. Spröer, I. Yang, T. Wex, J. Korlach, P. Malfertheiner and S. Suerbaum (2018). "Genome and methylome variation in *Helicobacter pylori* with a *cag* Pathogenicity Island during early stages of human infection." Gastroenterology 154(3): 612-623.

Nell, S., L. Kennemann, S. Schwarz, C. Josenhans and S. Suerbaum (2014). "Dynamics of Lewis b binding and sequence variation of the *babA* adhesin gene during chronic *Helicobacter pylori* infection in humans." MBio 5(6): e02281-02214.

Nobusato, A., I. Uchiyama and I. Kobayashi (2000). "Diversity of Restriction-Modification gene homologues in *Helicobacter pylori*." Gene 259(1-2): 89-98.

Noto, J. M. and R. M. Peek Jr. (2012). "The *Helicobacter pylori cag* Pathogenicity Island." Methods Mol Biol 921: 41-50.

O'Toole, P. W., M. C. Lane and S. Porwollik (2000). "*Helicobacter pylori* motility." Microbes Infect 2(10): 1207-1214.

Olbermann, P., C. Josenhans, Y. Moodley, M. Uhr, C. Stamer, M. Vauterin, S. Suerbaum, M. Achtman and B. Linz (2010). "A global overview of the genetic and functional diversity in the *Helicobacter pylori* cag pathogenicity island." PLoS Genet 6(8): e1001069.

Parsonnet, J., G. D. Friedman, N. Orentreich and H. Vogelman (1997). "Risk for gastric cancer in people with CagA positive or CagA negative *Helicobacter pylori* infection." Gut 40(3): 297-301.

Peleteiro, B., A. Bastos, A. Ferro and N. Lunet (2014). "Prevalence of *Helicobacter pylori* infection worldwide: a systematic review of studies with national coverage." Dig Dis Sci 59(8): 1698–1709.

Rao, D. N., D. T. Dryden and S. Bheemanaik (2014). "Type III Restriction-Modification enzymes: A historical perspective." Nucleic Acids Res 42(1): 45-55.

Roberts, R. J., M. Belfort, T. Bestor, A. S. Bhagwat, T. A. Bickle, J. Bitinaite, R. M. Blumenthal, S. K. Degtyarev, D. T. Dryden, K. Dybvig, K. Firman, E. S. Gromova, R. I. Gumport, S. E. Halford, S. Hattman, J. Heitman, D. P. Hornby, A. Janulaitis, A. Jeltsch, J. Josephsen, A. Kiss, T. R. Klaenhammer, I. Kobayashi, H. Kong, D. H. Kruger, S. Lacks, M. G. Marinus, M. Miyahara, R. D. Morgan, N. E. Murray, V. Nagaraja, A. Piekarowicz, A. Pingoud, E. Raleigh, D. N. Rao, N. Reich, V. E. Repin, E. U. Selker, P. C. Shaw, D. C. Stein, B. L. Stoddard, W. Szybalski, T. A. Trautner, J. L. Van Etten, J. M. Vitor, G. G. Wilson and S. Y. Xu (2003). "A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes." Nucleic Acids Res 31(7): 1805-1812.

Roberts, R. J., T. Vincze, J. Posfai and D. Macelis (2015). "REBASE--a database for DNA restriction and modification: enzymes, genes and genomes." Nucleic Acids Res 43: D298-D299.

Russell, D. W. and N. D. Zinder (1987). "Hemimethylation prevents DNA replication in *E. coli*." Cell 50(7): 1071-1079.

Salama, N. R., M. L. Hartung and A. Muller (2013). "Life in the human stomach: persistence strategies of the bacterial pathogen *Helicobacter pylori*." Nat Rev Microbiol 11(6): 385-399.

Salama, N. R., G. Otto, L. Tompkins and S. Falkow (2001). "Vacuolating cytotoxin of *Helicobacter pylori* plays a role during colonization in a mouse model of infection." Infect Immun 69(2): 730-736.

Salaun, L., B. Linz, S. Suerbaum and N. J. Saunders (2004). "The diversity within an expanded and redefined repertoire of phase-variable genes in *Helicobacter pylori*." Microbiology 150(4): 817-830.

Schadt, E. E., O. Banerjee, G. Fang, Z. Feng, W. H. Wong, X. Zhang, A. Kislyuk, T. A. Clark, K. Luong, A. Keren-Paz, A. Chess, V. Kumar, A. Chen-Plotkin, N. Sondheimer, J. Korlach and A. Kasarskis (2013). "Modeling kinetic rate variation in third generation DNA sequencing data to detect putative modifications to DNA bases." Genome Res 23(1): 129-141.

Schmidt, D., M. D. Wilson, C. Spyrou, G. D. Brown, J. Hadfield and D. T. Odom (2009). "ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions." Methods 48(3): 240-248.

Schreiber, S., M. Konradt, C. Groll, P. Scheid, G. Hanauer, H. O. Werling, C. Josenhans and S. Suerbaum (2004). "The spatial orientation of *Helicobacter pylori* in the gastric mucus." Proc Natl Acad Sci USA 101(14): 5024-5029.

Scott, D. R., D. Weeks, C. Hong, S. Postius, K. Melchers and G. Sachs (1998). "The role of internal urease in acid resistance of *Helicobacter pylori*." Gastroenterology 114(1): 58-70.

Secka, O., Y. Moodley, M. Antonio, D. E. Berg, M. Tapgun, R. Walton, A. Worwui, V. Thomas, T. Corrah, J. E. Thomas and R. A. Adegbola (2014). "Population genetic analyses of *Helicobacter pylori* isolates from gambian adults and children." PLoS One 9(10): e109466.

Seib, K. L., F. E. Jen, A. L. Scott, A. Tan and M. P. Jennings (2017). "Phase variation of DNA methyltransferases and the regulation of virulence and immune evasion in the pathogenic *Neisseria*." Pathog Dis 75(6): ftx080.

Severin, P. M. D., X. Zou, H. E. Gaub and K. Schulten (2011). "Cytosine methylation alters DNA mechanical properties." Nucleic Acids Res 39(20): 8740–8751.

Shapiro, E. D., A. T. Berg, R. Austrian, D. Schroeder, V. Parcells, A. Margolis, R. K. Adair and J. D. Clemens (1991). "The protective efficacy of polyvalent pneumococcal polysaccharide vaccine." N Engl J Med 325(21): 1453-1460.

Shimizu, T., T. Akamatsu, H. Ota and T. Katsuyama (1996). "Immunohistochemical detection of *Helicobacter pylori* in the surface mucous gel layer and its clinicopathological significance." Helicobacter 1(4): 197-206.

Smith, H. O. and K. W. Wilcox (1970). "A restriction enzyme from *Hemophilus influenzae* I. Purification and general properties." J Mol Biol 51(2): 379-391.

Solnick, J. V., L. M. Hansen, N. R. Salama, J. K. Boonjakuakul and M. Syvanen (2004). "Modification of *Helicobacter pylori* outer membrane protein expression during experimental infection of Rhesus macaques." Proc Natl Acad Sci USA 101(7): 2106-2111.

Spohn, G. S., V. (2001). Motility, chemotaxis, and flagella. *Helicobacter pylori*: Physiology and genetics. H. L. T. Mobley, G. L. Mendz and S. L. Hazell. Washington (DC), ASM Press.

Srikhanta, Y. N., S. J. Dowideit, J. L. Edwards, M. L. Falsetta, H. J. Wu, O. B. Harrison, K. L. Fox, K. L. Seib, T. L. Maguire, A. H. Wang, M. C. Maiden, S. M. Grimmond, M. A. Apicella and M. P. Jennings (2009). "Phasevarions mediate random switching of gene expression in pathogenic *Neisseria*." PLoS Pathog 5(4): e1000400.

Srikhanta, Y. N., K. L. Fox and M. P. Jennings (2010). "The phasevarion: phase variation of Type III DNA methyltransferases controls coordinated switching in multiple genes." Nat Rev Microbiol 8(3): 196-206.

Srikhanta, Y. N., R. J. Gorrell, P. M. Power, K. Tsyganov, M. Boitano, T. A. Clark, J. Korlach, E. L. Hartland, M. P. Jennings and T. Kwok (2017). "Methylomic and phenotypic analysis of the ModH5 phasevarion of *Helicobacter pylori*." Sci Rep 7(1): 16140.

Srikhanta, Y. N., R. J. Gorrell, J. A. Steen, J. A. Gawthorne, T. Kwok, S. M. Grimmond, R. M. Robins-Browne and M. P. Jennings (2011). "Phasevarion mediated epigenetic gene regulation in *Helicobacter pylori*." PLoS One 6(12): e27569.

Srikhanta, Y. N., T. L. Maguire, K. J. Stacey, S. M. Grimmond and M. P. Jennings (2005). "The phasevarion: a genetic system controlling coordinated, random switching of expression of multiple genes." Proc Natl Acad Sci USA 102(15): 5547-5551.

Stingl, K., S. Muller, G. Scheidgen-Kleyboldt, M. Clausen and B. Maier (2010). "Composite system mediates two-step DNA uptake into *Helicobacter pylori*." Proc Natl Acad Sci USA 107(3): 1184-1189.

Styer, C. M., L. M. Hansen, C. L. Cooke, A. M. Gundersen, S. S. Choi, D. E. Berg, M. Benghezal, B. J. Marshall, R. M. Peek, Jr., T. Boren and J. V. Solnick (2010). "Expression of the BabA adhesin during experimental infection with *Helicobacter pylori*." Infect Immun 78(4): 1593-1600.

Suerbaum, S. and C. Josenhans (2007). "*Helicobacter pylori* evolution and phenotypic diversification in a changing host." Nat Rev Microbiol 5(6): 441-452.

Suerbaum, S., J. Maynard Smith, K. Bapumia, G. Morelli, N. H. Smith, E. Kunstmann, I. Dyrek and M. Achtman (1998). "Free recombination within *Helicobacter pylori*." Proc Natl Acad Sci USA 95(21): 12619-12624.

Suerbaum, S. and P. Michetti (2002). "*Helicobacter pylori* infection." N Engl J Med 347(15): 1175-1186.

Sutton, P. and J. M. Boag (2018). "Status of vaccine research and development for *Helicobacter pylori*." Vaccine 18: 30017.

Takeuchi, O., K. Hoshino, T. Kawai, H. Sanjo, H. Takada, T. Ogawa, K. Takeda and S. Akira (1999). "Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components." Immunity 11(4): 443-451.

Tegtmeyer, N., Y. Moodley, Y. Yamaoka, S. R. Pernitzsch, V. Schmidt, F. R. Traverso, T. P. Schmidt, R. Rad, K. Yeoh, H. Bow, J. Torres, M. Gerhard, G. Schneider, S. Wessler and S. Backert (2016). "Characterisation of worldwide *Helicobacter pylori* strains reveals genetic conservation and essentiality of serine protease HtrA." Mol Microbiol 99(5): 925-944.

Testerman, T. L., D. J. McGee and H. L. T. Mobley (2001). Adherence and colonization. *Helicobacter pylori*: Physiology and genetics. H. L. T. Mobley, G. L. Mendz and S. L. Hazell. Washington (DC), ASM Press.

Tomb, J. F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzegerald, N. Lee, M. D. Adams, E. K. Hickey, D. E. Berg, J. D. Gocayne, T. R. Utterback, J. D. Peterson, J. M. Kelley, M. D. Cotton, J. M. Weidman, C. Fujii, C. Bowman, L. Watthey, E. Wallin, W. S. Hayes, M. Borodovsky, P. D. Karp, H. O. Smith, C. M. Fraser and J. C. Venter (1997). "The complete genome sequence of the gastric pathogen *Helicobacter pylori*." Nature 388(6642): 539-547.

Unemo, M., M. Aspholm-Hurtig, D. Ilver, J. Bergström, T. Borén, D. Danielsson and S. Teneberg (2005). "The sialic acid binding SabA adhesin of *Helicobacter pylori* is essential for nonopsonic activation of human neutrophils." J Biol Chem 280(15): 15390-15397.

Vale, F. F., F. Megraud and J. M. Vitor (2009). "Geographic distribution of methyltransferases of *Helicobacter pylori*: evidence of human host population isolation and migration." BMC Microbiol 9: 193.

van Vliet, A. H. M., E. J. Kuipers, B. Waidner, B. J. Davies, N. de Vries, C. W. Penn, C. M. J. E. Vandenbroucke-Grauls, M. Kist, S. Bereswill and J. G. Kusters (2001). "Nickel-Responsive induction of urease expression in *Helicobacter pylori* is mediated at the transcriptional level." Infect Immun 69(8): 4891-4897.

Varga, M. G. and R. M. Peek Jr. (2017). "DNA transfer and toll-like receptor modulation by *Helicobacter pylori*." Curr Top Microbiol Immunol 400: 169-193.

Vasu, K. and V. Nagaraja (2013). "Diverse functions of restriction-modification systems in addition to cellular defense." Microbiol Mol Biol Rev 77(1): 53-72.

Warren, J. R. and B. Marshall (1983). "Unidentified curved bacilli on gastric epithelium in active chronic gastritis." Lancet 321(8336): 1273-1275.

Wen, Y., D. R. Scott, O. Vagin, E. Tokhtaeva, E. A. Marcus and G. Sachs (2018). "Measurement of internal pH in *Helicobacter pylori* by using green fluorescent protein fluorimetry." J Bacteriol 200(14): e00178-00118.

WHO (2017). "Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics."

Willhite, D. C. and S. R. Blanke (2004). "*Helicobacter pylori* vacuolating cytotoxin enters cells, localizes to the mitochondria, and induces mitochondrial membrane permeability changes correlated to toxin channel activity." Cell microbiol 6(2): 143-154.

Winter, J. A., D. P. Letley, K. W. Cook, J. L. Rhead, A. A. M. Zaitoun, R. J. M. Ingram, K. R. Amilon, N. J. Croxall, P. V. Kaye, K. Robinson and J. C. Atherton (2014). "A role for the vacuolating cytotoxin, VacA, in colonization and *Helicobacter pylori*–induced metaplasia in the stomach." J Infect Dis 210(6): 954-963.

Wirth, H. P., M. H. Beins, M. Yang, K. T. Tham and M. J. Blaser (1998). "Experimental infection of Mongolian gerbils with wild-type and mutant *Helicobacter pylori* strains." Infect Immun 66(10): 4856-4866.

Yamaoka, Y. (2008). "Increasing evidence of the role of *Helicobacter pylori* SabA in the pathogenesis of gastroduodenal disease." J Infect Dev Ctries 2(3): 174-181.

Yamaoka, Y., T. Kodama, O. Gutierrez, J. G. Kim, K. Kashima and D. Y. Graham (1999). "Relationship between *Helicobacter pylori iceA*, *cagA*, and *vacA* status and clinical outcome: Studies in four different countries." J Clin Microbiol 37(7): 2274-2279.

Yue, M., X. Han, L. De Masi, C. Zhu, X. Ma, J. Zhang, R. Wu, R. Schmieder, R. S. Kaushik, G. P. Fraser, S. Zhao, P. F. McDermott, F. X. Weill, J. G. Mainil, C. Arze, W. F. Fricke, R. A. Edwards, D. Brisson, N. R. Zhang, S. C. Rankin and D. M. Schifferli (2015). "Allelic variation contributes to bacterial host specificity." Nat Commun 6: 8754.

Zagari, R. M., S. Rabitti, L. H. Eusebi and F. Bazzoli (2017). "Treatment of *Helicobacter pylori* infection: A clinical practice update." Eur J Clin Invest 48(1): cl2857.

Zeng, M., X. H. Mao, J. X. Li, W. D. Tong, B. Wang, Y. J. Zhang, G. Guo, Z. J. Zhao, L. Li, D. L. Wu, D. S. Lu, Z. M. Tan, H. Y. Liang, C. Wu, D. H. Li, P. Luo, H. Zeng, W. J. Zhang, J. Y. Zhang, B. T. Guo, F. C. Zhu and Q. M. Zou (2015). "Efficacy, safety, and immunogenicity of an oral recombinant *Helicobacter pylori* vaccine in children in China: a randomised, double-blind, placebo-controlled, phase 3 trial." Lancet 386(10002): 1457-1464.

8. ACKNOWLEDGMENTS

At first, I would like to thank my supervisor Prof. Dr. Sebastian Suerbaum for introducing me into the interesting field that is *Helicobacter pylori* and the molecular microbiology. I am very grateful for your support, guidance and helpful discussions during all the PhD process. I would also like to express my gratitude for the opportunity to attend several international meetings, where I had the chance of meeting many high-level scientist, being updated in the field and expose my results.

Great thanks to Prof. Dr. Christine Josenhans for accepting to be my co-supervisor. I am grateful for all the constructive discussions, advices and scientific support during the course of this PhD thesis. Next, I would like to thank Prof. Dr. Burkhard Tümmler for all the nice inputs during the co-supervisors meetings.

I would like to thanks the Examination Committee for their time examining my thesis.

I would like to recognize all the effort done by all the co-authors of the manuscripts, whose effort was necessary for the success of the publications.

Many thanks to all (past and actual) members of the two "*Helis"* groups for all the great moments lived during these time. I would like to show my most sincere gratitude to Juliane Krebes for all the time she invested teaching all techniques to me when I started, and for being my mentor in the lab at the beginning of this thesis. Especial thanks go for Friederike Kops and Birgit Brenneke, for being the pillars of the *Hannover*-lab, always helpful and supportive, and for all the funny moments lived in the lab (*i.e. all the singing, especially Disney songs*). Of course, I have to mention the *Munich*-team. Florent and Anny, for all the good moments together (*I cannot even mention all the stupid things we do, there are too many*), you two are always ready to help (*personally and scientifically*). And Gudrun, your amazing assistance and your knowledge keep the lab running (*I don't know how we would have survived without your help*).

To my "*pendejos*", because we built a fantastic family and lived thousands of adventures together that I know they won't stop. Because you're the best and you know it! *Pendejos conquering the world!*

A mi "*Kuadrilla Muniquesa*", por acogerme tan rápido cuando llegué a esta ciudad. Por todas las aventuras, noches de fiesta, excursiones al monte, momentos de turisteo, barbacoas y picinics, etc… En especial a mis "*vascas y cia"*, Lau y Carol, por las risas, el apoyo constante y todos esos momentos de cotilleos varios.

Nire Gasteizko kuadrillarentzat! A pesar de la distancia, la *Kuadri* siempre esta ahí. Mila esker por estos 25 años de amistad (*y alguno menos para las incorporaciones tardías, pero que valen igual que 25!*) y por todos los que nos quedan! GORA LA KUADRILLA! Eta Jul, zure laguntzagatik eta etengabeko euskarriagatik.

A Pedro, que aunque has llegado en la etapa final ha sido la más estresante; y has sabido entenderme, escucharme y hacerme reir.

Y por supuesto, el gracias más especial es para la familia (Tato, Maite, Ekaitxo, Izei, Abel, Rubén, Yone, Mapi, Leyre, Tía Pili, Abuela…). Pero sobre todo para Aita y Ama, porque siempre me habéis apoyado y me habéis hecho sentir segura en todo momento.

ESKERRIK ASKO!

GRACIAS!

THANKS!

DANKE!