Aus dem

Lehrstuhl für Medizinische Mikrobiologie und Krankenhaushygiene, Max-von-Pettenkofer-Institut

Institut der Ludwig-Maximilians-Universität München

Functional and computational characterization of methyltransferases of Helicobacter pylori

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

vorgelegt von
Wilhelm Dominik Gottschall

aus
München / Deutschland

Jahr
2026

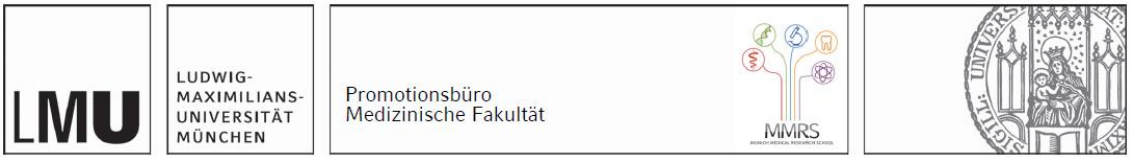
Mit Genehmigung der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

Erstes Gutachten:	Prof. Dr. Sebastian Suerbaum
Zweites Gutachten:	Priv. Doz. Dr. Christian Schulz
Drittes Gutachten:	Priv. Doz. Dr. Maximilian Münchhoff
Viertes Gutachten:	Prof. Dr. Gunnar Schotta

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 02.03.2026

Affidavit



Affidavit

Gottschall, Wilhelm

Surname, first name

Pettenkoferstr. 9A

Street

80336 Munich, Germany

Zip code, town, country

I hereby declare, that the submitted thesis entitled:

Functional and computational characterization of methyltransferases of *Helicobacter pylori*

……

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the dissertation presented here has not been submitted in the same or similar form to any other institution for the purpose of obtaining an academic degree.

Munich, 03.03.26

place, date

Wilhelm Gottschall

Signature doctoral candidate

Confirmation of congruency

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Promotionsbüro
Medizinische Fakultät

MMRS

Confirmation of congruency between printed and electronic version of the doctoral thesis

Gottschall, Wilhelm
Surname, first name

Pettenkoferstr. 9A
Street

80336 Munich, Germany
Zip code, town, country

I hereby declare, that the submitted thesis entitled:

Functional and computational characterization of methyltransferases of *Helicobacter pylori*

……

is congruent with the printed version both in content and format.

Munich, 03.03.26
place, date

Wilhelm Gottschall
Signature doctoral candidate

Table of content

List of abbreviations

^{m6}A: N6-methyladenine

^{m4}C: N4-methylcytosine

^{m5}C: 5-methylcytosine

CAT: chloramphenicol acetyl transferase

CDS: coding sequences

CH_3: methyl group

DEGs: differentially expressed genes

dsDNA: double-stranded DNA

FC: fold change

FCS: fetal calf serum

gDNA: genomic DNA

Km: kanamycin

KO: knockout

MTase: methyltransferase

MuGENT: multiplex genome editing

NGS: next generation sequencing

qPCR: quantitative PCR

PM: point mutant

REase: restriction endonuclease

RM system: restriction modification system

TSS: transcriptional start site

List of publications

Ailloud F, Gottschall W, Suerbaum S. Methylome evolution suggests lineage-dependent selection in the gastric pathogen *Helicobacter pylori*. Commun Biol. 2023 Aug 12;6(1):839. doi: 10.1038/s42003-023-05218-x. PMID: 37573385; PMCID: PMC10423294.

Gottschall W, Ailloud F, Josenhans C, Suerbaum S. The *Helicobacter pylori* orphan ATTAAT-specific methyltransferase M.Hpy99XIX plays a central role in the coordinated regulation of genes involved in iron metabolism. mBio 2025 (in press, accepted 08.05.2025)

1. Chapter 1 Introductory summary

1.1 Summary

Helicobacter pylori is unique among bacteria because it is known to have one of the largest numbers of restriction-modification (RM) systems in its genome. To better understand how RM systems contribute to *H. pylori*'s evolution, we analyzed 541 *H. pylori* genomes representing the seven major phylogeographic populations of this pathogen, and demonstrated that type II RM systems are the most conserved in this species. We also showed that type II sequence motifs face different levels of natural selection. Positive as well as negative correlations could be observed between the presence of certain type II methyltransferases (MTases) and their cognate target motifs undergoing methylation. These results suggest there to be an evolutionary relationship between RM systems and their corresponding target sequences. Additionally, we investigated the evolution of the ACGT motif recognized by the M.Hpy99XI MTase. Our analysis showed that this motif is the result of expansions and contractions due to mutations and deamination events driven by environmental conditions specific to certain locations.

MTases have been implicated in influencing the transcriptome through gene regulation. To further characterize the role of MTases in transcription, we decided to study the strictly orphan (i.e., always lacking a matching endonuclease) MTase M.Hpy99XIX since it cannot fulfill the classical function in phage defense. The removal of the M.Hpy99XIX MTase in *H. pylori*, resulting in a loss of ATTAAT methylation, has been shown to have a significant influence on gene expression and phenotype. Mutation of a single ATTAAT motif in the promoter region of the iron-acquisition gene *frpB1* resulted in a variety of effects on phenotype and the expression of some genes related to iron metabolism. Our research suggests that M.Hpy99XIX regulates gene expression via a two-step model. Firstly, we demonstrate that genes are directly regulated by the methylation of ATTAAT motifs located in the promoter region. Secondly, metabolic pathways are affected by the indirect, downstream consequences of deregulating genes that are directly controlled by ATTAAT methylation. The ferric uptake regulator (Fur) has been reported to regulate iron homeostasis, and we propose that M.Hpy99XIX is also involved in this regulation.

1.2 My contribution to the publications

1.2.1 Contribution to paper I

My first contribution to the first paper (co-author) was to build a database comprising 541 *H. pylori* genomes from the seven major phylogeographic populations. Samples were either obtained from a public database or sequenced in-house. I performed quality control of the genomes and assigned them to the different phylogeographic populations. Using the REBASE database, I then contributed to the assembly of a collection of 96 *H. pylori* RM systems, including their described type and target motif. Subsequently, I annotated the genome collection using Prokka and analyzed target motif frequencies using the Biostrings Bioconductor R package. Furthermore, I investigated gain and loss mechanisms of RM systems by searching for direct flanking repeats and determined their sequences. I also contributed to the analysis of the ACGT motif pattern by generating a smaller model database of 40 genomes, which were randomly selected and balanced across four representative phylogeographic populations (hpAfrica1, hpAfrica2, hpAsia2, and hpEastAsia). Throughout the process of obtaining the results for this paper, I contributed to their interpretation. I also helped revise the manuscript by proofreading and spell-checking it.

1.2.2 Contribution to paper II

For the second paper (first author), I contributed to the early stages of the study design by developing a script that can extract promoter regions of any *H. pylori* genome using reference data (Sharma et al., 2010). This script was used to perform an analysis of target motif frequency in *H. pylori*'s promoters, which showed that ATTAAT was the most enriched motif in these regions. Moreover, I analyzed the positional conservation of this motif relative to transcription start sites and compared it to known elements such as the Pribnow box. Further analysis of the database built for the first manuscript also uncovered that the M.Hpy99XIX MTase, which recognizes the ATTAAT motif, was the only orphan MTase among the known RM systems in *H. pylori*. I validated this initial result using the newly published HpGP genome database (Thorell et al., 2023), and characterized the presence and absence of the M.Hpy99XIX MTase, as well as its mechanisms of inactivation. To functionally characterize the role of this MTase, I designed several gene expression experiments aimed at understanding the overall effect of the MTase on the transcriptome as well as the effect of specific ATTAAT motifs on individual genes. Specifically, I performed RNA-seq analysis of a knockout mutant and a complementation mutant of the M.Hpy99XIX MTase in *H. pylori* strain J99 (the sequencing was done with a contract service provider). I generated point mutants of ATTAAT motifs in multiple genes and

analyzed them with qPCR. I validated several key results using qPCR in a second strain, *H. pylori* N6. To connect these results to phenotypic effects, I built a collection of published RNA-seq dataset associations with known transcriptional regulators in *H. pylori* and compared it to the results obtained for M.Hpy99XIX. To demonstrate these phenotypic effects, I developed growth assays to test the effect of different environmental conditions on the growth of *H. pylori*. All throughout this project, I was involved in designing the experiments and the interpretation of the results. I was involved in writing the original draft of the manuscript and subsequent revisions.

1.3 Introduction to *Helicobacter pylori*

1.3.1 The gastric pathogen *H. pylori*

H. pylori is a spiral-shaped, Gram-negative bacterium that colonizes the human stomach from an early age (Suerbaum & Michetti, 2002). If left untreated, the infection can persist for life, potentially leading to conditions such as peptic ulcers, gastritis, gastric cancer, and gastric mucosa-associated lymphoid-tissue (MALT) lymphoma (Suerbaum & Michetti, 2002). The bacterium produces several virulence factors, including CagA, VacA, and BabA, which contribute to inflammation, tissue damage, and carcinogenesis (Faass et al., 2023; Malfertheiner et al., 2023; Yamaoka, 2010). It is estimated that approximately half of the world's population is currently infected with *H. pylori*; however, the infection rate has been decreasing in recent decades (Malfertheiner et al., 2023). The current treatment regimen for eradicating *H. pylori* involves a strong acid suppressant combined with multiple antibiotics (Malfertheiner et al., 2023). However, *H. pylori*'s increasing resistance to antibiotics is a growing concern (Malfertheiner et al., 2023).

1.3.2 Genomic diversity in *H. pylori*

Mutation and recombination are the primary drivers of genomic diversity in *H. pylori* (Didelot et al., 2013; Falush et al., 2001). Suerbaum et al. (1998) showed that recombination between different *H. pylori* strains occurs frequently, with imports averaging a size of 417 base pairs (bp) (Falush et al., 2001), and more often than in other bacteria. In addition to the interstrain recombination, intrastrain diversification through mutation is thought to help *H. pylori* adapt to the human stomach (Bjorkholm et al., 2001; Suerbaum & Josenhans, 2007). The core genome of *H. pylori* is relatively small, comprising many accessory genes such as virulence factors and RM systems (Ailloud et al., 2021;

Gressmann et al., 2005). These are suggested to contribute an additional level of transcriptomic variation, enhancing the ability to adapt (Ailloud et al., 2021).

Genomic diversity is used to study the evolution of *H. pylori*. Genomic data from 370 *H. pylori* isolates enabled the classification into seven distinct populations (Falush et al., 2003). These populations correspond to human migrations and are thought to have originated in Africa (Linz et al., 2007; Suerbaum & Josenhans, 2007).

1.4 Introduction to Restriction-Modification systems

1.4.1 Types of RM systems

RM systems are composed of different enzyme subunits (Roberts et al., 2003). These subunits include an endonuclease, which cleaves DNA at a specific position; an MTase, which adds a methyl group to a specific residue at a specific position in the DNA; and can include a specificity subunit, which recognizes a motif in the DNA (Roberts et al., 2003). There are three different types of RM systems, categorized by the organization of the individual enzymes (Roberts et al., 2003). Type I RM systems consist of multisubunit complexes containing two MTase subunits, two endonuclease subunits, and one specificity subunit (Murray, 2000; Roberts et al., 2003). Type II RM systems are the most common, consisting of individual MTase and endonuclease enzymes that recognize an identical target motif (Pingoud et al., 2005; Roberts et al., 2003). Type III RM systems consist of two MTase and two endonuclease subunits, with the specificity being integrated into the MTase subunit (Dryden et al., 2001; Roberts et al., 2003). Type IV RM systems are a specific type of endonucleases that recognize and cleave modified residues (Roberts et al., 2003; Stewart et al., 2000). An extensive overview of the various RM systems is available on the REBASE database (Roberts et al., 2023).

1.4.2 Types of DNA methylation in prokaryotes

DNA MTases add methyl groups to specific sequences of DNA. Currently, only cytosine and adenine are known to be methylated. Cytosine can be methylated at the fourth or fifth position, whereas adenine is methylated in the sixth position (Sanchez-Romero & Casadesus, 2020). DNA methylation can be studied using various methods, including bisulfite sequencing, restriction enzyme digestion followed by next-generation sequencing, single-molecule, real-time (SMRT) sequencing (Flusberg et al., 2010), and Oxford Nanopore sequencing (Beaulaurier et al., 2019; Patel et al., 2024).

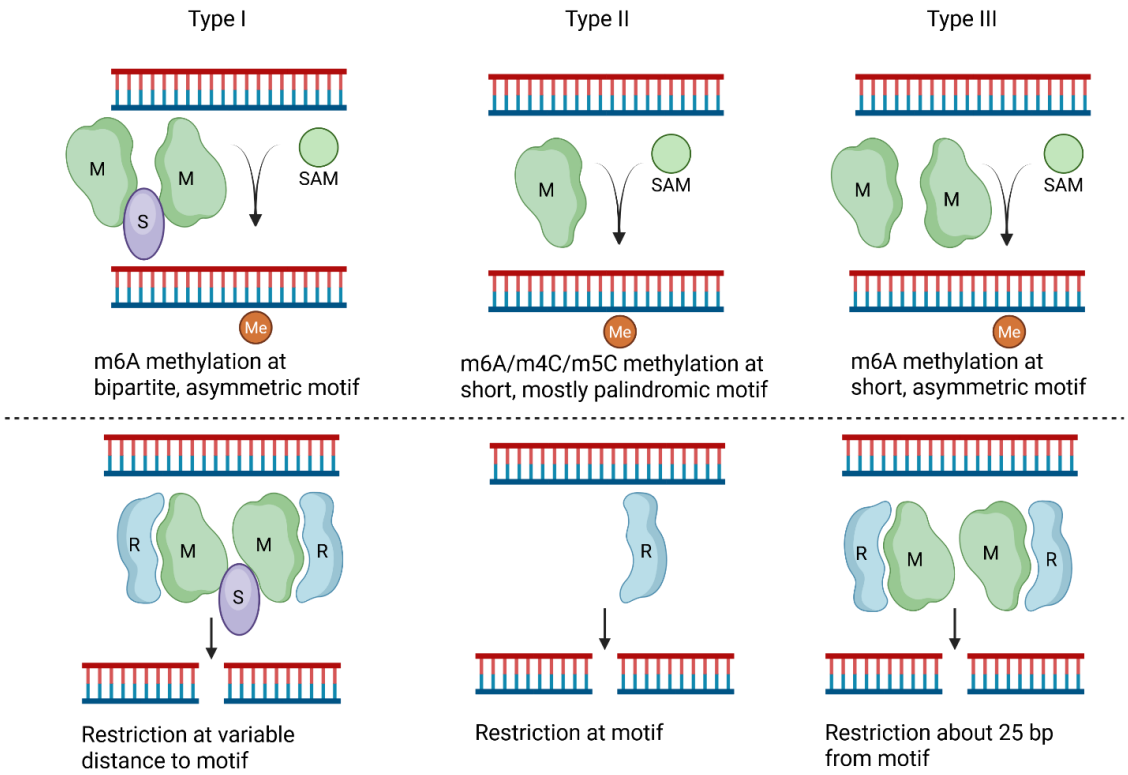


Figure 1: Classification of RM systems according to their function in *H. pylori*. M: MTase; R: endonuclease; S: specificity subunit; SAM: S-adenosyl-methionine. Figure created with BioRender.com.

1.5 DNA and RNA methylation in bacterial phenotype

1.5.1 DNA adenine methylation, bacterial replication, and phase variation

DNA methylation has been shown to influence various processes, including gene regulation, virulence, and adaptation (Heithoff et al., 1999). The MTase that has been studied most extensively is Dam, which recognizes the GATC target motif (Boye & Lobner-Olesen, 1990). It has been shown to control the initiation of DNA replication and regulate DnaA levels in *Escherichia coli* (Boye & Lobner-Olesen, 1990). Another example of how methylation regulates gene expression can be seen in the control of pyelonephritis-associated pili phase variation in *E. coli* (Braaten et al., 1994). Two sites that are methylated by Dam were found to exhibit differential methylation in the regulatory region of the *pap* operon, enabling the bacteria to transition between ON and OFF states of *pap* expression (Braaten et al., 1994). Furthermore, Heithoff et al. (1999) described the essential role of Dam methylation in the virulence of *S.* Typhimurium. Deletion of the *dam* gene resulted in an avirulent mutant that was effective as a vaccine in the murine model (Heithoff et al., 1999).

1.5.2 Additional DNA methylation systems and their role in pathogenesis

In addition to adenine methylation, which was mentioned previously, Kumar et al. (2018) describe how N4-cytosine methylation regulates gene expression and pathogenesis in *H. pylori*. Methylation affects adherence to host AGS cells, as well as inducing inflammation and apoptosis in AGS cells (Kumar et al., 2018). Another type of cytosine methylation, involving an *H. pylori* m5C MTase, was shown to significantly influence transcription, as well as affecting adherence to gastric epithelial cells, competence for DNA uptake, morphology, and susceptibility to copper (Estibariz et al., 2019). This effect on transcription was traced back to the methylation of a motif in the promoter region (Estibariz et al., 2019). Yano et al. (2020) demonstrated that different *H. pylori* MTases have functions in motility, oxidative stress tolerance, and DNA damage repair, and that they interact with each other. MTases can alter the specificity of their target motif, thereby adding plasticity to this regulatory system (Krebes et al., 2014; Yano et al., 2020).

MTases have also been shown to play a role in pathogenesis in other bacteria. For instance, a conserved MTase in *C. difficile* has been shown to regulate sporulation and pathogenesis (Oliveira et al., 2020). In *P. aeruginosa*, an MTase has been shown to regulate nitric oxide homeostasis and virulence (Han et al., 2022). In *A. baumannii*, an MTase has been shown to play a role in the formation of persister cells (Kim et al., 2022).

1.5.3 Description of the phasevarion and influence on phenotype

The activity of MTases can change for a variety of reasons. One such mechanism is phase variation, whereby MTase activity can switch between an “ON” and “OFF” state (Krebes et al., 2014; Srikhanta et al., 2005). Such switching can result in phenotypic changes in terms of virulence, immune evasion, and resistance to oxidative stress (Srikhanta et al., 2011). Fox et al. (2007) demonstrated that these systems evolved from type III RM systems to become epigenetic regulators of transcription.

To date, 17 different phase-variable type III RM system *modH* alleles have been characterized in *H. pylori*; these differ only in their DNA recognition domain (Srikhanta et al., 2011). The most common of these is ModH5, which has been shown to regulate genes involved in virulence, such as those encoding motility and outer membrane proteins (Srikhanta et al., 2017). ModH5 has also been shown to regulate the *flaA* promoter in a methylation-dependent manner (Srikhanta et al., 2017).

1.5.4 Effect of RNA methylation on the bacterial phenotype

Methylation has been observed in most types of RNA, and its influence on the bacterial phenotypes is described below for a selection of these. Methylation of tRNA results in the formation of 1-methylguanosine at position 37 and has been shown to prevent frameshifting during translation (Bjork et al., 1989). Methylation of rRNA occurs close to the exit tunnel of the ribosome and moderately affects cell fitness (Sergiev et al., 2008). Modification of mRNA regulates mRNA translation and alters the genetic code (Hoernes et al., 2016). No phenotypes have yet been associated with RNA methylation in *H. pylori*, although the existence of RNA MTases can be inferred from homology.

1.6 Role of iron in regulation of iron provisioning and host interaction in *H. pylori*

1.6.1 Importance of iron for *H. pylori* infection and pathogenesis

Iron is essential for *H. pylori* and is required for its growth and pathogenicity, yet its availability is limited within the gastric niche that the bacterium colonizes. Consequently, *H. pylori* has evolved a mechanism to take up iron from humans (Husson et al., 1993). This acquisition may be linked to stomach colonization and the virulence of the infection (Husson et al., 1993). Competition for iron within the gastric niche not only determines *H. pylori*'s survival but also enhances its pathogenicity (Noto et al., 2013).

1.6.2 Iron acquisition and regulation in *H. pylori*

As there is limited free iron within the gastric niche, *H. pylori* acquires iron from the human host in the form of human lactoferrin (Husson et al., 1993). This study showed that *H. pylori* does not grow in the presence of human transferrin, bovine lactoferrin, or hen ovotransferrin, suggesting a high level of specificity in this mechanism (Husson et al., 1993). Once inside the cell, iron is stored in ferritin (Pfr), which plays a key role in iron metabolism and gastric colonization (Waidner et al., 2002). Deletion of *pfr* resulted in mutants being unable to colonize Mongolian gerbils (Waidner et al., 2002). Pfr stores iron, enabling *H. pylori* to grow under iron-limited conditions and protecting it from acid-amplified iron toxicity (Waidner et al., 2002).

The ferric uptake regulator protein (Fur) regulates iron homeostasis in *H. pylori*, and it has been suggested that it has additional functions due to the lack of regulators in *H. pylori* (Ernst et al., 2005). Under iron-replete conditions, Fur represses iron acquisition genes and activates iron storage by Pfr, as well as oxidative stress defense genes (Ernst et al., 2005). Fur specifically binds to promoters, as demonstrated for the *pfr* promoter

(Waidner et al., 2002). Fur regulation is important for colonization, particularly in response to changes in stomach conditions, such as the pH (Gancz et al., 2006). Studies have shown that Fur is necessary for the colonization of Mongolian gerbils and that most of the genes it regulates are involved in the acid stress response (Gancz et al., 2006).
More recent studies on Fur have described the Fur box sequences in promoters and updated and expanded the holo- and apo-Fur conformation regulons (Pich et al., 2012; Vannini et al., 2024). The Fur box core sequence was found to be a 7-1-7 motif with dyad symmetry, and mutations in this sequence were found to result in decreased Fur regulation (Pich et al., 2012). Fur has been shown to bind specifically to promoters containing the Fur box sequence (Pich et al., 2012). Using new methods such as RNA-seq and ChIP-seq, the Fur regulon was updated to expand on previous findings (Vannini et al., 2024). This study reported new coding sequences, non-coding RNAs, toxin-antitoxin systems, and transcripts within open reading frames that are regulated by Fur in response to iron availability (Vannini et al., 2024).

1.6.3 Host interaction and iron-mediated pathogenesis in *H. pylori*

In addition to internal regulation, *H. pylori* also disrupts the host epithelial cells to obtain iron for its metabolic needs (Tan et al., 2011). This iron acquisition is made possible by the virulence factor CagA, which is injected by *H. pylori* to increase transferrin internalization (Tan et al., 2011). Another virulence factor, VacA, has been shown to facilitate *H. pylori* colonization to the apical cell surface (Tan et al., 2011). The resulting iron deficiency in epithelial cells has been found to accelerate the progression of gastric carcinogenesis arising from *H. pylori* infection due to enhanced virulence (Noto et al., 2013). This increase in virulence was described as resulting from the limitation of iron (Noto et al., 2013).
Iron metabolism has also been identified as being interconnected with metal homeostasis. The SlyD metallochaperone, which was initially described for its role in regulating nickel transport, manages metal cofactors and influences iron-sulfur cluster biosynthesis and core metabolic pathways, suggesting a broad role in metal homeostasis (Denic et al., 2023). The interconnection of iron and nickel uptake may impact *H. pylori*'s ability to adapt to gastric conditions, given that SlyD fulfills multiple functions (Denic et al., 2021). SlyD combines a peptidyl-prolyl isomerase, a chaperone, and metal-binding functions and has been shown to be essential for the colonization of mice (Denic et al., 2021).

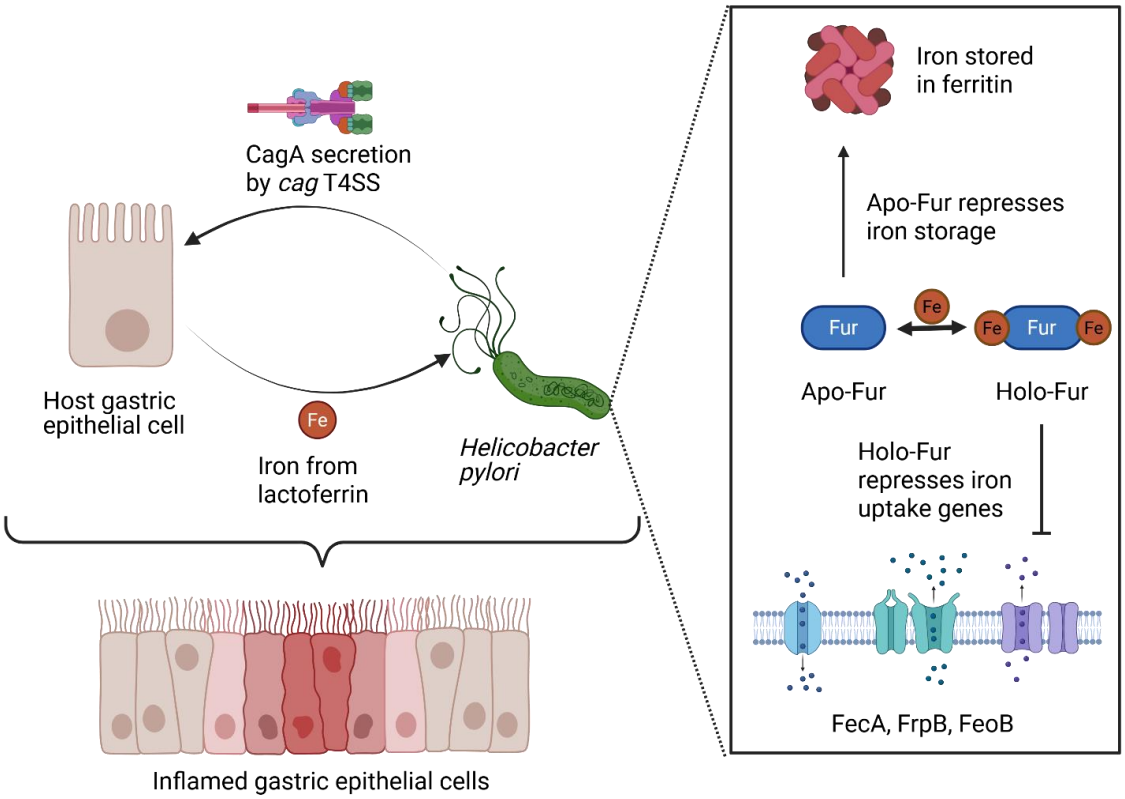


Figure 2: Role of iron in regulation of iron provisioning and bacteria-host interaction in *H. pylori*. Figure created with BioRender.com.

1.7 Research Problems and Aims

At the start of this thesis, the high genetic diversity of *H. pylori* had already been extensively characterized, from large differences between phylogeographic populations to local variations between gastric niches. While the complete methylome of several *H. pylori* strains was known and selected methyltransferases had been characterized, knowledge about the distribution of MTases in the H. pylori global population was patchy, and only a few examples of connections between methylation and gene regulation had been elucidated. In addition, the extent to which target motif frequency varied between *H. pylori* strains had not yet been explored, nor was it clear how methylation might influence the evolution of motif patterns. Furthermore, multiple *H. pylori* MTases had been experimentally associated with the regulation of gene expression, but the mechanism by which methylation induced wider transcriptomic changes beyond their direct effect on promoter regions was unclear.

To address these issues, we aimed to systematically characterize the diversity of the *H. pylori* methylome and searched for signs of evolutionary selective pressures by compar-

ing the distribution of MTases and the frequency of their cognate motifs across phylogeographic populations of *H. pylori*. Specifically, we intended to determine whether the methylome could be considered a distinct layer of epigenetic diversity and to what extent it is evolutionarily intertwined with the layer of genetic diversity. Additionally, we aimed to improve our understanding of the association between methylation and transcription and determine whether and how large downstream indirect transcriptomic effects occur. To achieve that, we first investigated the prevalence of orphan MTases, which do not fulfill a classical function in phage defense, as well as the frequency of potentially regulatory target motifs within promoter regions, and then performed an extensive transcriptomic and functional characterization of a strictly orphan MTase that is highly conserved in *H. pylori*.

communications biology

ARTICLE

https://doi.org/10.1038/s42003-023-05218-x OPEN

Methylome evolution suggests lineage-dependent selection in the gastric pathogen *Helicobacter pylori*

Florent Ailloud [1,2]✉, Wilhelm Gottschall[1] & Sebastian Suerbaum [1,2]✉

The bacterial pathogen *Helicobacter pylori*, the leading cause of gastric cancer, is genetically highly diverse and harbours a large and variable portfolio of restriction-modification systems. Our understanding of the evolution and function of DNA methylation in bacteria is limited. Here, we performed a comprehensive analysis of the methylome diversity in *H. pylori*, using a dataset of 541 genomes that included all known phylogeographic populations. The frequency of 96 methyltransferases and the abundance of their cognate recognition sequences were strongly influenced by phylogeographic structure and were inter-correlated, positively or negatively, for 20% of type II methyltransferases. Low density motifs were more likely to be affected by natural selection, as reflected by higher genomic instability and compositional bias. Importantly, direct correlation implied that methylation patterns can be actively enriched by positive selection and suggests that specific sites have important functions in methylation-dependent phenotypes. Finally, we identified lineage-specific selective pressures modulating the contraction and expansion of the motif ACGT, revealing that the genetic load of methylation could be dependent on local ecological factors. Taken together, natural selection may shape both the abundance and distribution of methyltransferases and their specific recognition sequences, likely permitting a fine-tuning of genome-encoded functions not achievable by genetic variation alone.

[1] Medical Microbiology and Hospital Epidemiology, Max von Pettenkofer Institute, Faculty of Medicine, LMU Munich, Munich, Germany. [2] German Center for Infection Research (DZIF), Partner Site Munich, Munich, Germany. ✉email: ailloud@mvp.lmu.de; suerbaum@mvp.lmu.de

Methylation of DNA is a common epigenetic marker found in nearly all bacteria[1]. It involves the transfer of a methyl group from S-adenosyl-methionine to different positions of the DNA molecule by a DNA methyltransferase (MTase). In bacterial genomes, N^6-methyl-adenine (^{m6}A), C^5-methyl-cytosine (^{m5}C), and N^4-methyl-cytosine (^{m4}C) modifications can be observed. Methyltransferases are often parts of restriction-modification (RM) systems. Presently, four types of RM systems have been described in bacteria[1,2]. Type I RM systems are multimeric enzymes with separate restriction, methylation, and specificity subunits[3]. Type II RM systems have separate restriction endonuclease and methyltransferase enzymes, with the exception of type IIG systems where both activities are either performed by a single protein or with the help of an additional specificity subunit[4]. Type III RM systems also have distinct restriction endonuclease and methyltransferase enzymes, but the endonuclease needs to bind the methyltransferase first in order to be active[5]. Type IV RM systems do not contain a methyltransferase and, unlike the other systems, restrict methylated DNA[6].

Helicobacter pylori is responsible for one of the most prevalent bacterial infections worldwide, affecting more than one-half of the human population[7,8]. It typically leads to chronic active gastritis, which can progress to further complications, such as peptic ulcers, MALT lymphoma, or gastric adenocarcinoma[9]. *H. pylori* is characterized by extensive inter-strain diversity which is the product of a high mutation rate, frequent recombination due to natural transformation, and a large and diverse repertoire of RM systems[10]. Such diversity is thought to be critical to *H. pylori*'s lifelong persistence and exceptional aptitude to adapt to the gastric environment and to evade the host immune responses[11]. To date, various functions have been attached to RM systems and methylation in bacteria[10]. In addition to its central role in distinguishing self from non-self DNA as part of the defense against phages, methylation has also been connected to transcriptional regulation, chromosome replication, stress response, antibiotic resistance, and virulence[12]. In *H. pylori*, several different methyltransferases have been associated with the regulation of gene expression[13–16]. In particular, the M.Hpy99III methyltransferase targeting the GCGC motif has been shown to influence cell morphology, expression of outer membrane proteins, and copper resistance[15]. Nevertheless, the transcriptomic and phenotypic effects associated with GCGC methylation were highly variable between strains, suggesting that genetic background plays a central role in determining the outcome of DNA methylation. The functions of other methyltransferases in *H. pylori* have only been assessed in single strains, and thus strain-specific effects could not be estimated.

On average, two to three RM systems are found in prokaryotic genomes[17,18]. In striking contrast, over 30 can be observed in a given *H. pylori* genome[19]. Only very few methyltransferases belong to the core genome and are strictly conserved in *H. pylori*[15,20]. Accordingly, the majority of methyltransferases are only found in subgroups of strains and thus belong to the accessory genome. This results in a very large number of possible combinations of RM systems, and a highly diverse methylome between strains[21–25]. The variability of RM systems in bacteria is a combination of different mechanisms of horizontal transfer. The target recognition domains (TRD) of type I and type III RM systems can be swapped via recombination to generate new specificities[23,26–28]. Alternatively, complete type II RM systems can be gained and lost by horizontal gene transfer[17,29]. In *H. pylori*, the global frequency and phylogenetic distribution of known RM systems as well as the influence of horizontal transfer have not yet been characterized exhaustively.

Methylation patterns are a combination of many individually methylated target motifs. Across the genome, single methylated motifs located in promoters, coding sequences or translation start sites have been associated with regulation of gene expression in *H. pylori*[14,15,30]. Consequently, changes in the position or frequency of motifs within methylation patterns have the potential to dramatically alter the effects of methyltransferases. Considering the genetic diversity of *H. pylori*, methylation patterns are likely to be substantially different between strains and lineages but such variability has not been investigated yet.

We propose that the methylome, the combination of a diverse repertoire of methyltransferases and a variable distribution of target sequences, represents an entire complex layer of (epigenetic) diversity, distinct yet intertwined with the nucleotide sequence (genetic) diversity of *H. pylori*. Furthermore, the phenotypes that have so far been associated with methylation in *H. pylori* suggest that this layer could contribute to rapid phenotypic diversification and adaptation to the ever-changing gastric environment. To determine the potential contribution of each methyltransferase to epigenetic diversity, we characterized the distribution of methyltransferases and the genomic patterns of methylated target motifs in *H. pylori*. Using a large collection of genomes representative of the geographical diversity of *H. pylori*, we show that type II RM systems are the most conserved in this species and that type II motifs are differentially affected by natural selection. In particular, we detected positive or negative correlations of motif density with the frequency of several type II methyltransferases across phylogeographic populations suggesting some direct evolutionary interplay between RM systems and methylation patterns. Finally, we reconstructed the complex evolution of the ACGT motif targeted by the M.Hpy99XI enzyme and characterized the striking contraction and expansion of this motif following geographically specific environmental factors.

Table 1 Distribution of 96 *H. pylori* RM systems in a globally representative collection of 541 *H. pylori* genomes.

RM system type	Total	Mean frequency (%)	Min-Max frequency (%)[a]
Type I	47	3.0	0.2–12.8
Type II	31	25.3	1.1–100
Type IIG	13	4.7	1.5–12.6
Type III	5	2.2	0.2–5.5

[a]Min-Max frequency indicates the least and most frequent RM systems of a specific type.

Results

Systematic analysis of the diversity of 96 methyltransferases in *H. pylori*. To quantify the variability of methyltransferases in *H. pylori*, we analyzed the distribution of 96 genes related to target-sequence specificity of RM systems (Table 1; Supplementary Data 1) in a collection of 541 genomes representative of the worldwide diversity of *H. pylori* (Supplementary Data 2). Within this collection, we sequenced 32 new *H. pylori* strains from the lesser characterized hpNEAfrica population and 31 additional strains from the hpAfrica1 population.

Most active type I, IIG, and III methyltransferases or specificity subunits were only detected in a small fraction of *H. pylori* genomes, with an average frequency below 5% (Table 1, Fig. 1a; Supplementary Data 3). Type II methyltransferases were by far the most widespread in *H. pylori*. In particular, a subset of ten type II genes was observed in more than 75% of the genome collection. Only two MTases, M.Hpy99III and M.HpyI were present in all genomes. M.Hpy99III targets the motif GCGC and

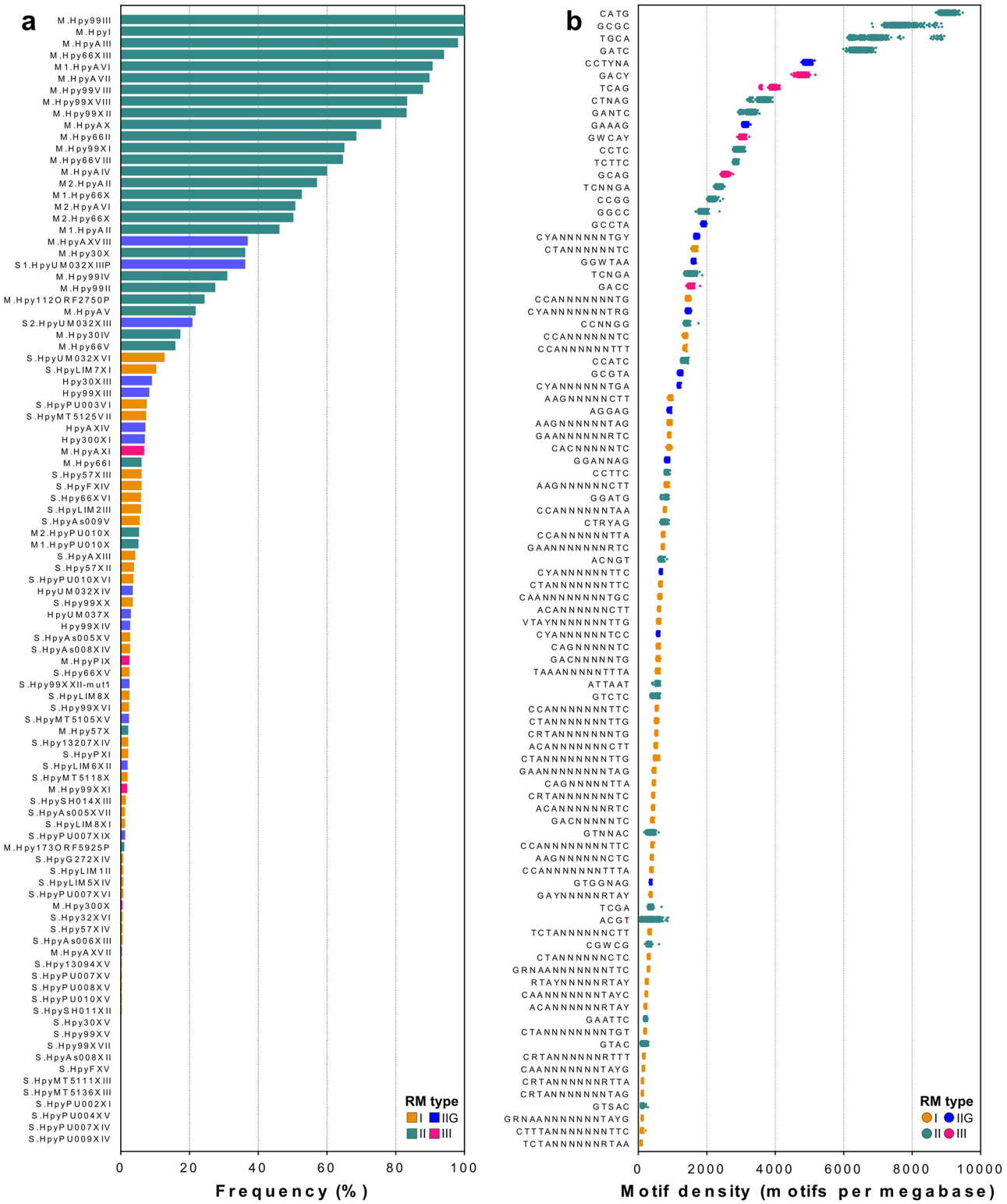


Fig. 1 Distribution of methyltransferases, target-sequence specificity subunits, and target motifs in *H. pylori*. a The frequency of 96 genes related to target-sequence specificity of RM systems was calculated across a collection of 541 *H. pylori* genomes. The gene names are indicated on the *y*-axis. Bars are colored according to the R-M system type. **b** The frequency of 92 target motifs was calculated across the same collection of *H. pylori* genomes. The motif sequences are indicated on the *y*-axis. Each dot represents a single genome and is colored according to R-M system type.

has been associated with the regulation of gene expression[15], while M.HpyI targets CATG motifs and is part of the *iceA* RM system, a potential marker for *H. pylori* strains associated with gastric carcinoma[31,32]. On the opposite of the frequency spectrum, nine type II methyltransferases were detected in less than 25% of genomes. This includes the paired M1.HpyPU010X and M2.HpyPU010X methyltransferases, which methylate the exact same target-sequence motif GGATG and are part of the rare group of RM systems regulated by a controller subunit in *H. pylori*[33,34].

Table 2 Direct flanking repeats in six type II RM systems from *H. pylori*.

MTase	Repeat length (bp)	Inter-strain identity (%)	Intra-strain identity (%)	Prevalence in MTases (%)
M.Hpy99XI	95	86.1	86.2	67.0
M.Hpy66II	115	97.1	98.4	2.5
M.Hpy99IV	395	94.5	97.9	41.7
M.Hpy66I	376	93.1	97.8	21.2
M1.HpyAII/M2.HpyAII	65	92.9	90.9	59.6
M.HpyAIV	118	95.5	93.3	65.8

On average, type I specificity subunits and type II methyltransferases were more conserved (95.9-96.4%) than type IIG and III methyltransferases (91.9–92.8%) (Supplementary Fig. 1a, b). One of the main differences between the type II RM systems and the others is the localization of the target recognition domain (TRD). In type I and a subset of type IIG systems, the TRDs are located in specificity subunits, detached from the endonuclease and methyltransferases genes. Within the same specificity subunit allelic backbone, distinct target sequences can be obtained by recombination between different TRDs. Using phylogenetic and consensus analysis, we were able to group the TRDs from 47 type I S subunits in only five allelic backbones (Supplementary Fig. 2a). Likewise, the five type IIG S subunits actually shared a single backbone (Supplementary Fig. 2b). In type III and a different subset of type IIG systems, the TRDs are located in the methyltransferase. Similar to S subunits, the TRDs in those systems can recombine within similar allelic backbones. We identified three backbones among seven TRDs for type IIG methyltransferases (Supplementary Fig. 2c) and three backbones among five TRDs for type III methyltransferases (Supplementary Fig. 2d). In particular, one type III backbone corresponded to the previously characterized *modH* methyltransferase for which 14 TRDs have been identified so far, although a majority have not been associated with a target sequence yet[35,36]. Consequently, the low frequency of type I, IIG, and III methyltransferases or specific subunits is likely the result of competition for the limited amount of allelic backbones available in *H. pylori*.

In type II systems, the TRDs are not able to recombine and thus methyltransferase loci are associated with a single target sequence and do not compete with each other. Instead, the diversity of type II RM systems is typically based on the gain and loss of specific systems. By examining the local context of type II gene clusters, we identified six type II systems flanked by direct repeats that can potentially lead to spontaneous deletion events (Table 2; Supplementary Data 4)[37–39]. These results were further supported by the observation of a single copy of the repeats at similar genomic locations in strains where these RM systems are absent (Supplementary Fig. 3). Flanking repeats displayed both high inter- and intra-strain identity supporting the possibility of intramolecular recombination. Interestingly, repeats were not systematically found in all alleles of each affected methyltransferase (Table 2). Other type II systems did not display similar repeats and thus are likely gained or lost only through natural transformation. Overall, the frequency of type II methyltransferases was moderately correlated with their nucleotide identity, suggesting that the less prevalent enzymes were simply acquired more recently (Supplementary Fig. 1c). As *H. pylori* is considered to be constitutively competent[40,41], such RM systems could still be transferred by homologous recombination following natural transformation.

Methylation patterns follow different evolutionary trajectories. Variability in the frequency of target motifs has the potential to affect the role of MTases through changes in methylation patterns across the genome of *H. pylori*. Therefore, we determined the frequency of 92 target motifs in our *H. pylori* genome collection. To account for differences in genome length, the total number of motifs for each strain was normalized by the length of its genome and scaled to obtain the density in motifs per megabase (Fig. 1b). The results spread over a 100-fold range of frequencies with the lowest motif density calculated for TCTANNNNNNRTAA (84 motifs/Mb), methylated by a type I system, and the highest obtained for CATG (9036 motifs/Mb), methylated by a type II system. A similar pattern was observed across the whole dataset, with low densities associated with type I motifs and higher densities associated with type II motifs. Intriguingly, the two target sequences with the highest motif densities, CATG and GCGC, are recognized by the two only RM systems whose MTases are fully conserved in *H. pylori* (Fig. 1b). Additionally, we compared the motif frequencies we obtained from whole genomes to ones obtained from a core gene alignment (Supplementary Fig. 4). Seven motifs showed a frequency increase of >25% in the whole versus core comparisons, including the type II motifs ATTAAT and TCGA. However, these motifs have a fairly low frequency overall (min: 91 motif/Mb, max 915 motif /Mb) and thus the differences in absolute number of motifs were relatively small (92 motifs/Mb difference on average). Consequently, the frequency of the majority of motifs appears to be influenced by the evolution of the whole genome rather than the gain and loss of motifs through the accessory genome.

Next, we measured the dependence between motif densities and RM system frequency using the distance correlation method in order to detect non-linear associations (Fig. 2a). A moderate correlation was detected between the two variables (distance correlation measurement dCor=0.48; right-tailed permutation test with 1000 bootstraps $p = 0.009$), which suggests some interdependence between the function of RM systems and the density of their respective motifs. In particular, high-density motifs (>5000 motifs per Mb) were specifically associated with high frequency (>50%) type II RM systems.

The high mutation rate (i.e., approx. 10^{-5} mutations per site per year) characterizing *H. pylori* has the potential to rapidly change methylation patterns[42–44]. Accordingly, we assessed the genetic variability of each motif across the species. Based on a core-genome alignment built from our collection of 541 *H. pylori* genomes, we calculated the average number of motifs shared between genomes and scaled it to the mean number of motifs per genome to determine the stability of each motif pattern (Fig. 2b). The stability of target motifs in *H. pylori* ranged from 25 to 88% with an overall mean of 70%. A moderate correlation was observed between motif stability and motif density (dCor = 0.46, $p = 0.0009$). Interestingly, only motifs methylated by type II RM systems and with very low density (<500 motifs per Mb) displayed a stability below 50%, suggesting that the genomic patterns of these motifs carry a higher genetic load than type II motifs with a higher density (>5000 motifs per Mb) which, in contrast, systematically had a stability above 70%.

In order to look for further evidence of selection pressures on methylation patterns, we calculated the expected frequencies of

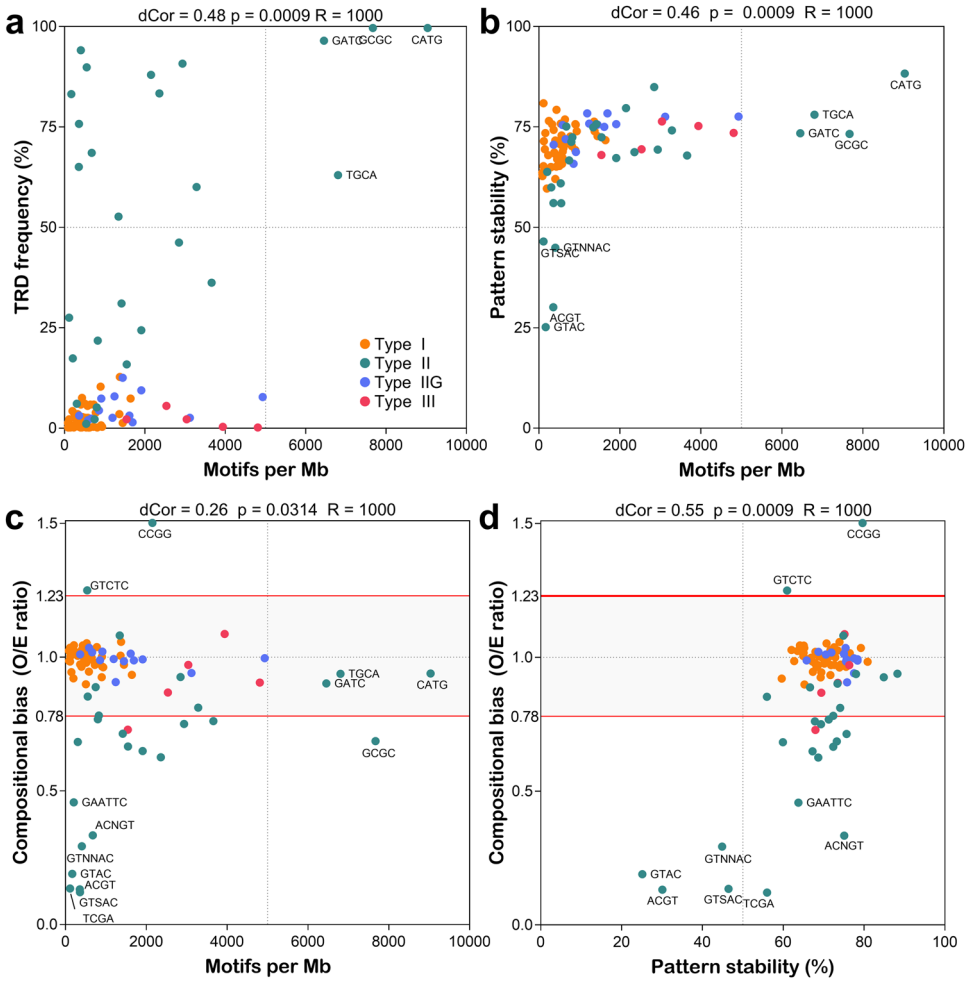


Fig. 2 Interaction between methylome attributes. The correlation between different methylome variables was measured by distance correlation analysis with 1000 bootstrap replicates (R). *p*-values and the distance correlation coefficients are indicated above each scatter plot. **a** Motif density ~ TRD frequency **b** Motif density ~ Pattern stability. **c** Motif density ~ Compositional bias. **d** Pattern stability ~ Compositional bias. Dots are colored according to the RM system enzymatic type. Selected data points are annotated with the corresponding target motif. Cut-offs for under- (<0.78) and over- (>1.23) represented are indicated by red horizontal lines for compositional bias data.

target motifs using probabilistic models based on the nucleotide composition of *H. pylori*[45,46]. By comparing expected and observed frequencies (i.e., compositional bias CB), we determined which motifs were either under- (CB < 0.78) or over-represented (CB > 1.23) and thus potentially under selection (Fig. 2c). As shown in other prokaryotic genomes[18], 22% of the target motifs of type II R-M systems were strongly under-represented with a compositional bias below 0.5. In contrast, one type II motif, CCGG, was over-represented with a compositional bias above 1.5. Additionally, we replicated the compositional bias calculations using two alternative methods and confirmed these observations (Supplementary Fig. 5). Compositional bias did not appear to be a good predictor of motif density and only a weak correlation was observed between these variables (dCor = 0.26, $p = 0.03$). Nevertheless, high-density type II target sequences only displayed a limited amount of under-representation confirming that those motifs are genetically maintained within the species. Compositional bias was, however, strongly associated with pattern stability (dCor = 0.55, $p = 0.0009$). This correlation suggests that the extremely high instability of some motif patterns is likely the result of natural selection pressures that ultimately lead to the removal and an overall under-representation of the motif (Fig. 2d).

Type II methyltransferases have direct positive or negative selective effects on methylation patterns. *H. pylori* is known to exhibit phylogeographic patterns reflecting the ones of its human host, owing to their long co-evolutionary association[47]. Phylogeographic populations of *H. pylori* are genetically distinct from each other, with the most well-known example being the virulence-associated *cag* pathogenicity island (*cag*PAI). For instance, the *cag*PAI displays phenotypical variation between Eastern and Western strains in the first super-lineage of *H. pylori*, while being completely absent in the second super-lineage at the origin of the hpAfrica2 population[48,49]. Epigenetic variations across phylogeographic populations of *H. pylori* have not been studied in comparable detail. Accordingly, we performed clustering analysis to examine the distribution heterogeneity of 31 type II methyltransferases in seven major phylogeographic populations of *H. pylori* (Fig. 3).

The clustering of populations according to their frequency patterns of methyltransferases (Fig. 3) mimicked the phylogeny of *H. pylori*[50]. This suggests that human host migration and geographic isolation contributed to the variability of type II RM systems in *H. pylori*[51]. Furthermore, the clustering of methyltransferases according to their frequencies in *H. pylori* populations (Fig. 3) revealed three distinct clusters: (i) a cluster of 11

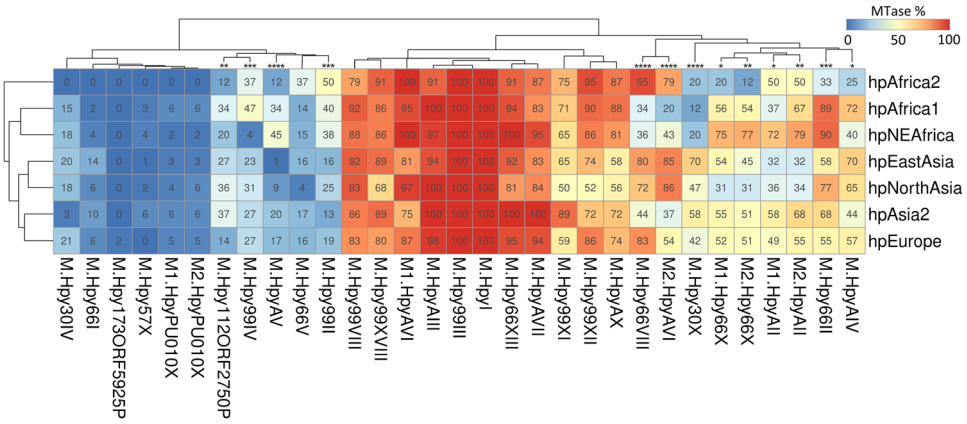


Fig. 3 Geographical variation of type II methyltransferase frequency in *H. pylori*. The frequency of 31 methyltransferases was calculated in 7 geographical populations of *H. pylori* within a collection of 541 genomes. The heatmap is color-coded from blue to red according to the proportion indicated in each cell and is clustered in both axes. Methyltransferases with significant variation between populations are indicated by asterisks (Chi-square *p*-value, * <0.05, ** <0.01, *** <0.001, ****<0.0001).

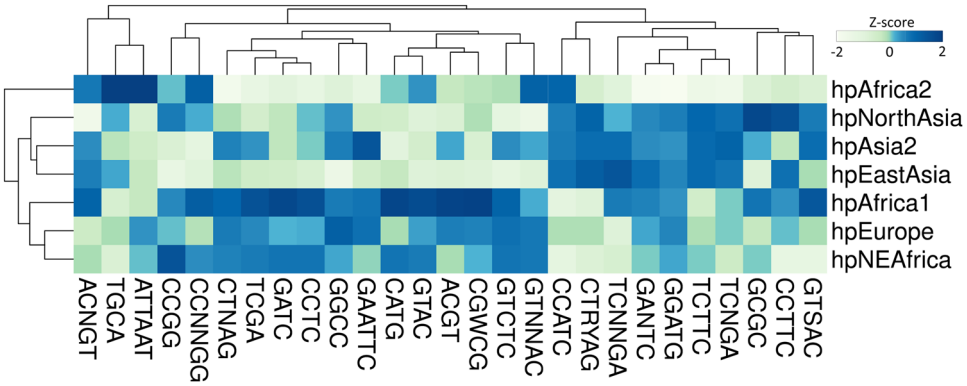


Fig. 4 Geographical variation of type II target motif frequency in *H. pylori*. The density of 27 target motifs (motifs/Mb) was calculated in 7 geographical populations of *H. pylori* within a collection of 541 genomes. The heatmap is color-coded according to the *Z*-score, representing the standard deviation calculated for each motif separately.

"common" enzymes with frequencies ranging from 50 to 100% depending on the phylogeographic population, (ii) a cluster of 11 "rare" enzymes below 50% frequency in every population, and (iii) a cluster of nine "variable" enzymes displaying large variations of frequency across all populations.

We detected significant variation of frequencies between phylogeographic populations of *H. pylori* for 13 methyltransferases (Chi-square test, $p < 0.05$), suggesting that these genes may be affected by positive or negative selection in specific lineages. In the cluster of "rare" enzymes, several methyltransferases, M.HpyAV ($C^{m5}CTTC/GA^{m6}AGG$), M.Hpy99II ($GTS^{m6}AC$) and M.Hpy99IV ($^{m4}CCNNGG$) were overall more frequent in distinct African populations than in Asian populations. In the cluster of "variable" enzymes, M.Hpy66VIII ($TGC^{m6}A$) and M2.HpyAVI ($^{m5}CCTC$) displayed similar patterns across phylogeographic populations. Both were strongly associated with hpEastAsia and hpNorthAsia as well as with hpAfrica2, but were far less common in the other African populations, hpAfrica1 and hpNEAfrica. Additionally, M.Hpy30X ($^{m4}CTNAG$) appeared specifically depleted in all African populations, while M.Hpy66II ($A^{m4}CNGT$) was clearly enriched in the hpAfrica1 and close relative hpNEAfrica populations.

Next, we asked whether motif densities would behave in a comparable way and performed a similar analysis (Fig. 4). Surprisingly, every motif showed a significant variation of density between populations (Kruskal–Wallis test, $p < 0.05$). Because methylation patterns are intertwined with the genome, they are similarly affected by genetic drift and natural selection. Therefore, differences are expected when comparing divergent lineages. Nevertheless, many motifs displayed comparable patterns of density. The largest cluster was composed of motifs with higher densities in hpAfrica1 and related hpNEAfrica and hpEurope populations, while the second largest cluster contained motifs with increased densities in all Asian lineages. On the contrary, only a small number of motifs appeared to have expanded in the hpAfrica2 population. Overall, the largest absolute variation of density was observed for the TGCA motif with ~3000 motifs in hpAfrica1 and ~4500 motifs in hpAfrica2. On a relative scale, the greatest variation was observed for the ACGT motif with ~50 motifs in hpEastAsia and ~400 motifs in hpAfrica1, representing an 8-fold difference.

Finally, we sought to investigate the relationship between methyltransferase and target motif density across phylogeographic populations in order to determine if methylation can directly influence the evolution of motif patterns. Consequently, we used logistic regression to investigate how the frequency of methyltransferases affects the motif density in phylogeographic populations. We found significant positive and negative interactions for four and three methyltransferases, respectively (Supplementary Data 5). In this context, a positive interaction indicates that the motif density increases as the methyltransferase frequency increases too while a negative interaction

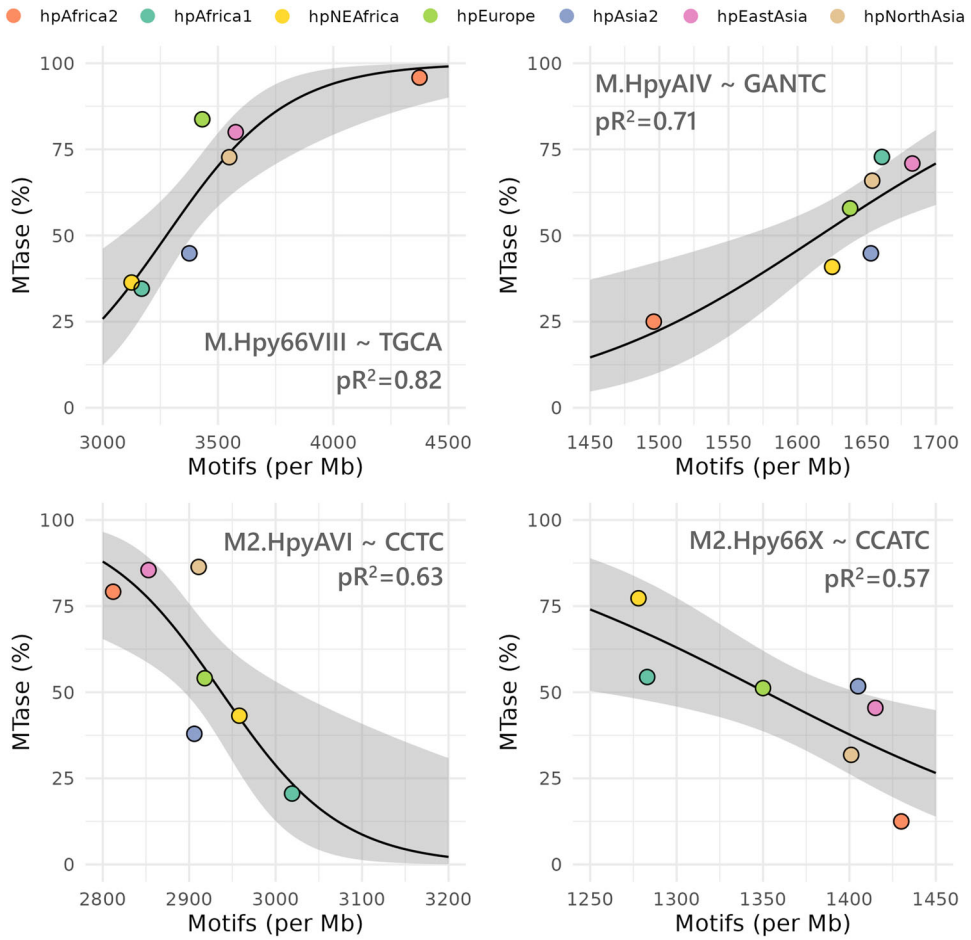


Fig. 5 Positive and negative interaction of type II methyltransferases with motif patterns. The interactions between methyltransferases and motifs were estimated using a generalized linear model with a quasibinomial distribution and a logit link function. A 95% confidence interval is indicated by a gray ribbon. Goodness-of-it was assessed using a chi-square test ($p < 0.05$). Pseudo-R^2 (McFadden) was calculated using the ratio of residual deviance over the null deviance and is indicated on the plots.

indicates the opposite. Two examples of each type of interaction are displayed in Fig. 5. Interestingly, the hpAfrica1 and hpAfrica2 populations were at the opposite ends of the spectrum for each interaction. Notably, the motif density of TGCA increased from ~3000 to 4500 with the frequency of the cognate methyltransferase M.Hpy66VIII going from ~30 to ~100%. In contrast, an increase of ~50% of the M2.HpyAVI frequency was associated with a decrease from ~3000 to 2800 in the CCTC motif density.

The presence of the cognate restriction endonuclease in type II RM systems has been shown to lead to the avoidance of palindromic motifs in several bacterial species[52–54]. The Hpy99III (GCGC) and HpyI (CATG) type II systems are known to be facultatively lacking an endonuclease (i.e., orphan methyltransferases) in *H. pylori*[15,55]. In the same way as for methyltransferases, we tested the relationship between the frequency of the endonuclease and the motif density for these two RM systems but did not find any significant interaction (Supplementary Fig. 6).

These results suggest a direct selective effect of methylation on individually methylated motifs. Furthermore, the existence of both positive and negative interactions indicates that the evolution of motif patterns is highly specific to each methyltransferase and implies that distinct RM systems might fulfill specific functions with overarching effects on the fitness of *H. pylori*.

Lineage-specific expansion and contraction of the type II m5c motif ACGT. Among type II motifs, the ACGT motif displayed the largest relative variation of density between phylogeographic populations of *H. pylori*. However, these differences were not correlated with the frequencies of the cognate methyltransferase across populations (Supplementary Data 3). For instance, the Hpy99XI methyltransferase was present in 65% of the hpEastAsia strains which only contained around 53 ACGT motifs/Mb, but was observed in 71% of the hpAfrica1 genomes which contained 295 motifs/Mb (Supplementary Fig. 7). In our global analysis of methylated motifs in *H. pylori*, the ACGT target sequence was also one of the most highly unstable and under-represented motifs (Fig. 2). In order to determine the underlying causes for the peculiar evolution of ACGT, we first compared the genomic patterns of this motif to determine the overlap across four representative phylogeographic populations (hpAfrica1, hpAfrica2, hpAsia2, and hpEastAsia). A minimal number of motifs were shared across all strains and most motifs were completely specific to each population (Fig. 6a). This result echoes the low pattern stability observed previously and suggests that the ACGT motif underwent rapid evolution. Intriguingly, the proportion of motifs shared between populations did not quite reflect the evolutionary history of *H. pylori*. For instance, the number of motifs shared between hpAfrica1 and hpAfrica2 was higher than between hpAfrica1 and hpEastAsia. At the same time, the number of motifs specific to each population was also highly variable

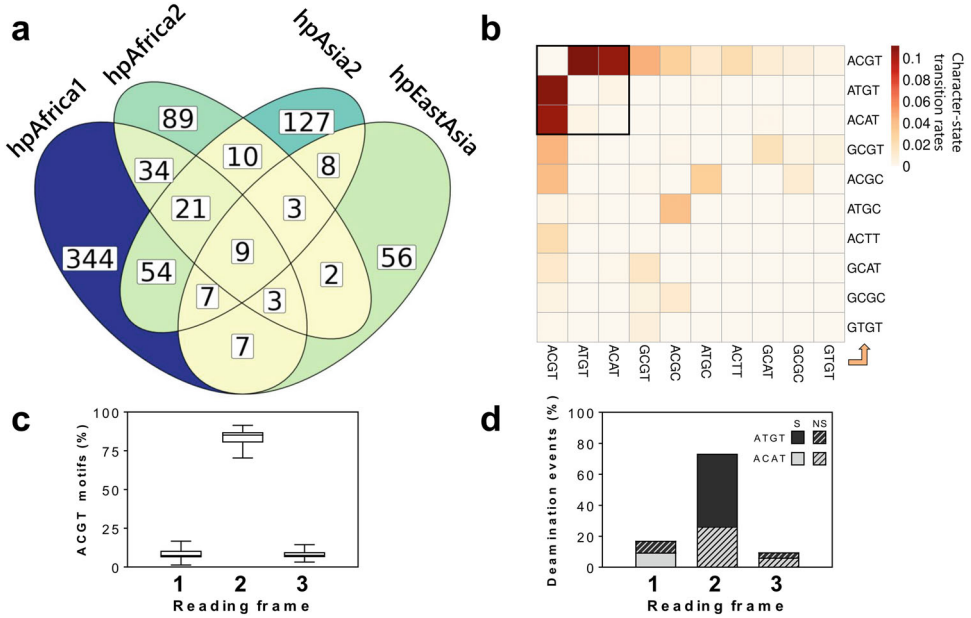


Fig. 6 Genomic patterns of the ACGT target motif. a Overlap of motif patterns between representative populations of *H. pylori*. All genomes included in this analysis were compared in a pairwise fashion to determine how many motifs are present at the same genomic coordinates in each pair and the average between populations is presented. Sections of the Venn diagram are colored from yellow to blue according to the number of overlapping motifs. **b** Character-state transition rates between ACGT and its most frequent allelic variants. The direction of character-state change is displayed from the *x*-axis to the *y*-axis. **c** Distribution ACGT motifs located within coding sequences according to their reading frames and represented by a box plot. The median is indicated by the central line across the box. The lower and upper hinges represent the 25th and 75th percentile, respectively. The ends of the lower and upper whiskers represent the minimum and maximum data points, respectively. **d** Effect of state transition between ACGT and ATGT/ACAT according to the reading frame (Syn: synonymous; Non-syn: non-synonymous).

which suggests overall that the genomic patterns of the ACGT were shaped by very specific differences in local environments.

Next, we used a representative subset of ten genomes each from the four major phylogeographic populations of *H. pylori* to perform ancestral-state reconstruction analysis and understand how ACGT evolved during the divergence of these lineages. Based on the ACGT motif patterns observed in modern strains of *H. pylori*, the ancestral-state reconstruction method can identify the ancestral phylogenetic nodes in which individual motifs were either gained or lost. By combining the reconstruction of all ~7000 unique positions in which ACGT motifs were observed in our subset of 40 representative genomes, we determined the transition rates between ACGT motifs and its most frequent allelic variants (Fig. 6b). The transition matrix was heavily skewed toward the ATGT and ACAT variants. The M.Hpy99XI enzyme, targeting the ACGT motif, is a ^{m5}C methyltransferase. Spontaneous deamination of 5-methylcytosine to thymine is known for making ^{m5}C motifs more prone to mutations compared to ^{m6}A or ^{m4}C motifs[56]. Interestingly, in the case of the ACGT motif, deamination would produce either ATGT or ACAT variants (depending on the affected DNA strand), suggesting that deamination played a major role in the evolution of this motif. To understand the potential fitness cost of the ACGT motif mutations, we focussed next on motifs located within coding sequences. In particular, we determined in which reading frame those motifs are positioned and found that they are heavily biased towards the second frame (i.e., motif starting at the second base of codons) (Fig. 6c). Furthermore, the majority of deamination events resulted in synonymous mutations, with no effect on the encoded protein sequences (Fig. 6d). Mutations in the second frame were skewed toward synonymous ATGT variants, corresponding to deamination of ACGTs motif on the sense strand of coding sequences (Fig. 6d). These results indicate that deamination of methylated ACGT motifs would mostly result in silent mutations and would likely not cause any secondary effects outside of the loss of methylation (and vice versa, gain of ACGT motifs from ATGT/ACAT would be mostly silent). Consequently, the evolution of the ACGT motif itself seems to be partially constrained by the purifying selection influencing coding sequences, affecting the rate at which each strand gets deaminated and minimizing its effect on protein-coding sequences. Accordingly, the main effect of ACGT changes is solely the loss of methylation markers and thus any natural selection pressure associated with the evolution of A^{m5}CGT motif patterns is presumably operating at the epigenetic level without being hindered by its effects at the genetic level.

Finally, we used our ancestral-state dataset to reconstruct the expansion and contraction of the ACGT motif patterns at each major ancestral node as well as at the root (Fig. 7). Our analysis indicates that ~130 ACGT motifs were present in the common ancestor of *H. pylori* at the root, dated at ca. 100,000 years according to previous studies[49]. From there, the motif pattern expanded at the hpAfrica1/Asian node whereas it contracted in the hpAfrica2 node mainly because of deamination. The expansion continued between the hpAfrica1/Asian node and the hpAfrica1 leading to a higher number of ACGT motifs observed in modern Africa1 strains. On the opposite, the motif patterns contracted between the hpAfrica1/Asian and the Asian nodes leading to a severe reduction in the number of motifs. Interestingly, the contraction seemingly slowed down in the hpAsia2 node but appeared to accelerate in the East-Asia node, explaining the unusually low number of motifs seen in the modern East-Asian *H. pylori* population. Overall, the population-specific trends of contraction and expansion across those genetically distinct geographical populations indicate that external factors, such as the host physiology or the local environment, are having a selective effect on the evolution of the ACGT methylation patterns. Furthermore, the strong positive selection

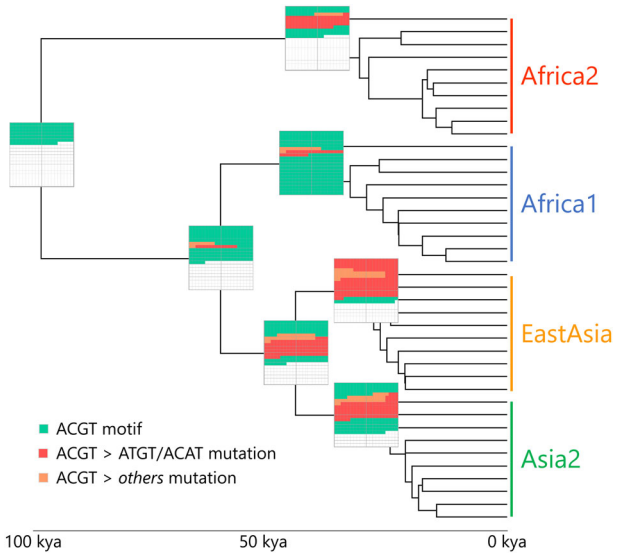


Fig. 7 Ancestral reconstruction of the ACGT motif patterns across *H. pylori* major phylogeographic populations. A time-scaled tree was created using 40 representative strains from the hpAfrica1, hpAfrica2, hpEastAsia, and hpAsia2 populations. The ancestral states for each genomic position displaying ACGT motifs in modern strains were first predicted separately by stochastic character mapping and then concatenated to generate the ancestral patterns. A graphical representation of the ACGT motif patterns is given by boxes located at each ancestral node. Green squares represent new individual ACGT motifs, while red and orange squares indicate ACGT motifs lost through ATGT/ACAT mutations and other mutations, respectively.

pressures which are likely required to expand and maintain the ACGT motifs in hpAfrica1, in contrast to the negative selection pressures needed to purge most of the motifs in hpEastAsia, would suggest the ACGT motif may have a role in the evolutionary fitness of *H. pylori*.

Discussion

H. pylori is a bacterial pathogen with exceptionally high genetic diversity whose phylogenetic structure reflects the one of its host[10,47]. The long and intimate association between *H. pylori* and humans[47] has likely been sustained by the ability of this organism to rapidly adapt to the harsh environment characteristic of the gastric habitat[11]. The diversity of *H. pylori* is believed to be the main factor underlying its capacity for host adaptation. While the genetic variability of *H. pylori* has been extensively characterized, from its worldwide phylogeographic structure to its within-host diversity across stomach niches[10], the evolution of its methylome has not yet been investigated in the context of its global population structure. In this study, we undertook an in-depth characterization of the distribution of RM systems as well as the evolution of methylation patterns across the phylogenetic spectrum of *H. pylori*.

We started by investigating the frequency of the 96 RM systems characterized so far in *H. pylori* among a genetically diverse collection of 541 genomes. Type II methyltransferases were by far the most conserved in the species. This result is likely tied to their genetic organization and the separation of the endonuclease and methyltransferase activities. In particular, this organization can lead to post-segregational killing from residual endonuclease activity after the loss of type II RM systems[38,57]. Nevertheless, the two methyltransferases that are fully conserved in *H. pylori* are not always attached to an endonuclease (i.e., orphan MTases)[15,32], indicating that these methyltransferases are likely maintained in the genome of *H. pylori* because they provide an evolutionary benefit. Furthermore, for both systems, the presence of the endonuclease did not appear to influence methylation patterns. On the other hand, the frequency of other types of RM systems (I, IIG, and III) was strongly limited by the fact that multiple TRDs have to compete for a limited number of methyltransferase or specificity subunit allelic backbones. The specificity of type I RM systems is determined by two TRDs located in the S subunits[23,58]. The shuffling of TRDs via recombination can produce an almost infinite number of target motifs[28,59], which is reflected by the large number of type I motifs found in *H. pylori*. For type IIG and type III RM systems, the specificity is determined by single TRDs, which can also be transferred between allelic backbones. Intriguingly, only a single backbone was identified among all the type IIG RM systems controlled by a specificity subunit. The most variable type III RM system in *H. pylori* is *modH* with 17 distinct TRD identified so far[35]. However, the target motifs have only been characterized for three TRDs of *modH*, indicating that type III systems need to be characterized further.

The target motifs methylated by type II methyltransferases were overall more frequent than the ones from other types of RM systems. Globally, the frequency of methyltransferases was only moderately correlated with the densities of their cognate target sequences. As a notable exception, the only two universally conserved *H. pylori* MTases recognize the two by far most abundant motifs (GCGC and CATG). Similarly, high-density type II motifs, such as CATG and GCGC, were characterized by high pattern stability and limited compositional bias. Both the MTases M.HpyI ($C^{m6}ATG$) and M.Hpy99III ($G^{m5}CGC$) have been shown to influence the expression of many genes in *H. pylori*[15,16]. The regulation of gene expression via methylation has also been shown for other *H. pylori* methyltransferases[13,14,16], as well as in other bacterial species[60–62]. While several studies have pointed out a potential role of target motifs within promoter elements or coding sequences[14–16,30], the transcriptional mechanisms have not been clearly characterized yet. Moreover, the strain specificity of gene regulation observed in these studies suggests a role of variable methylation patterns. On the opposite, the under-representation of multiple type II motifs in the genome of *H. pylori* was strongly associated with the lower stability of their motif patterns. Consequently, these highly unstable motifs are likely under selection pressures that gradually remove them from the species. Overall, the large variation in compositional bias and pattern stability among target motifs indicates that natural selection pressures do not affect the methylome globally but are more likely specific for each methyltransferase and cognate motif, suggesting these serve different functions in the biology of *H. pylori*. Motif avoidance is a phenomenon known for causing the depletion of restriction sites in bacterial genomes in order to limit self-restriction while still maintaining active endonucleases required for phage defense[45,52–54,63]. Interestingly, prophages in *H. pylori* also display a phylogeographic structure[64–66]. This effect is typically more pronounced in type II RM systems, which is likely due to their fixed TRD and higher diversity compared to other types[63]. Despite the presence of many cognate endonucleases, type II motifs are obviously not all affected similarly by motif avoidance. This could be explained by target motifs having distinct susceptibilities to self-restriction, or, the existence of additional selection pressures on specific motifs, counter-acting motif avoidance. Furthermore, context-dependent mutations may also be responsible for the instability of some motifs. For example, 5-methylcytosine is prone to spontaneous deamination and thus typically displays an increased mutation rate. Nevertheless, m5c motifs show large differences in terms of compositional bias

which also suggests that specific motifs are either maintained or lost via additional factors.

As the evolution of *H. pylori* is intimately linked to human migrations and geographical isolation, we investigated the distribution of type II methyltransferases and target motifs among the phylogeographic populations of *H. pylori*. Our analysis revealed a subset of nine methyltransferases showing large variations of frequency between populations. Four of these methyltransferases are flanked by direct repeats, which most likely contribute to their higher variability. While the frequency of methyltransferases was likely heterogeneous between ancestral populations of *H. pylori* due to founder effects following human migration events, the distribution of these methyltransferases in modern panmictic populations of *H. pylori* is either explained by either pure genetic drift or geographically dependent fitness effects. Subsequently, we identified correlations between methyltransferase frequency and motif density for seven type II RM systems, suggesting that methylation can indeed shape motif patterns via natural selection. Interestingly, we observed both positive and negative correlations. Negative correlations indicate that the presence of the methyltransferase leads to the elimination of its cognate motif. In addition, selection pressure leading to the depletion or enrichment of motifs might also be driven by the host immune system[67,68]. By contrast, positive correlations imply that the presence of the methyltransferase leads to a positive selection of its cognate motif. Direct selective effects on methylation patterns leading to the enrichment and/or maintenance of a motif in the genome have not been described and suggest that specific methylation patterns can contribute to the evolutionary success of *H. pylori*. This effect was particularly evident for the M.Hpy66VIII methyltransferase targeting the motif TGCA. In this case, the gradient of methyltransferase frequency and motif density distinguished the hpAfrica1/hpNEAfrica, hpEastAsia/hpNorthAsia, and hpAfrica2 populations. These specific groups of populations are highly divergent from each other. In particular, the CagA virulence factor is functionally distinct in Western versus Eastern populations of *H. pylori* while the *cag*PAI T4SS is completely absent in hpAfrica2 since it descends from a separate super-lineage than the other populations. To date, the role of M.Hpy66VIII has not been investigated but the maintenance of this methyltransferase and TGCA motif patterns in hpAfrica2 is likely related to specific local environmental factors.

In specific cases, local environmental factors may have selective effects on methylation patterns leading to geographical variation, independently of the frequency of the methyltransferase. Our regression analysis suggests that fluctuations in MTase frequency could only account for differences in motif densities in ~20% of the cases. In particular, the ACGT motif displayed the highest relative change in density across phylogeographic populations but showed no correlation with methyltransferase frequency. Demographic bottlenecks and the rapid evolution of *H. pylori* following repeated human migration events[69–71] most likely precipitated the evolution of the ACGT methylation pattern in the species. We hypothesize that the striking difference in the evolution of the ACGT motif between the hpAfrica1 and hpAfrica2 populations could be related to the acquisition of the *cag* pathogenicity island in the former[49]. The *cag*PAI is thought to provide a fitness advantage to *H. pylori* and to have contributed to the spread of hpAfrica1 through Africa and subsequently to other regions of the world[48]. How could the low density of the ACGT motif observed in Asian populations be explained in this scenario? Our analyses suggest that the deamination of 5-methylcytosine was one of the main drivers of motif depletion for ACGT. Since it is well established that Asian variants of *cag*PAI components, and CagA in particular, are associated with stronger inflammation and ultimately carcinogenicity[72–74], we speculate that increased host cell interaction and inflammatory response may have contributed to increased deamination and hence caused loss of ACGT motifs in Asian populations. Furthermore, because of the placement of ACGT motifs within coding sequences, transitions between ACGT and its deaminated variants are mostly silent, facilitating the evolvability of this motif. The evolution of the ACGT motif in the methylome context is thus strongly separated from its genomic context and thus any hypothetical selective effects involved in this process would be mainly driven by the epigenetic status of this motif rather than its genetic sequence. As speculated above, the various trajectories taken by the ACGT motif pattern from the common ancestor to the modern populations suggest that local environmental cues can greatly affect the genetic load of methylation and thus the epigenetic landscape of *H. pylori*. The fact that the GCGC motif was neither underrepresented nor unstable additionally points to the specificity of evolutionary pressures affecting 5-methylcytosines.

In conclusion, the methylome of *H. pylori* is a major contributor to its overall variability. Because the evolution of methylation patterns is constrained by their genetic sequence and the distribution of RM systems is influenced by their gene organization, the methylome is completely intertwined with the genetic variation of *H. pylori* and dependent on the phylogeographic structure of the species. Yet, the methylome is also shaped independently by selection pressures able to expand or contract motif patterns as a direct result of methylation, and environmental factors whose selective effects appear dependent on specific motifs and lineages. Third-generation sequencing technologies have permitted the rapid discovery of many new methyltransferases and the characterization of their target sequences in diverse bacterial species. Quantitative frameworks, such as the one expanded in this study, will contribute to the identification of methyltransferases whose functions extend beyond the standard phage defense model.

Methods

Construction of a worldwide *H. pylori* genome collection. Genome assemblies of *H. pylori* were acquired from the Enterobase database[75]. An additional 63 isolates from the hpAfrica1 and hpNEAfrica population[76] were sequenced on an Illumina MiSeq (2 x 300bp) and assembled with spades 3.15.4 using the –careful and –only-assembler parameters[77] in order to complete the collection. Phylogeographical population assignments were obtained from previous population genetic studies[70,71,76,78]. Sequences with quality (>1000 ambiguous bases) and assembly (>100 contigs) issues were discarded. Based on the *H. pylori* MLST scheme[79], closely related strains were identified (>3 identical MLST alleles) and discarded. The 541 genomes selected and analyzed for this study are listed in Supplementary Data 2, including their phylogeographical population.

Type I, II(M/G), and III RM system gene sequences. Genes modulating target-sequence specificity (i.e., genes containing the TRD region) in *H. pylori* were collected from the REBASE database[80]. Depending on the type of RM systems, either the methyltransferase (type II and type III), the specificity subunit (type I and some type IIG), or the RM fusion (type IIG) were selected. The 96 genes analyzed in this study are listed in Supplementary Data 1, including their enzymatic characteristics. All RM systems analyzed in this study are encoded on the chromosome of *H. pylori*. The activities of 88 *H. pylori* methyltransferases analyzed in this study have been previously validated with PacBio Single Molecule Real-Time (SMRT) sequencing data in at least one *H. pylori* strain as indicated in Supplementary Data 1. Among the motifs not validated by PacBio data, two m5c motifs were validated by bisulfite sequencing, four motifs were validated by different methods (see additional details in Supplementary Data 1 and on REbase in corresponding *H. pylori* strains). The methylation of two motifs, GTCTC and CRTANNNNNNNTAG, has not yet been validated experimentally in *H. pylori* and has only been inferred by homology with methylases from other species.

Distribution of RM system genes in the genome collection. The genome collection was annotated using the *Helicobacter pylori* genus and species database from Prokka v1.14.5[81] and GNU parallel[81]. Homologs of the RM system genes were searched with the megablast algorithm implemented in BLAST + 2.12.0[82] against a database built with annotated coding sequences (CDS). BLAST hits on single CDS with above 80% nucleotide identity (or 90% for genes undergoing

domain movement) and 70% query coverage were considered positive. BLAST hits below 80% nucleotide identity, 70% query coverage, or including fragmented CDS were considered negative. The frequency for each gene was calculated for both the entire complete collection and for each geographical population. Results were compared to frequencies using blastp and amino acid sequences, instead of blastn and nucleotide sequences, and similar results were obtained (Supplementary Data 6). The frequency in each population was represented as a heatmap with the pheatmap 1.0.12 R package and clustered on both axes using the average-linkage clustering method. Variability between each population was tested using Pearson's chi-square test with Yates' continuity correction.

Analysis of target motifs frequencies and genomic patterns. Target-sequence motifs were detected individually in each genome using the Biostrings 2.58.0 Bioconductor R package. For paired methyltransferases that recognize either the exact same motif (M1/M2.HpyPU010X) or complementary non-palindromic motifs (M1/M2.Hpy66X, M1/M2.HpyAVI, and M1/M2.HpyAII), only the motif targeted by the first enzyme (i.e., M1) was considered. The frequency of each motif was scaled to the length of each genome to obtain a normalized motif/Mb unit and plotted either as a bar chart or as a heatmap with the pheatmap 1.0.12R package according to the *Z*-score calculated individually for each motif. The heatmap data was clustered on both axes using the average-linkage clustering method. Variability between each population was tested using Kruskal–Wallis rank sum test (non-parametric one-way analysis of variance). Expected frequencies and compositional bias for each motif were calculated using the methods of Burge and co-authors[46], Pevzner and co-authors[83], and a maximum-order Markov chain model implemented in CBcalc[45]. Additionally, a core gene alignment was created using Roary[84] with -i 80 and -cd 90 parameters and used to calculate the motif frequency in the core genome.

A consensus genome alignment was created by mapping assemblies onto the reference strain BCM300 (RefSeq NZ_LT837687) with BWA 0.7.17[85] (bwa mem with default parameters) and generating consensus sequences with bcftools 1.15.1[86]. Uncovered regions of the reference sequence were masked using the genomecov and substract tools from bedtools 2.30.0[87]. Coverage and pairwise identity data are available for each strain included in the consensus alignment in Supplementary Data 7. Pattern stability was determined by calculating the average number of motifs shared between all pairs of genomes (i.e., motifs located at the same position) in the consensus alignment and reporting it as a proportion of the mean number of motifs per genome. The consensus genome alignment was also used to produce the Venn diagram representing the overlap of ACGT motifs between phylogeographical populations and the gene reading frame analysis of ACGT motifs.

The dependence between methyltransferase frequencies, motif frequencies, pattern stability, and compositional bias was calculated using the distance correlation method implemented in the dcor.test() function from the energy R package. The distance correlation method is a non-parametric test of multivariate independence with the statistical significance evaluated by permutation bootstrap. *p*-values from a right-tailed test are reported.

The interaction between type II methyltransferases and motif densities across geographic populations was determined using a generalized linear model with a quasibinomial distribution and a logit link function, implemented in the glm R package[88]. Goodness-of-it was assessed using a chi-square test ($p < 0.05$) and a pseudo-R^2 (McFadden) calculated with the ratio of residual deviance over the null deviance.

Ancestral-state reconstruction of the ACGT motif patterns. A random selection of 40 genomes belonging to the main phylogeographical populations hpAfrica1, hpAfrica2, hpAsia2, and hpEastAsia was used as a representative group of *H. pylori* diversity to perform ancestral-state reconstruction analysis. The smaller size of this group compared to the main genome collection ensured a balanced number of genomes per population and reduced the computational complexity of the analysis. A core-genome alignment was created as described above and a phylogenetic tree was produced using iqtree[89] and the TVM + F + R7 substitution model determined by ModelFinder[90]. A time-calibrated tree was generated using the ape R package[91]. Marginal reconstruction of ancestral states was carried out using the stochastic mapping method implemented in the phytools R package[92]. An evolutionary model with fully independent ("all-rates-different") transition rates was selected based on AIC scores and quality of reconstruction. The transition matrix Q was fitted using a continuous-time reversible Markov model (Q = "empirical") and the prior distribution pi on the root of the tree was estimated using the tip character states. Reconstruction was performed for each position containing an ACGT motif in the core-genome alignment. Each allelic variant was considered a distinct state within the reconstruction. The global character-state transition rate matrix was obtained by averaging the transition rates of all individual reconstruction events (the ten most frequent events are displayed in Fig. 6). Gain and loss of ACGT motifs were estimated by first selecting the state with the highest likelihood at each major node of the *H. pylori* tree for each reconstruction (i.e., root and TMRCA nodes of each phylogeographical populations). The selected ancestral states for each node were then added up across all reconstructions and classified into three groups: (1) ACGT motifs, (2) ACAT/ATGT variants (i.e., deamination), and (3) all other allelic variants.

Statistics and reproducibility. All data were analyzed using R version 4.1.2 or GraphPad Prism version 7.04. The dependence between methyltransferase frequency, motif frequency, pattern stability, and compositional bias was evaluated across $n = 31$ methyltransferases using the distance correlation method (right-tailed test). The variability of $n = 31$ methyltransferase frequencies between $n = 7$ geographic population of *H. pylori* was tested using Pearson's chi-square test with Yates' continuity correction. The variability of $n = 27$ target motif frequencies between $n = 7$ geographic populations of *H. pylori* was tested using the Kruskal–Wallis rank sum test. The interaction between type II methyltransferases and motif densities across $n = 7$ geographic populations was determined using a generalized linear model with a quasibinomial distribution and a logit link function with the goodness-of-it was assessed using a chi-square test and a pseudo-R^2 (McFadden) calculated with the ratio of residual deviance over the null deviance.

Reporting summary. Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The dataset supporting the conclusions of this article is available in the NCBI SRA repository, BioProject accession no. PRJNA914092. Source data for the main figures can be found in Supplementary Data 8.

Received: 9 February 2023; Accepted: 4 August 2023;
Published online: 12 August 2023

References

1. Blow, M. J. et al. The epigenomic landscape of prokaryotes. *PLoS Genet*. **12**, e1005854 (2016).
2. Anton, B. P. & Roberts, R. J. Beyond restriction modification: epigenomic roles of DNA methylation in prokaryotes. *Annu. Rev. Microbiol* **75**, 129–149 (2021).
3. Loenen, W. A., Dryden, D. T., Raleigh, E. A. & Wilson, G. G. Type I restriction enzymes and their relatives. *Nucleic Acids Res.* **42**, 20–44 (2014).
4. Pingoud, A., Wilson, G. G. & Wende, W. Type II restriction endonucleases–a historical perspective and more. *Nucleic Acids Res.* **42**, 7489–7527 (2014).
5. Rao, D. N., Dryden, D. T. & Bheemanaik, S. Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Res.* **42**, 45–55 (2014).
6. Loenen, W. A., Dryden, D. T., Raleigh, E. A., Wilson, G. G. & Murray, N. E. Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res.* **42**, 3–19 (2014).
7. Hooi, J. K. Y. et al. Global prevalence of *Helicobacter pylori* infection: systematic review and meta-analysis. *Gastroenterology* **153**, 420–429 (2017).
8. Malfertheiner, P. et al. *Helicobacter pylori* infection. *Nat. Rev. Dis. Prim.* **9**, 19 (2023).
9. Suerbaum, S. & Michetti, P. *Helicobacter pylori* infection. *N. Engl. J. Med.* **347**, 1175–1186 (2002).
10. Ailloud, F., Estibariz, I. & Suerbaum, S. Evolved to vary: genome and epigenome variation in the human pathogen *Helicobacter pylori*. *FEMS Microbiol. Rev.* **45**, https://doi.org/10.1093/femsre/fuaa042 (2021).
11. Suerbaum, S. & Josenhans, C. *Helicobacter pylori* evolution and phenotypic diversification in a changing host. *Nat. Rev. Microbiol.* **5**, 441–452 (2007).
12. Sanchez-Romero, M. A., Cota, I. & Casadesus, J. DNA methylation in bacteria: from the methyl group to the methylome. *Curr. Opin. Microbiol.* **25**, 9–16 (2015).
13. Kumar, R., Mukhopadhyay, A. K., Ghosh, P. & Rao, D. N. Comparative transcriptomics of *H. pylori* strains AM5, SS1 and their *hpyAVIBM* deletion mutants: possible roles of cytosine methylation. *PLoS ONE* **7**, e42303 (2012).
14. Kumar, S. et al. N4-cytosine DNA methylation regulates transcription and pathogenesis in *Helicobacter pylori*. *Nucleic Acids Res.* **46**, 3429–3445 (2018).
15. Estibariz, I. et al. The core genome ^{m5}C methyltransferase JHP1050 (M.Hpy99III) plays an important role in orchestrating gene expression in *Helicobacter pylori*. *Nucleic Acids Res.* **47**, 2336–2348 (2019).
16. Yano, H. et al. Networking and specificity-changing DNA methyltransferases in *Helicobacter pylori*. *Front. Microbiol.* **11**, 1628 (2020).
17. Oliveira, P. H., Touchon, M. & Rocha, E. P. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.* **42**, 10618–10631 (2014).
18. Rusinov, I., Ershova, A., Karyagina, A., Spirin, S. & Alexeevski, A. Lifespan of restriction-modification systems critically affects avoidance of their recognition sites in host genomes. *BMC Genomics* **16**, 1084 (2015).
19. Vasu, K. & Nagaraja, V. Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol. Mol. Biol. Rev.* **77**, 53–72 (2013).
20. Xu, Q., Morgan, R. D., Roberts, R. J. & Blaser, M. J. Identification of Type II restriction and modification systems in *Helicobacter pylori* reveals their

substantial diversity among strains. *Proc. Natl Acad. Sci. USA* **97**, 9671–9676 (2000).
21. Krebes, J. et al. The complex methylome of the human gastric pathogen *Helicobacter pylori*. *Nucleic Acids Res.* **42**, 2415–2432 (2014).
22. Nell, S. et al. Genome and methylome variation in *Helicobacter pylori* with a cag Pathogenicity Island during early stages of human infection. *Gastroenterology* **154**, 612–623 (2018).
23. Furuta, Y. et al. Methylome diversification through changes in DNA methyltransferase sequence specificity. *PLoS Genet.* **10**, e1004272 (2014).
24. Lee, W. C. et al. The complete methylome of *Helicobacter pylori* UM032. *BMC Genomics* **16**, 424 (2015).
25. Estibariz, I. et al. *In vivo* genome and methylome adaptation of cag-negative *Helicobacter pylori* during experimental human infection. *mBio* **11**, e01803–20 (2020).
26. Gann, A. A., Campbell, A. J., Collins, J. F., Coulson, A. F. & Murray, N. E. Reassortment of DNA recognition domains and the evolution of new specificities. *Mol. Microbiol.* **1**, 13–22 (1987).
27. Dimitriu, T., Szczelkun, M. D. & Westra, E. R. Evolutionary ecology and interplay of prokaryotic innate and adaptive immune systems. *Curr. Biol.* **30**, R1189–R1202 (2020).
28. Furuta, Y. & Kobayashi, I. Movement of DNA sequence recognition domains between non-orthologous proteins. *Nucleic Acids Res* **40**, 9218–9232 (2012).
29. Koonin, E. V., Makarova, K. S. & Wolf, Y. I. Evolutionary genomics of defense systems in Archaea and Bacteria. *Annu Rev. Microbiol* **71**, 233–261 (2017).
30. Meng, B., Epp, N., Wijaya, W., Mrazek, J. & Hoover, T. R. Methylation motifs in promoter sequences may contribute to the maintenance of a conserved (m5)C methyltransferase in *Helicobacter pylori*. *Microorganisms* **9**, https://doi.org/10.3390/microorganisms9122474 (2021).
31. Yamaoka, Y. et al. Relationship between *Helicobacter pylori iceA*, *cagA*, and *vacA* status and clinical outcome: Studies in four different countries. *J. Clin. Microbiol.* **37**, 2274–2279 (1999).
32. Xu, Q. et al. Functional analysis of *iceA1*, a CATG-recognizing restriction endonuclease gene in *Helicobacter pylori*. *Nucleic Acids Res* **30**, 3839–3847 (2002).
33. Kita, K., Tsuda, J. & Nakai, S. Y. C.EcoO109I, a regulatory protein for production of EcoO109I restriction endonuclease, specifically binds to and bends DNA upstream of its translational start site. *Nucleic Acids Res.* **30**, 3558–3565 (2002).
34. Negri, A. et al. Regulator-dependent temporal dynamics of a restriction-modification system's gene expression upon entering new host cells: single-cell and population studies. *Nucleic Acids Res.* **49**, 3826–3840 (2021).
35. Srikhanta, Y. N. et al. Phasevarion mediated epigenetic gene regulation in *Helicobacter pylori*. *PLoS ONE* **6**, e27569 (2011).
36. Srikhanta, Y. N. et al. Methylomic and phenotypic analysis of the ModH5 phasevarion of *Helicobacter pylori*. *Sci. Rep.* **7**, 16140 (2017).
37. Bzymek, M. & Lovett, S. T. Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. *Proc. Natl Acad. Sci. USA* **98**, 8319–8325 (2001).
38. Kobayashi, I. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res.* **29**, 3742–3756 (2001).
39. Furuta, Y., Abe, K. & Kobayashi, I. Genome comparison and context analysis reveals putative mobile forms of restriction-modification systems and related rearrangements. *Nucleic Acids Res.* **38**, 2428–2443 (2010).
40. Baltrus, D. A. & Guillemin, K. Multiple phases of competence occur during the *Helicobacter pylori* growth cycle. *FEMS Microbiol. Lett.* **255**, 148–155 (2006).
41. Corbinais, C. et al. ComB proteins expression levels determine *Helicobacter pylori* competence capacity. *Sci. Rep.* **7**, 41495 (2017).
42. Morelli, G. et al. Microevolution of *Helicobacter pylori* during prolonged infection of single hosts and within families. *PLoS Genet.* **6**, e1001036 (2010).
43. Kennemann, L. et al. *Helicobacter pylori* genome evolution during human infection. *Proc. Natl Acad. Sci. USA* **108**, 5033–5038 (2011).
44. Didelot, X. et al. Genomic evolution and transmission of *Helicobacter pylori* in two South African families. *Proc. Natl Acad. Sci. USA* **110**, 13880–13885 (2013).
45. Rusinov, I. S., Ershova, A. S., Karyagina, A. S., Spirin, S. A. & Alexeevski, A. V. Comparison of methods of detection of exceptional sequences in prokaryotic genomes. *Biochem. (Mosc.)* **83**, 129–139 (2018).
46. Burge, C., Campbell, A. M. & Karlin, S. Over- and under-representation of short oligonucleotides in DNA sequences. *Proc. Natl Acad. Sci. USA* **89**, 1358–1362 (1992).
47. Falush, D. et al. Traces of human migrations in *Helicobacter pylori* populations. *Science* **299**, 1582–1585 (2003).
48. Olbermann, P. et al. A global overview of the genetic and functional diversity in the *Helicobacter pylori cag* pathogenicity island. *PLoS Genet.* **6**, e1001069 (2010).
49. Moodley, Y. et al. Age of the association between *Helicobacter pylori* and man. *PLoS Pathog.* **8**, e1002693 (2012).
50. Vale, F. F. & Vitor, J. M. Genomic methylation: a tool for typing *Helicobacter pylori* isolates. *Appl Environ. Microbiol.* **73**, 4243–4249 (2007).
51. Vale, F. F., Megraud, F. & Vitor, J. M. Geographic distribution of methyltransferases of *Helicobacter pylori*: evidence of human host population isolation and migration. *BMC Microbiol.* **9**, 193 (2009).
52. Karlin, S., Burge, C. & Campbell, A. M. Statistical analyses of counts and distributions of restriction sites in DNA sequences. *Nucleic Acids Res.* **20**, 1363–1370 (1992).
53. Gelfand, M. S. & Koonin, E. V. Avoidance of palindromic words in bacterial and archaeal genomes: a close connection with restriction enzymes. *Nucleic Acids Res.* **25**, 2430–2439 (1997).
54. Rocha, E. P. C., Danchin, A. & Viari, A. Evolutionary role of Restriction/Modification systems as revealed by comparative genome analysis. *Genome Res.* **11**, 946–958 (2001).
55. Figueiredo, C. et al. Genetic organization and heterogeneity of the *iceA* locus of *Helicobacter pylori*. *Gene* **246**, 59–68 (2000).
56. Shen, J. C., Rideout, W. M. 3rd & Jones, P. A. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res.* **22**, 972–976 (1994).
57. Naito, T., Kusano, K. & Kobayashi, I. Selfish behavior of Restriction-Modification systems. *Science* **267**, 897–899 (1995).
58. Murray, N. E. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol. Mol. Biol. Rev.* **64**, 412–434 (2000).
59. Furuta, Y., Kawai, M., Uchiyama, I. & Kobayashi, I. Domain movement within a gene: a novel evolutionary mechanism for protein diversification. *PLoS ONE* **6**, e18819 (2011).
60. Chao, M. C. et al. A cytosine methyltransferase modulates the cell envelope stress response in the cholera pathogen. *PLoS Genet.* **11**, e1005739 (2015).
61. Haakonsen, D. L., Yuan, A. H. & Laub, M. T. 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 (2015).
62. Kahramanoglou, C. et al. Genomics of DNA cytosine methylation in *Escherichia coli* reveals its role in stationary phase transcription. *Nat. Commun.* **3**, 886 (2012).
63. Callens, M., Pradier, L., Finnegan, M., Rose, C. & Bedhomme, S. Read between the lines: diversity of nontranslational selection pressures on local codon usage. *Genome Biol. Evol.* **13**, https://doi.org/10.1093/gbe/evab097 (2021).
64. Munoz, A. B., Stepanian, J., Trespalacios, A. A. & Vale, F. F. Bacteriophages of *Helicobacter pylori*. *Front. Microbiol.* **11**, 549084 (2020).
65. Vale, F. F. et al. Dormant phages of *Helicobacter pylori* reveal distinct populations in Europe. *Sci. Rep.* **5**, 14333 (2015).
66. Vale, F. F. et al. Genomic structure and insertion sites of *Helicobacter pylori* prophages from various geographical origins. *Sci. Rep.* **7**, 42471 (2017).
67. Balzarolo, M. et al. m6A methylation potentiates cytosolic dsDNA recognition in a sequence-specific manner. *Open Biol.* **11**, 210030 (2021).
68. Tsuchiya, H., Matsuda, T., Harashima, H. & Kamiya, H. Cytokine induction by a bacterial DNA-specific modified base. *Biochem. Biophys. Res. Commun.* **326**, 777–781 (2005).
69. Thorpe, H. A. et al. Repeated out-of-Africa expansions of *Helicobacter pylori* driven by replacement of deleterious mutations. *Nat. Commun.* **13**, 6842 (2022).
70. Thorell, K. et al. Rapid evolution of distinct *Helicobacter pylori* subpopulations in the Americas. *PLoS Genet.* **13**, e1006546 (2017).
71. Munoz-Ramirez, Z. Y. et al. A 500-year tale of co-evolution, adaptation, and virulence: *Helicobacter pylori* in the Americas. *ISME J.* **15**, 78–92 (2021).
72. Higashi, H. et al. Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the Tyrosine phosphorylation sites. *Proc. Natl Acad. Sci. USA* **99**, 14428–14433 (2002).
73. Naito, M. et al. Influence of EPIYA-repeat polymorphism on the phosphorylation-dependent biological activity of *Helicobacter pylori* CagA. *Gastroenterology* **130**, 1181–1190 (2006).
74. Hayashi, T. et al. Differential mechanisms for SHP2 binding and activation are exploited by geographically distinct *Helicobacter pylori* CagA oncoproteins. *Cell Rep.* **20**, 2876–2890 (2017).
75. Zhou, Z. et al. The EnteroBase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity. *Genome Res.* **30**, 138–152 (2020).
76. Nell, S. et al. Recent acquisition of *Helicobacter pylori* by Baka pygmies. *PLoS Genet.* **9**, e1003775 (2013).
77. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
78. Moodley, Y. et al. *Helicobacter pylori*'s historical journey through Siberia and the Americas. *Proc Natl Acad Sci USA* **118**, https://doi.org/10.1073/pnas.2015523118 (2021).
79. Achtman, M. et al. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol. Microbiol.* **32**, 459–470 (1999).
80. Roberts, R. J., Vincze, T., Posfai, J. & Macelis, D. REBASE–a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* **43**, D298–D299 (2015).

81. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
82. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
83. Pevzner, P. A., Borodovsky, M. & Mironov, A. A. Linguistics of nucleotide sequences. I: The significance of deviations from mean statistical characteristics and prediction of the frequencies of occurrence of words. *J. Biomol. Struct. Dyn.* **6**, 1013–1026 (1989).
84. Page, A. J. et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).
85. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
86. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, https://doi.org/10.1093/gigascience/giab008 (2021).
87. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
88. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33**, 1–22 (2010).
89. Minh, B. Q. et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
90. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
91. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
92. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2011).

Acknowledgements

We thank Iratxe Estibariz for the early discussions about the variability of methylation patterns and Christine Josenhans for constructive comments on the manuscript. Funding was provided from the Deutsche Forschungsgemeinschaft (project number 158989968-grants SFB900/A1 and SFB900/Z1 to S.S.), by the Bavarian Ministry of Science and the Arts through project HelicoPredict in the framework of the research network bayresq.net, and the German Center for Infection Research (DZIF grants 06.824 and 06.709).

Author contributions

F.A. and S.S. designed the study. F.A. and W.G. performed the analysis. F.A., W.G., and S.S. interpreted the results. F.A. and S.S. wrote the manuscript. F.A., W.G., and S.S. revised the manuscript. All authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

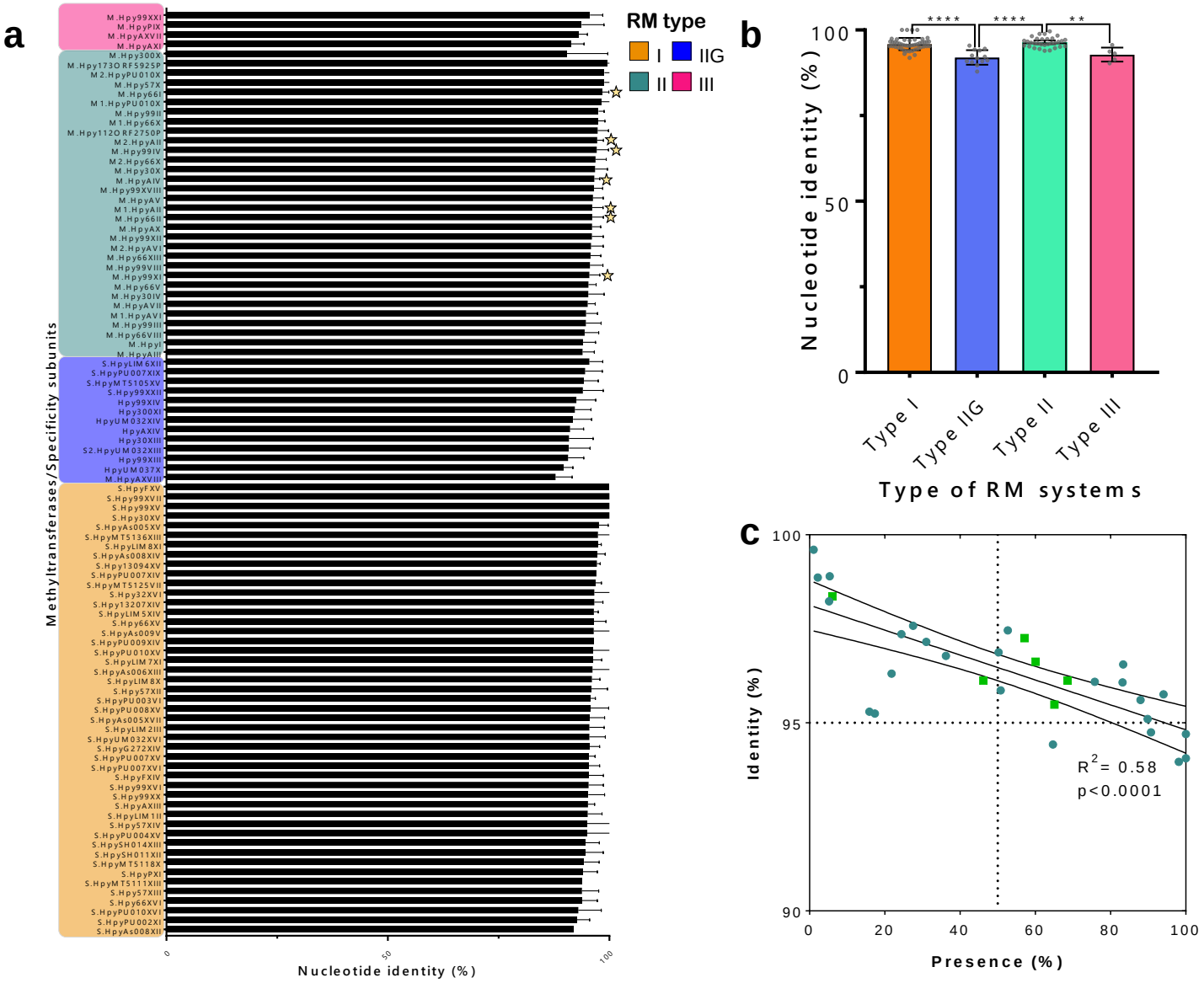
The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-023-05218-x.

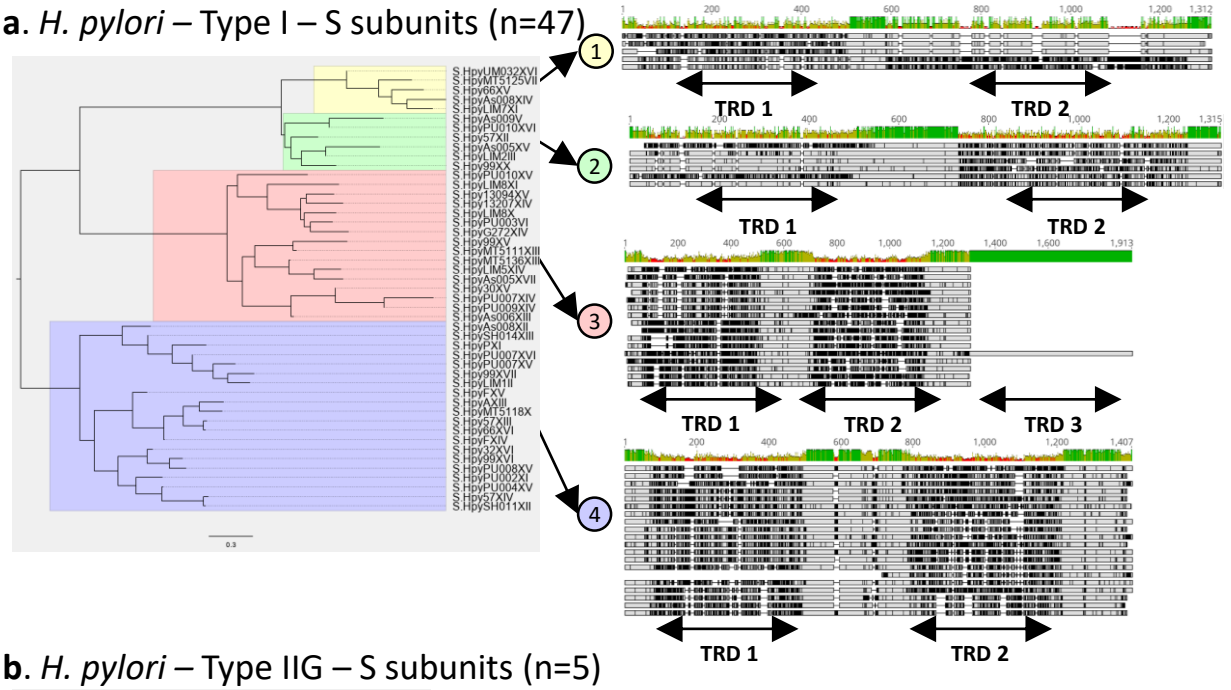
Correspondence and requests for materials should be addressed to Florent Ailloud or Sebastian Suerbaum.

Peer review information *Communications Biology* thanks Maria Antonia Sanchez-Romero, Andreas E. Zautner, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary handling editors: Tobias Goris and George Inglis. A peer review file is available.

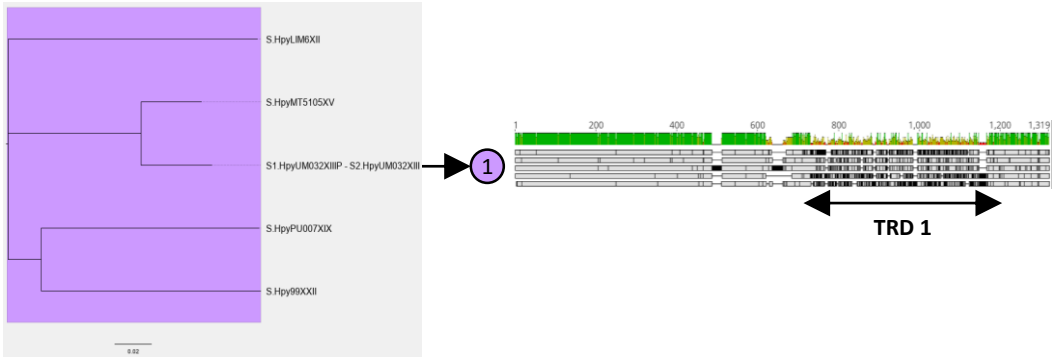


Supplementary Figure 1 – Sequence conservation of methyltransferases in *H. pylori*. **a**. Nucleotide diversity of 96 methyltransferases and specificity subunits. Type II methyltransferases flanked by direct repeats are indicated by a star. Error bars represent 95% confidence intervals. **b**. Average identity in each type (Dunn's test two-sided **** $p < 0.0001$, ** $p < 0.01$). Error bars represent 95% confidence intervals. **c**. Scatter plot of the frequency and nucleotide identity of type II methyltransferases. Methyltransferases flanked by direct repeats are indicated by a square symbol. The correlation coefficient obtained by linear regression is indicated.

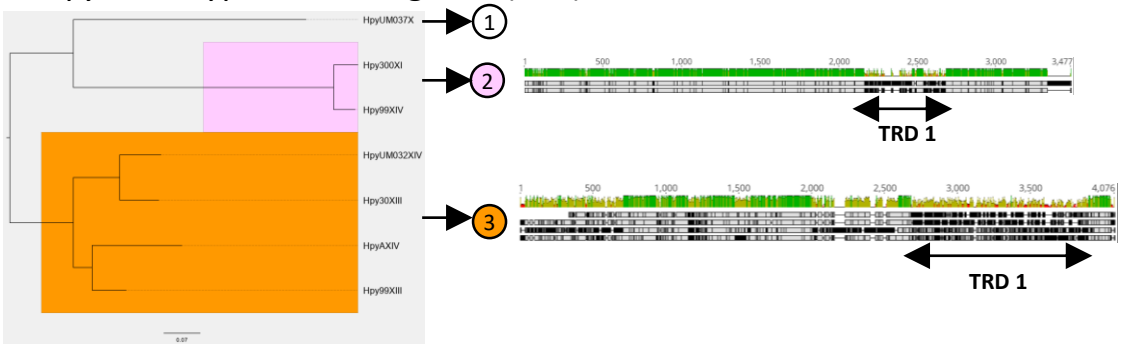
a. *H. pylori* – Type I – S subunits (n=47)



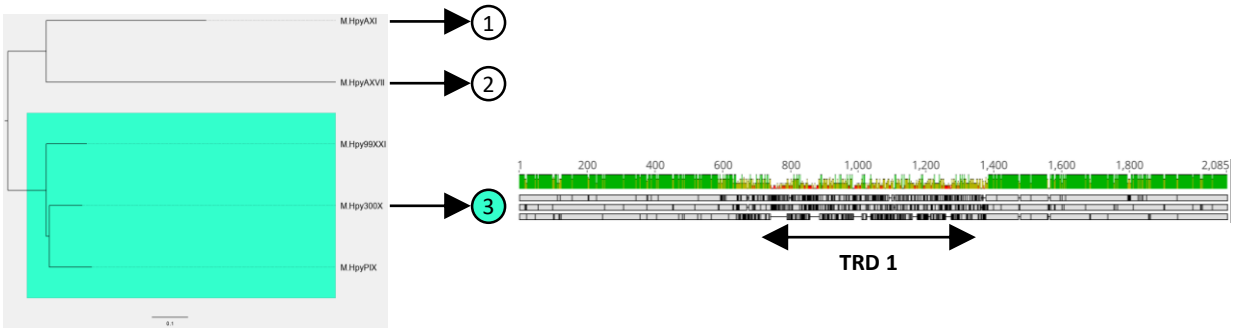
b. *H. pylori* – Type IIG – S subunits (n=5)



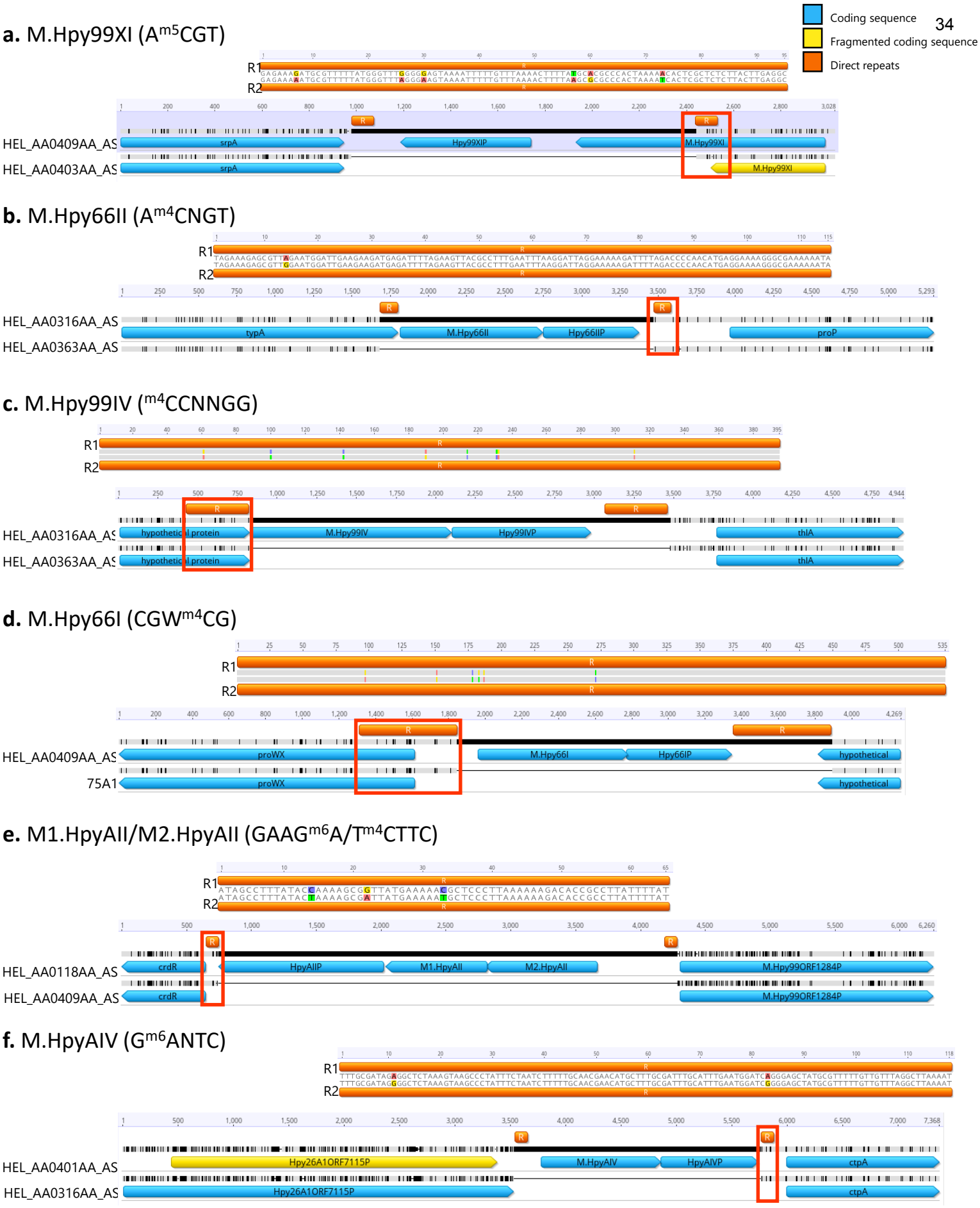
c. *H. pylori* – Type IIG – RM genes (n=7)



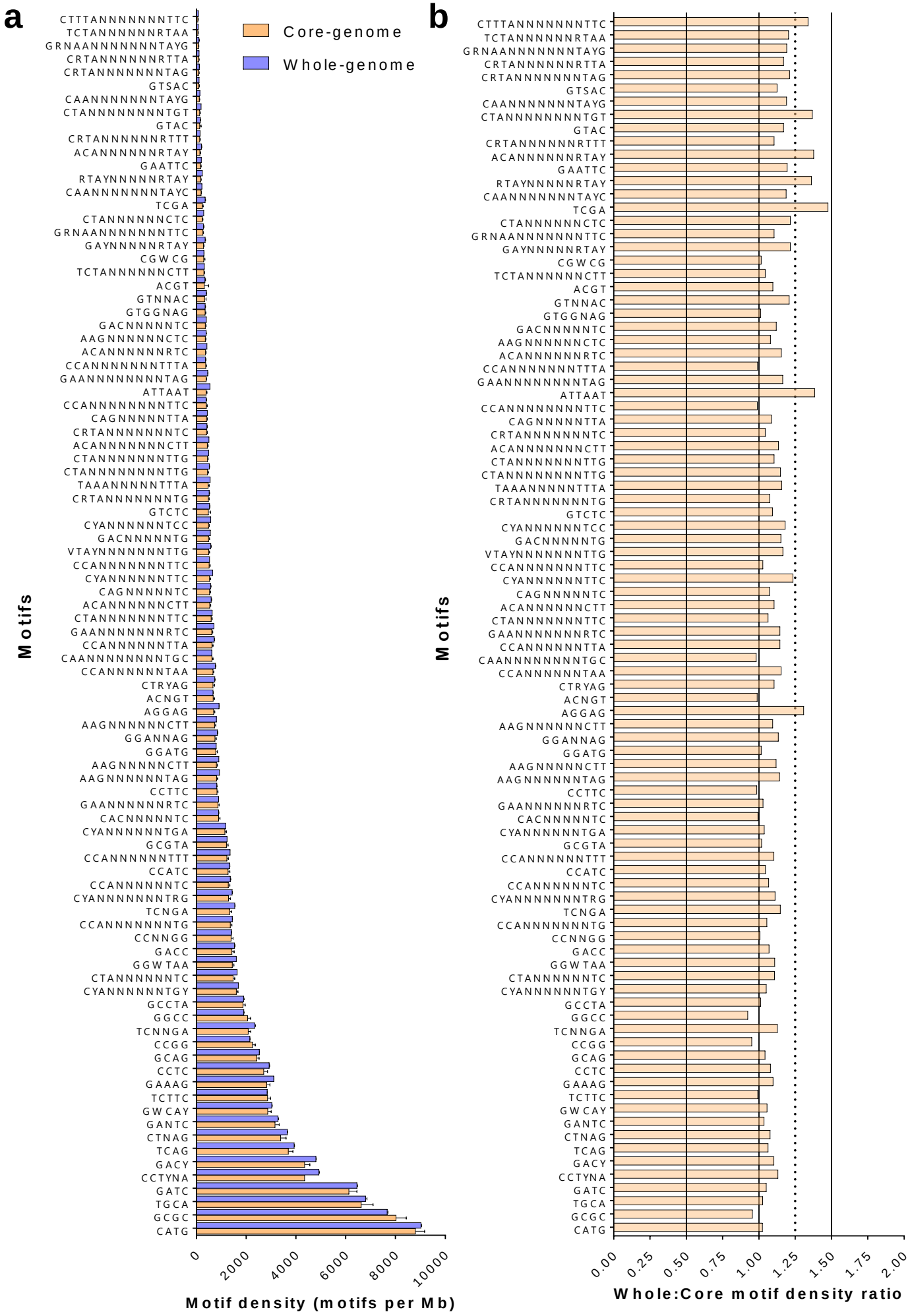
d. *H. pylori* – Type III – M genes (n=5)



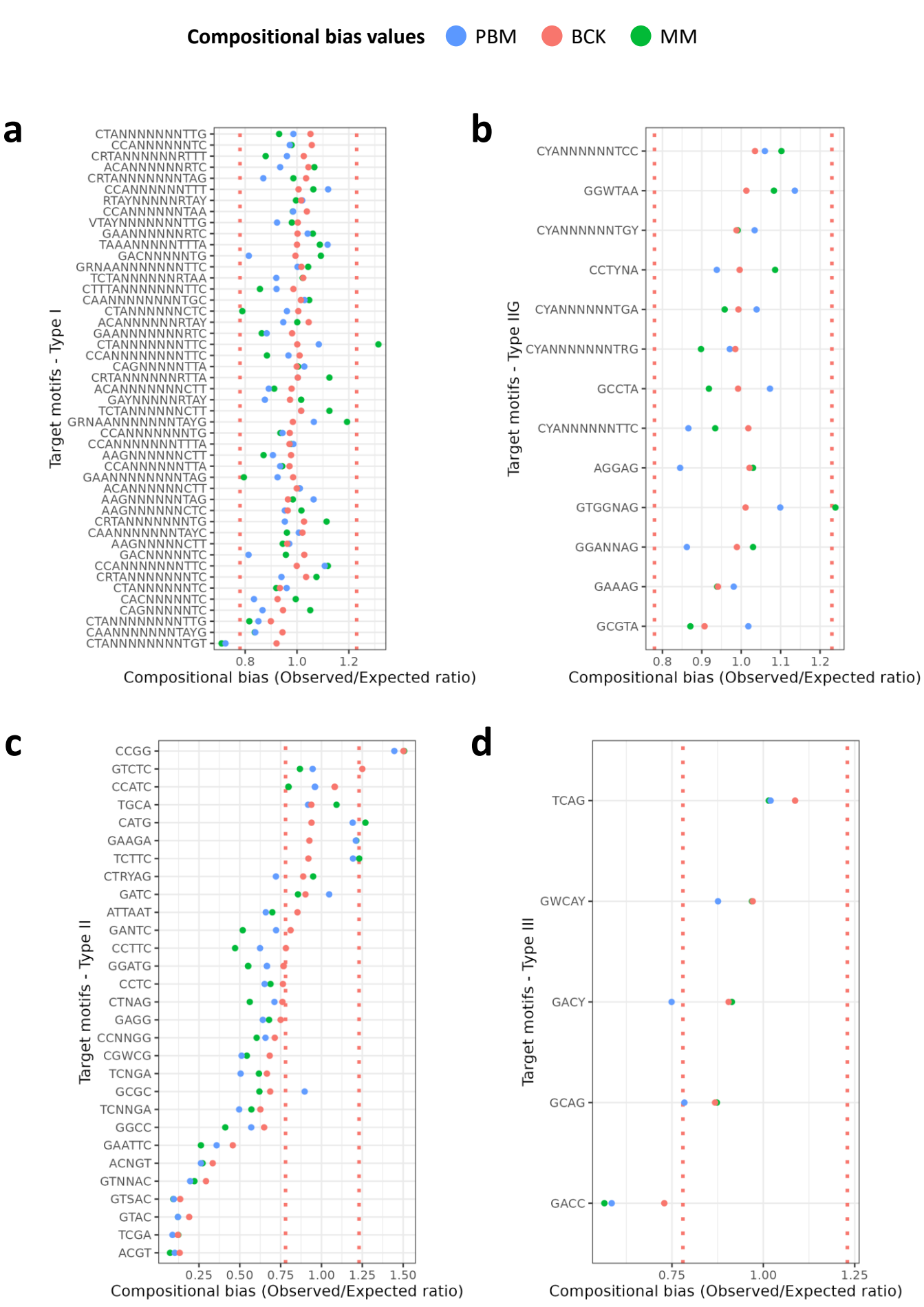
Supplementary Figure 2 – Phylogeny of type I, IIG, and III RM systems in *H. pylori*. Clusters of methyltransferases determined by phylogeny and confirmed by inspection of multiple alignments are indicated by colour boxes. For each cluster, a visual representation of the multiple alignment is provided (black lines indicate positions which are different between strains). An identity diagram is located on top of each alignment and shows conserved (in green) and highly variable (in red) positions. On the bottom of the alignment, the position of target-recognition domains is indicated by horizontal arrows. The diversity of target-recognition domains within a conserved methyltransferase backbone can be appreciated by remarking the conserved positions of the backbone interspersed by clusters of variable positions from the domains. **a**. Phylogeny of S-subunits from type I RM systems. **b**. Phylogeny of S subunits from type IIG RM systems. **c**. Phylogeny of RM genes from type IIG RM systems. **d**. Phylogeny of methyltransferases from type III RM systems.



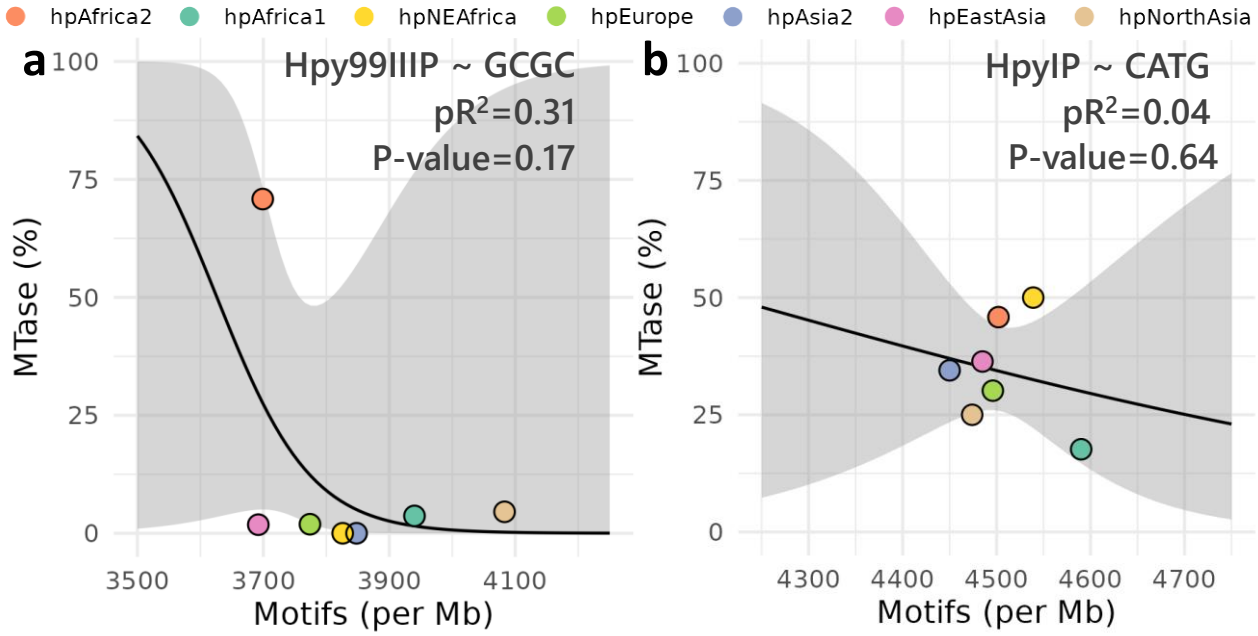
Supplementary Figure 3 – Direct repeats flanking six different type II RM systems. The genetic context in strains in which the given type II RM system is either present (top) or absent (bottom) is shown as a pairwise alignment. Complete and fragment coding sequences are coloured in blue and yellow, respectively. The direct flanking repeats are indicated by orange boxes and the repeat remaining in both strains is highlighted by a red box. An alignment of the left and right flanking repeats is also provided to visualized the similarity between them. A. M.Hpy99XI (A^{m5}CGT). B. M.Hpy66II (A^{m4}CNGT). C. M.Hpy99IV (^{m4}CCNNGG). D. M.Hpy66I (CGW^{m4}CG). E. M1.HpyAII/M2.HpyAII (GAAG^{m6}A/T^{m4}CTTC). F. M.HpyAIV (G^{m6}ANTC).



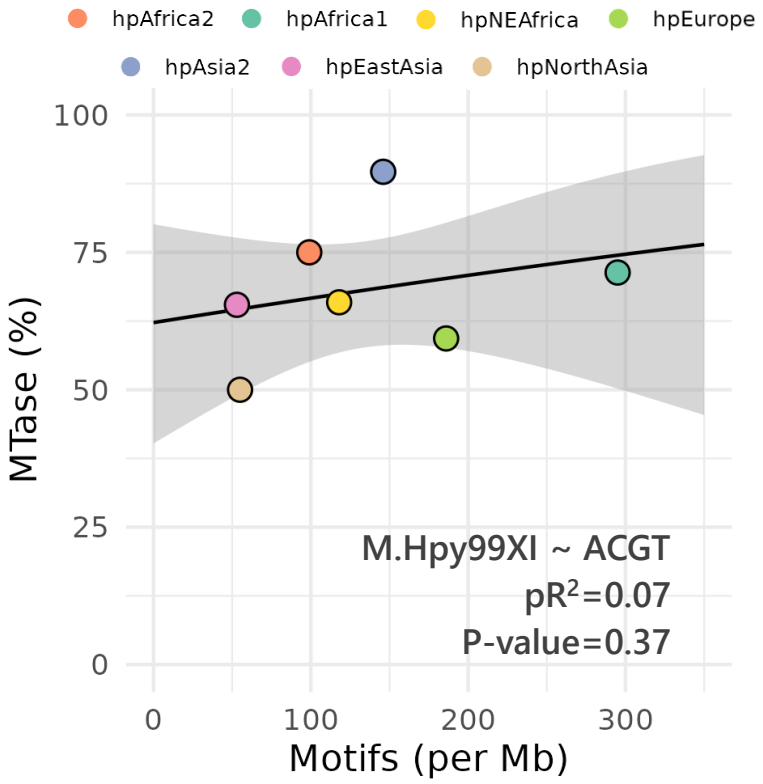
Supplementary Figure 4 – Comparison of target motif density in whole genome and core genes. **a**. Motif density (motifs/Mb) of 92 target sequences calculated in a core gene alignment. Error bars represent 95% confidence intervals. **b**. Ratio of motif density calculated on whole genomes versus core genes.



Supplementary Figure 5 – Compositional bias calculation for target motifs in *H. pylori* using three different methods. **a**. Type I motifs. **b**. Type IIG motifs. **c**. Type II motifs. **d**. Type III motifs. MM: method based on maximum order Markov chain. PBM: method based on Pevzner and co-authors. BCK: method based on Burge and co-authors. Under- and over- represented cutoffs are represented by vertical lines.



Supplementary Figure 6 – Interaction between endonuclease frequency and motif density in two type II RM systems. **a**. Hpy99IIIP **b**. HpyIP. Pseudo R^2 (calculated with the Mc Fadden method) and p-values are indicated. 95% confidence interval is indicated by a grey ribbon.



Supplementary Figure 7 – Interaction between methyltransferase frequency and motif density for the Hpy99XI RM system. Pseudo R^2 (calculated with the Mc Fadden method) and p-values are indicated. 95% confidence interval is indicated by a grey ribbon.

AMERICAN SOCIETY FOR MICROBIOLOGY | mBio

Genetics and Molecular Biology | Research Article

The *Helicobacter pylori* orphan ATTAAT-specific methyltransferase M.Hpy99XIX plays a central role in the coordinated regulation of genes involved in iron metabolism

Wilhelm Gottschall,[1] Florent Ailloud,[1,2] Christine Josenhans,[1,2] Sebastian Suerbaum[1,2]

AUTHOR AFFILIATIONS See affiliation list on p. 18.

ABSTRACT *Helicobacter pylori* genomes contain a large and variable portfolio of methyltransferases (MTases), creating a highly diverse methylome. Here, we characterize a highly conserved ATTAAT-specific MTase, M.Hpy99XIX (*H. pylori* strain J99, alternative designations in other strains, M.HpyAVII and M.HpyPVII), the only *H. pylori* MTase never associated with an endonuclease ("orphan" MTase). Inactivation of M.Hpy99XIX resulted in a significant change in the transcription of >100 genes, despite the fact that only a small subset of their promoter regions contained an ATTAAT target motif. Patterns of transcriptional change showed significant correlations with changes reported for *H. pylori* mutants in the regulators involved in iron regulation. MTase inactivation also caused a higher susceptibility to diverse metal ions as well as iron chelation and oxidative stress. These phenotypes could be traced back to the methylation of single motifs in the promoter regions of iron transporters *frpB1* and *fecA1*. Altogether, methylation of individual motifs in promoters can have a large downstream effect, causing major changes to metabolic pathways. These findings suggest that the methylome represents a universal and dynamic interface connecting genome diversity and transcriptional regulation. Very recently, a new ecospecies of *H. pylori*, Hardy, has been reported. M.Hpy99XIX is present in the majority of "normal" (Ubiquitous) *H. pylori* strains, whereas no single Hardy strain contained this gene, consistent with other reported differences between Hardy and Ubiquitous strains related to iron/metal homeostasis. ATTAAT methylation is intricately connected with the bacterial transcriptional network, highlighting the important role of bacterial epigenetic modifications in bacterial physiology and pathogenesis.

IMPORTANCE *Helicobacter pylori* has one of the largest repertoires of methyltransferases. Methylation has been associated with multiple functions in *H. pylori*, including the defense against foreign DNA and transcriptional regulation. Regulation of gene expression by methylation has the potential to influence many distant genes across the genome via target motifs in proximity to transcription start sites. Here, we sought to understand the role of M.Hpy99XIX, an orphan methyltransferase targeting the ATTAAT motif that is highly conserved in *H. pylori*. We show that by directly regulating specific genes involved in iron uptake via methylated ATTAAT motifs, M.Hpy99XIX has a significant effect on iron homeostasis by triggering the canonical iron regulatory pathway. Furthermore, we show that M.Hpy99XIX appears to have been acquired after the split between the two ecospecies of *H. pylori*, suggesting that its role in the tuning of iron homeostasis might have contributed to this divergence.

KEYWORDS *Helicobacter pylori*, DNA methylation, transcriptome, bacterial epigenetics, promoters, genomics

Editor Carmen Buchrieser, Institut Pasteur, Paris, France

Address correspondence to Sebastian Suerbaum, suerbaum@mvp.lmu.de.

Wilhelm Gottschall and Florent Ailloud contributed equally to this article. The co-first authors mutually agreed upon the order in which they are listed in the byline, and the last author concurred with their decision.

The authors declare no conflict of interest.

See the funding table on p. 19.

Received 25 April 2025
Accepted 8 May 2025
Published 9 June 2025

Helicobacter pylori is a genetically highly diverse bacterial species that can colonize the stomach mucosa of humans for many decades and can lead to severe diseases such as peptic ulcers, gastric adenocarcinoma, and lymphoma of the mucosa-associated lymphoid tissue (1–3). *H. pylori* has a large repertoire of restriction-modification (RM) systems (4–7). While the pangenome of *H. pylori* comprises more than 100 methyltransferases (MTases), only a minority of these is highly conserved across the diverse phylogeographical populations within the species (8). Each strain carries a specific repertoire of RM systems, acting on a unique pattern of target motifs in the genome, resulting in virtually unique methylomes (8, 9). In the last decade, research on DNA methylation and its role in bacterial physiology has been accelerated by the development of novel technologies which permit the detection of methylation at single-base resolution (in particular, single-molecule real-time sequencing [10] and, more recently, Oxford Nanopore sequencing [11]). For example, methylation of the motif GCGC by the MTase M.Hpy99III, one of the two MTases present in all *H. pylori* strains, is associated with multiple important phenotypes, such as adhesion to gastric cells, natural competence, bacterial cell shape, or copper tolerance (12). Similarly, three other, less conserved MTases have also been connected to important traits such as motility and tolerance to oxidative stress (13).

DNA methylation has been shown to play an important role in gene regulation in many bacteria (14). For example, the *dam* methyltransferase regulates the phase variation of the pyelonephritis-associated pili in *Escherichia coli* (15), and the transcription of TraJ, a transcriptional activator of the transfer operon in *Salmonella enterica* (16). The phenotypic effects connected to methylation in *H. pylori*, in some cases, have been shown to be mediated by changes in gene expression (12, 13, 17, 18). However, the connection between methylation and transcription appears to be highly complex, and significant variation was observed among different MTases and between strains of *H. pylori* (12, 13). For the M.Hpy99III MTase, transcriptional regulation is based on the methylation of GCGC motifs within promoter regions (12). This has been demonstrated for four distinct genes using targeted mutagenesis and reporter assays. Because genes carrying a target motif within their promoter generally account only for a small proportion of genes regulated by MTases (12, 13, 17, 18), additional regulatory mechanisms and large downstream transcriptional effects are likely at play.

In bacterial genomes, MTase genes are frequently coupled with genes coding for a cognate endonuclease, recognizing the same sequence motif as the MTase, constituting RM systems. These are thought to be primarily involved in defense against phages and invading foreign DNA (19, 20). In type II RM systems, the endonuclease is not essential for the activity of the MTase, and therefore, type II MTase genes can persist independently as so-called "orphan" genes (21). Since selection pressures acting on orphan MTases are not influenced by those acting on a cognate endonuclease anymore, orphan MTases constitute an ideal model to better understand the mechanisms connecting methylation, transcription, and phenotypic effects.

Here, we determined that M.Hpy99XIX is the only strictly orphan MTase in *H. pylori* and that its target motif, ATTAAT (9), is enriched in promoter regions. M.Hpy99XIX is encoded by gene *jhp0430* in *H. pylori* strain J99 and has previously been studied in other strain contexts under alternative names M.HpyAVII and M.HpyPVII (9, 13). In this study, we found a large group of genes to be differentially regulated after inactivation of ATTAAT methylation, disrupting the iron homeostasis regulatory pathway on the transcriptomic and phenotypic levels. These phenotypic effects could be recapitulated by the targeted modification of a single ATTAAT motif located within the promoter region of an iron acquisition gene, demonstrating how DNA methylation can have large effects on transcriptional networks and core bacterial processes by modulating only a very limited number of genes directly. On a larger scale, this study substantiates the role of the methylome as an additional source of diversity in *H. pylori*, interconnected with the genome through individual sequence motifs, and with the transcriptome via methylation-based regulation.

RESULTS

M.Hpy99XIX is the only strictly orphan MTase in *H. pylori*

Methyltransferases without a coupled endonuclease, frequently referred to as orphan MTases, cannot fulfill their presumed original function as part of RM systems, such as defense against invading DNA. In bacteria, orphan MTases have been frequently associated with transcriptional regulation (14). To determine the proportion of *H. pylori* MTases that are not part of canonical RM systems, we analyzed a collection of ~400 geographically diverse genomes of *H. pylori* that was the basis of our previous characterization of the methylome diversity in *H. pylori* (8). In particular, we determined the frequency of all known *H. pylori* type II MTase and cognate endonuclease genes in this genome collection (22). When no cognate endonuclease could be found, the genes neighboring the MTase locus were examined to confirm that they were not homologs of endonucleases.

Remarkably, we could only identify a single MTase gene, M.Hpy99XIX (9), that was never associated with an endonuclease across the whole species (Fig. 1A). The M.Hpy99XIX MTase is highly conserved and predicted to be active in ~90% of strains in the collection. We also analyzed a collection of 1,012 finished *H. pylori* genomes, combining the recently released HpGp dataset and other *H. pylori* reference sequences (see Methods) (23). In this collection, the M.Hpy99XIX MTase was predicted to be functional in ~94% of the isolates and found exclusively at a single locus, between the genes *hopJ* (*jhp0429*), coding for an outer membrane protein, and *hepT* (*jhp0431*), encoding a DD-heptosyltransferase (Fig. 1B). Subsequent analysis showed that the MTase gene was fragmented in 43 of the 1,012 strains (4.3%) due to nonsense mutations, which likely render it inactive. In a smaller subset of only 10 strains (1%), the MTase gene was completely absent with no apparent remnants or excision site. While the strains affected by nonsense mutations were not obviously genetically related, the strains lacking the M.Hpy99XIX gene all belonged to a single *H. pylori* subpopulation, hspIndigenousAmerica.

The ATTAAT motif is the only MTase target sequence enriched in promoter regions in *H. pylori*

The M.Hpy99XIX MTase targets the palindromic motif ATTAAT and catalyzes the N6-methyladenosine modification at the fifth position (9). In *H. pylori* and other bacterial species, methylation of single motifs located within promoter regions has been associated with regulation of the downstream transcriptional unit (12, 24). Compared to some other motifs modified by type II enzymes, the ATTAAT motif is relatively infrequent in the *H. pylori* genome with an average of 553 motifs per megabase (8) and thus is not likely to be randomly observed near promoter regions. Consequently, we determined how often this motif appears in promoter regions and thus could potentially directly act on downstream gene expression.

We calculated ratios (as fold change [FC]) between the relative motif frequencies within promoters and the corresponding motif frequencies in the complete genome, for the ATTAAT motif and selected other known type II motifs (Fig. 1C). Consequently, this analysis describes whether a given motif is either overrepresented ($\log_2$FC > 0) or underrepresented ($\log_2$FC < 0) within promoter regions. Almost all type II motifs were significantly underrepresented in promoter regions, whereas the ATTAAT motif was significantly enriched with a positive $\log_2$FC of 0.98, signifying that it was observed almost twice as frequently in promoter regions compared to the rest of the genome. On average, the ATTAAT motif was observed in 37 ± 5 promoter regions per strain, representing approximately 5% of the promoter regions in *H. pylori* (Table S1).

Regulatory motifs are often located in fixed positions relative to the transcription start sites. Based on all promoters carrying ATTAAT motifs, we determined the probability distribution function of the motif position and of the Pribnow box (TATAAT) typically

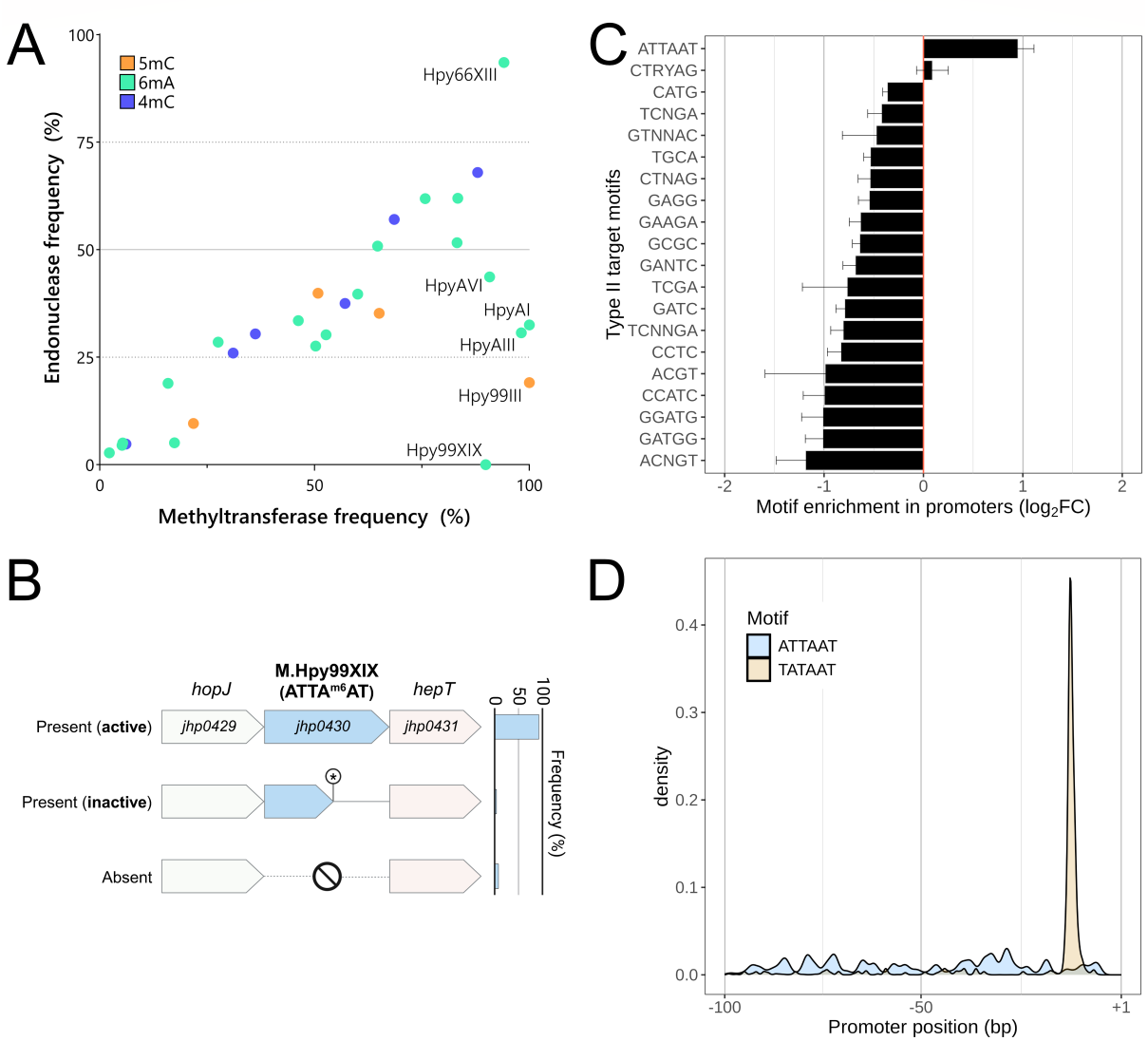


FIG 1 Genetic characterization of the *H. pylori* M.Hpy99XIX methyltransferase and its target motif ATTAAT. (A) Frequencies of 26 type II *H. pylori* methyltransferase (*x*-axis) and cognate endonuclease (*y*-axis) genes in a collection of 398 geographically diverse *H. pylori* genomes. Frequencies were normalized across four geographical populations of *H. pylori* (hpAfrica1, hpAfrica2, hpEurope, and hpAsia2). (B) Genetic context and frequency of the complete, inactivated, and empty locus in a distinct collection of 1,012 *H. pylori* genomes. The complete locus is by far the most frequent and is always flanked by the *hopJ* (*jhp0429*) and *hepT* (*jhp0431*) genes. (C) Enrichment analysis of type II motifs performed by calculating the ratio of motif frequencies in promoter regions and in the whole *H. pylori* genome. Log_2 fold changes (FCs) >1 and <1 indicate motifs more and less frequent in promoters compared to the rest of the genome, respectively. Error bars indicate the standard deviations across all *H. pylori* genomes analyzed. The first 20 motifs with the highest absolute log_2 fold changes are shown. (D) Positional bias analysis of the ATTAAT target motif (blue) and the TATA box (orange) within promoter regions. The ATTAAT motif does not appear at any preferred position.

located at −10 as a reference (25). In contrast to TATAAT, no specific positions were favored by the ATTAAT motif within the promoter region (Fig. 1D).

ATTAAT methylation by M.Hpy99XIX has a large effect on the *H. pylori* transcriptome

In order to determine whether M.Hpy99XIX and the methylation of its target motif ATTAAT have a role in gene expression, we compared the transcriptomes of wild-type (WT) *H. pylori* strain J99 (26), an isogenic M.Hpy99XIX knockout (KO) mutant, and a functionally complemented strain using RNA-Seq analysis (Table S2). Differential gene expression analysis revealed that 85 genes were downregulated in the ATTAAT methylation-deficient strain compared to the wild type (log_2 fold change < 1 and false discovery

rate [FDR] adjusted *P* value < 0.01), whereas 28 genes were upregulated (log_2 fold change > 1 and FDR adjusted *P* value < 0.01) (Fig. 2A), altogether representing ~8% of the total genes in the *H. pylori* J99 genome. The functional complementation (Fig. S1) restored the expression of 81.2% of downregulated and of 96.6% of upregulated genes to their wild-type expression levels. Ten and five new genes were down- and upregulated in the complemented strain compared to the wild type, respectively, possibly owing to the increased expression of the MTase in the complemented strain, which is a consequence of the insertion of the intact MTase gene in the highly transcribed urease gene cluster (Fig. S2).

Next, we asked whether the differential gene expression resulting from the absence of ATTA^{m6}AT methylation could be correlated with the position and local frequency of ATTAAT motifs across the genome. We calculated a motif enrichment score for ATTAAT in coding sequences (CDSs) (see Materials and Methods). An average of 0.3 ± 0.6 motifs was observed per CDS, and 77 CDSs showed significant enrichment. Nevertheless, no correlation was found between enrichment score and gene expression (Fig. S3; Table S1). Next, we turned to motifs residing in promoter regions. ATTAAT motifs could be detected in the promoter regions of 8% and 3% of up- and downregulated genes, respectively (Table S1). While no significant association was found (*P* = 0.068), downregulated genes that were linked to an ATTAAT motif in a promoter region leaned toward lower fold changes than other downregulated genes (Fig. 2B). No similar trend was observed for upregulated genes (*P* = 0.908). Furthermore, ATTAAT motifs located in promoter regions of differentially expressed genes were not located at a specific position relative to the transcriptional start site (TSS), yet five out of these eight motifs were found in the range of 67 to 80 bp upstream of the TSS (Fig. S4). Altogether, the data suggest that the majority of genes in our transcriptomic data set are not directly controlled by ATTA^{m6}AT methylation.

ATTA^{m6}AT-based regulation affects core transcriptional regulators

The regulatory pathways of *H. pylori* have been extensively characterized, and core transcriptional regulators have been described (27). Therefore, we asked whether

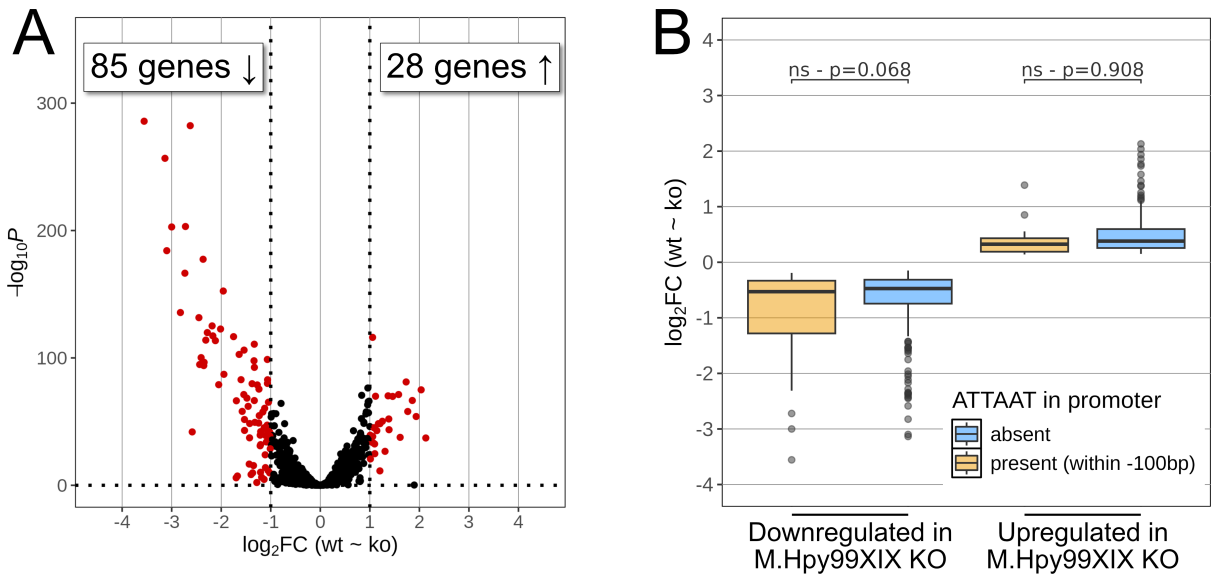


FIG 2 Effects of M.Hpy99XIX on the transcriptome of *H. pylori* strain J99. (A) Volcano plot of *H. pylori* J99 wild type compared to *H. pylori* J99 M.Hpy99XIX knockout strain. Negative and positive log_2 fold changes indicate a gene downregulated (left) or upregulated (right) in the knockout strain compared to the wild type, respectively (see also Table S2). Genes with an adjusted *P* value of <0.01 and an absolute log_2 fold change of >1 are shown in red. (B) The log_2 fold changes of downregulated (left) and upregulated (right) genes were compared (Student's *t*-test) according to the presence (orange) or absence (blue) of an ATTAAT motif within 100 bp of their transcription start sites (promoter regions).

significant overlaps existed between known *H. pylori* regulons and genes regulated through ATTA^{m6}AT methylation. Interestingly, we were able to find a significant association between genes regulated by ATTAAT methylation and four (ArsRS, HspR, acid, and iron responses [28–30]) out of the eight previously reported regulon data sets that we analyzed for correlations with our data set. There was no significant correlation with data sets for the transcriptional regulators HrcA, NikR, HP1021, and CrdRS (30–33) (Table S1).

Transcriptomic regulation connected to ATTA^{m6}AT methylation for many genes mirrored the changes associated with the ArsRS two-component system in *H. pylori* strain 26695 (Fig. S5A) (28) and showed inverse regulation to the ones that have been linked to the HspR regulator in *H. pylori* strain G27 (Fig. S5B) (30). Sixty-six percent (21 out of 32) of the genes downregulated in the ArsRS KO were also significantly downregulated in the M.Hpy99XIX KO strain (FDR adjusted *P* value <0.01 and $\log_2$FC < 0), while 39% (27 out of 70) of the genes upregulated in the ArsRS KO were significantly upregulated in the M.Hpy99XIX KO strain (FDR adjusted *P* value < 0.01 and $\log_2$FC > 0). By contrast, 80% (8 out of 10) of the genes downregulated in the HspR KO were upregulated in the M.Hpy99XIX KO strain, while 80% (8 out of 10) of the genes upregulated in the HspR KO were downregulated in the M.Hpy99XIX KO strain. When considering genes with at least a 50% difference in expression (equivalent to an absolute $\log_2$FC = 0.58 in the M.Hpy99XIX KO strain), a moderate correlation could be observed between the fold changes of differentially expressed genes in the M.Hpy99XIX KO strain with the ArsRS regulon (ρ = 0.59, *P* < 0.001) but not with the HspR regulon (ρ = 0.48, *P* = 0.1) (Fig. S6). For instance, the ferric uptake regulator (Fur, *jhp0397*) was downregulated both in the absence of ATTA^{m6}AT methylation and by ArsRS (28). In contrast, the chaperone and stress-response proteins, GroES (*jhp0009*) and DnaK (*jhp0101*), were upregulated in the M.Hpy99XIX KO strain, whereas they were downregulated by HspR (Fig. S5B).

Furthermore, ATTA^{m6}AT-based regulation also matched transcriptional changes associated with regulatory responses to low pH, which are largely dependent on ArsRS (Fig. 3A) (28) and extracellular iron content (Fig. 3B) (29). Specifically, 59% (10 out of 17) and 55% (17 out of 31) of genes induced and repressed at low pH were up- and downregulated in the M.Hpy99XIX KO strain, respectively, while 66% (50 out of 76) and 59% (29 out of 49) of genes induced and repressed in response to a high iron concentration were up- and downregulated in the M.Hpy99XIX KO strain, respectively. Moreover, strong correlations were observed between the fold changes in the M.Hpy99XIX KO strain with the acid-responsive regulon (ρ = 0.85, *P* < 0.001) and the iron-responsive regulon (ρ = 0.70, *P* < 0.001) (Fig. S6). To determine whether the association between M.Hpy99XIX and iron-based transcriptomic response is connected to Fur, we also compared ATTA^{m6}AT-based regulation with the Fur-dependent transcriptome (29) and observed that 71% (22 out of 31) of Fur-dependent iron-induced genes were up-regulated in the M.Hpy99XIX KO strain, while 63% (39 out of 62) of Fur-dependent iron-repressed genes were downregulated in the M.Hpy99XIX KO strain (Fig. 3B; Fig. S5C).

We note that strains, as well as exact conditions (e.g., with respect to iron), in the previous studies vary and differ from the ones used in this study. Regulons were previously studied by comparing the respective mutants with the wild-type bacteria. For the acid-responsive regulon, the medium conditions were compared between neutral pH (pH = 7) and low pH (pH = 5.3). For the iron-responsive regulon, a high iron [1 mM $(NH4)_2Fe(SO_4)_2$] condition was compared to an iron depletion (150 µM 2,2'-dipyridyl) condition (see Table S3 for further details).

These results highlight the interconnectedness of regulatory networks in *H. pylori* and suggest that the ATTAAT transcriptional response may play a role in fine-tuning specific regulatory pathways within *H. pylori*.

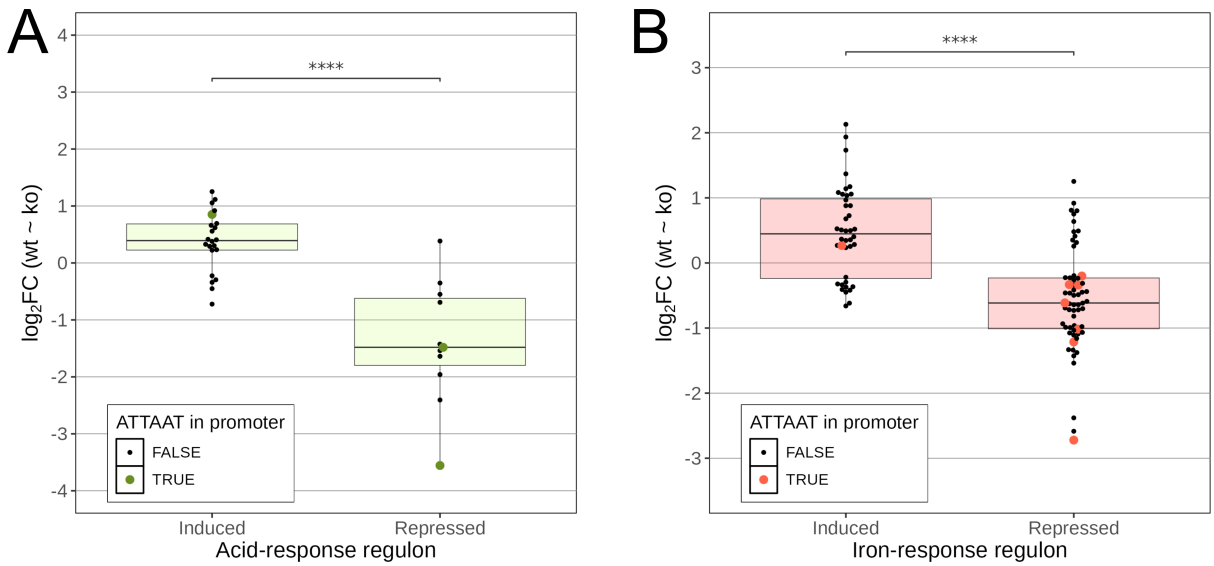


FIG 3 Comparison between transcriptional effects after inactivation of M.Hpy99XIX and environment-responsive regulons. (A) Comparison of RNA-Seq expression data of M.Hpy99XIX KO against WT with the acid-responsive regulon (28) split into induced and repressed genes. Student's *t*-test: ****$P < 0.00001$. (B) Comparison of RNA-Seq expression data of M.Hpy99XIX KO against WT with the iron-responsive regulon (29) split into induced and repressed genes. Student's *t*-test: ****$P < 0.00001$. Log_2 fold changes of significantly differentially expressed genes (adjusted *P* value < 0.01) are represented by individual points and by a box plot. The presence of an ATTAAT motif within 100 bp of the TSS of each gene is indicated by a larger colored point.

Transcriptional induction of iron transporter genes by ATTAAT methylation in promoters

To further understand the connection between the ATTAAT methylation-dependent transcriptional response and the major regulatory pathways, we looked for ATTAAT motifs within the promoter regions of genes that displayed a similar transcriptional behavior in the M.Hpy99XIX KO strain and in the acid- and iron-responsive regulons. ATTAAT motifs were observed in the promoter regions (within −100 bp of the TSS) in four (*jhp0021*, *jhp1005/hemH*, *jhp1392*, and *jhp1416*) and eight (*jhp0152*, *jhp0334/kgtP*, *jhp0626/fecA1*, *jhp0810/frpB1*, *jhp1107*, *jhp1112*, *jhp1350*, and *jhp1472*) genes differentially expressed in both the M.Hpy99XIX KO strain and the acid-responsive and iron-responsive regulons, respectively. In both cases, ATTAAT motifs were observed more frequently in down- than upregulated genes, indicating that ATTA^{m6}AT methylation in promoter regions generally leads to an increase in gene expression. In particular, seven genes downregulated in response to iron and in the M.Hpy99XIX KO strain carried an ATTAAT motif within their promoter region, including the predicted iron transporter gene *fecA1* (*jhp0626*) (34) and the experimentally confirmed iron transporter gene *frpB1* (*jhp0810*) (35, 36). This observation implies that the pleiotropic regulatory effects connected to the inactivation of the ATTAAT MTase might originate from a defect in iron transport and dysregulation of iron homeostasis.

Up to now, regulation of gene expression via methylation of promoter-bound target sequences has not been demonstrated for the ATTAAT motif. To this end, we generated an isogenic point mutant in which only the ATTAAT motif located in the promoter of the *frpB1* gene (at −80 bp) was mutated in order to prevent methylation at this specific location. Specifically, the methylatable adenine, in the fifth position of the motif (9), was changed to a thymine using targeted mutagenesis (Fig. 4A). Similar to the M.Hpy99XIX KO strain, the *frpB1* point mutant strain showed a significant decrease (>50%) in expression of the *frpB1* gene compared to the wild type (Fig. 4B). This indicates that the expression of the *frpB1* gene is specifically and directly increased by the methylation of the ATTAAT motif within its promoter.

Next, we repeated the experiment with the second transporter gene, *fecA1*. Because *fecA1* contains two ATTAAT motifs in its promoter region (located at −67 and −96 bp of

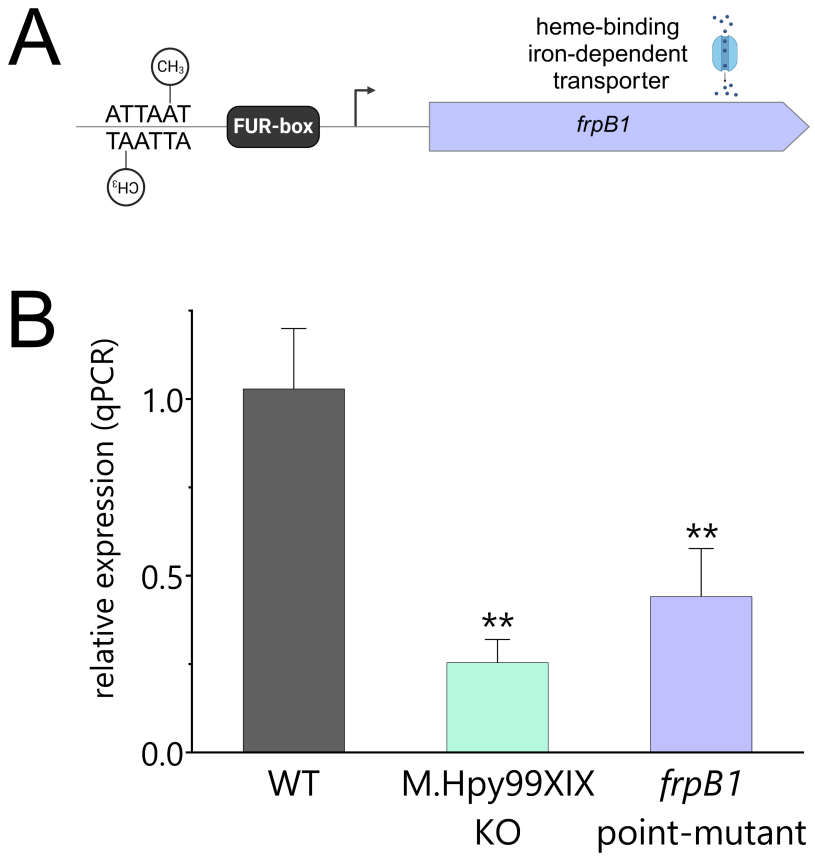


FIG 4 Characterization of the effect of ATTAAT methylation within the promoter region of *frpB1*. (A) Genetic context of the *frpB1* gene, coding for a heme-binding iron-dependent transporter. Upstream of the transcription start site is (i) a FUR box targeted by the transcription factor Fur and (ii) an ATTAAT motif. A point mutant strain was created in which the methylatable A within the motif was changed to a T via targeted mutagenesis in order to prevent methylation at this position. (B) Comparison of *frpB1* expression in the M.Hpy99XIX knockout and the point mutant to the wild-type strain by quantitative PCR. Student's *t*-test: $^{**}P < 0.01$. Error bars indicate the standard deviations of two to five replicates.

the *fecA1* TSS), we created two isogenic point mutants and determined that mutation in only one of the motifs (the one farther from the TSS) led to a lower expression of the *fecA1* gene (~50%) (Fig. S7). By contrast, the mutant of the motif closer to the TSS showed a twofold increase in *fecA1* expression compared to the wild type. A double mutant of both ATTAAT motifs showed a fourfold increase in *fecA1* transcript, suggesting that gene regulation can be adjustable, depending on the number and position of methylated motifs within promoter regions.

Critical regulatory elements are generally conserved across bacterial strains (37, 38). We looked at the sequence diversity of the two ATTAAT motifs involved in the control of expression of the *frpB1* and *fecA1* genes and found these motifs to be highly conserved at 95% and 82% in our *H. pylori* genome collection, respectively. The motif leading to reduced expression of *fecA1* was also highly conserved at 97%. Compared to the average 8% conservation of ATTAAT motifs in the same *H. pylori* genome collection, this indicates that these promoter-bound motifs could be maintained by selection due to their role in regulating gene expression.

M.Hpy99XIX regulates metal homeostasis and oxidative stress

Because FrpB1 is involved in iron acquisition, a decreased expression could have a downstream effect on iron metabolism as well as on associated pathways and regulatory

networks. Furthermore, since the *frpB1* gene only has a single ATTAAT motif in its promoter region, it represents a simplified model to understand the effects of ATTA^{m6}AT methylation. To determine the extent of this theoretical effect, we tested the tolerance of the ATTA^{m6}AT-deficient strains to different conditions.

First, we compared the effects of supplementation with iron [1 mM $(NH4)_2Fe(SO_4)_2$] to those of iron depletion (75 µM 2,2′-dipyridyl [dpp]). Compared to the *H. pylori* J99 wild-type strain, the isogenic M.Hpy99XIX KO strain and the *frpB1* point mutant showed a significant growth defect in both conditions (Fig. 5). Specifically, the KO and point mutant strains had a delay in growth, leading them to grow 70% and 25% slower than the wild type in iron-supplemented media. Under standard culture conditions, no significant differences in growth could be observed between the wild type, M.Hpy99XIX KO, and *frpB1* point mutant strains (Fig. S8).

Through interconnected regulatory networks, iron metabolism has been shown to be influenced by other metal ions. Consequently, we repeated the growth assay in the presence of high concentrations of copper, nickel, and manganese. Interestingly, all metal ions affected bacterial growth in both M.Hpy99XIX KO and *frpB1* point mutant strains. Nickel had the strongest effect, causing a 75% and 40% slower growth in the KO and point mutant, respectively. In addition, the allosteric regulation of Fur has been shown to be triggered by reactive oxygen species. Following exposure to paraquat, we observed that both the M.Hpy99XIX KO and *frpB1* point mutant strains exhibited a significant growth defect, suggesting they are more susceptible to oxidative stress.

In order to test whether the impact of ATTAAT methylation on iron homeostasis was limited to the specific strain tested (J99), we inactivated the orthologous gene M.Hpy99XIX in *H. pylori* strain N6 and tested the N6 wild-type strain and the ATTAAT methylation-deficient mutant for phenotypical effects under the same conditions of iron depletion and iron excess. These experiments with strain N6 yielded comparable results

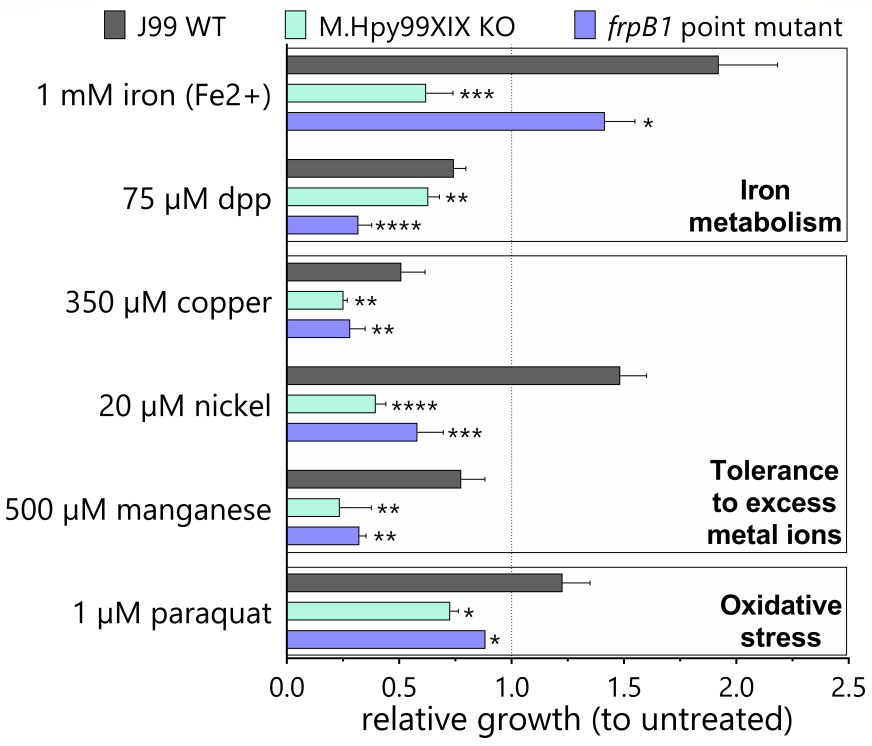


FIG 5 Impact of ATTAAT methylation on growth under environmental conditions related to iron metabolism. Growth of WT and mutants under iron, iron chelation (dpp), copper, nickel, manganese, and oxidative stress (paraquat) conditions was normalized to growth without treatment and compared after 20 h. Student's *t*-test: **P* < 0.05, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001. Error bars indicate the standard deviations of three to five replicates.

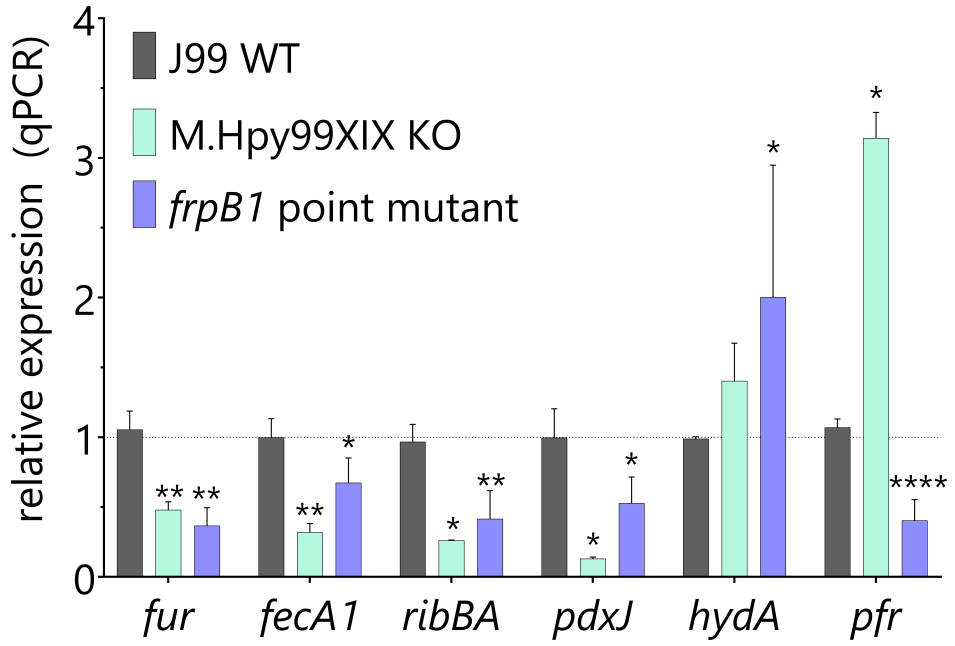


FIG 6 Indirect effects of ATTAAT methylation on the expression of genes related to iron metabolism. Six genes were selected based on their expression patterns in the RNA-Seq data set and their functional role in iron homeostasis. Gene expression was quantitated by qPCR in the J99 wild-type strain, the M.Hpy99XIX knockout, and the *frpB1* point mutant strain. Student's *t*-test for significant differences between wild-type strain and mutants: **P* < 0.05, ***P* < 0.01, *****P* < 0.0001. Error bars indicate the standard deviations of two to six replicates.

to those obtained with strain J99, which confirmed the conservation of this phenotypic effect between strains (Fig. S9).

Altogether, in the two mutant strains (M.Hpy99XIX KO and *frpB1* point mutant), iron homeostasis thus appeared to be dysregulated in the presence of a high or low extracellular iron content that is most likely leading to toxicity and reduced metabolic activity, respectively.

Indirect transcriptional effects mediated by direct ATTA^{m6}AT-based regulation of gene expression

The data shown in the previous sections demonstrate that the disruption of iron homeostasis observed in the ATTAAT methylation-deficient strains is rooted in changes in the Fur regulatory network. In the M.Hpy99XIX KO strain, we determined that a significant proportion of the transcriptomic effect was similar to the changes observed in response to extracellular iron content. We selected six genes that were differentially regulated in the M.Hpy99XIX KO strain and known from previous studies to be involved in iron metabolism (29) and analyzed their expression in the *frpB1* point mutant strain. All six genes carry a Fur box in their promoter (29, 39). Among the selected genes, two were upregulated in the M.Hpy99XIX KO strain, coding for Pfr, a ferritin protein involved in iron storage (40), and HydA, a subunit of a [Ni-Fe] hydrogenase (41). In the *frpB1* point mutant, *hydA* had a similar trend but was not significantly upregulated, whereas *pfr* showed a completely opposite expression pattern (Fig. 6). The remaining four genes were downregulated in the M.Hpy99XIX KO strain and code for FecA1, an iron transporter (35); RibBA and PdxJ, enzymes involved in riboflavin (42) and vitamin B_6 biosynthesis (43), respectively; and Fur (for ferric uptake regulator), a transcription factor mainly involved in the control of iron metabolism. In the *frpB1* point mutant, the genes all displayed a similar expression pattern as well, which indicates that the downregulation of *frpB1*, resulting from the lack of ATTAAT methylation in its promoter

region, leads to the downregulation of other genes interconnected to iron metabolism (Fig. 6).

In *H. pylori*, ferritin (Pfr) is also responsive to nickel, zinc, manganese, and copper levels (40). The discrepancy between the M.Hpy99XIX KO and the *frpB1* point mutant strains might thus be explained by general metabolic differences between strains, resulting themselves from transcriptional effects associated with different methylation profiles at other ATTAAT sites. Altogether, these results suggest that the group of differentially expressed genes observed in the M.Hpy99XIX KO strain is associated with (i) direct regulation of gene expression via promoter-bound ATTAAT methylation for a minority of genes and (ii) indirect transcriptional effects mediated by core regulators and alterations of the iron metabolism resulting from the function of directly regulated genes.

M.Hpy99XIX is restricted to the Ubiquitous *H. pylori* ecospecies

We showed that the M.Hpy99XIX MTase is highly conserved in *H. pylori* and is interconnected with the regulatory pathways governing iron homeostasis, suggesting it could be important for the adaptation of *H. pylori* to the stomach environment, in particular for survival under iron starvation and protection against iron toxicity. The strains in which M.Hpy99XIX is completely absent all belong to a recently described ancient group of *H. pylori*, the "Hardy" ecospecies, distinct from all other *H. pylori* strains (also termed "Ubiquitous" ecospecies) (44). An additional analysis of the 48 *H. pylori* genomes attributed to the Hardy ecospecies confirmed the complete absence of M.Hpy99XIX from this group. Interestingly, the emergence of this ecospecies is thought to be related to differences in human diets (i.e., carnivorous for Hardy ecospecies vs. omnivorous for Ubiquitous). Genetically, the Hardy ecospecies is characterized by 100 consistently differentiated genes from Ubiquitous *H. pylori*. Notably, differentiated genes include genes related to iron and metal homeostasis. In particular, an iron-dependent additional urease is found exclusively in Hardy *H. pylori*, and several unique mutations are observed in iron uptake genes and the Fur transcriptional regulator (44).

Among both gastric and non-gastric *Helicobacter* spp., orthologs of the M.Hpy99XIX MTase are only found in *Helicobacter cetorum* and *Helicobacter macacae*. In bacterial (non-*Helicobacter*) species, orthologs of M.Hpy99XIX are relatively rare and currently found in 15 Gram-negative and 6 Gram-positive species (according to the RM systems database REBASE, version 410) (22). Phylogenetically, the M.Hpy99XIX orthologs fall in two distinct clusters (Fig. 7A). The first cluster consists of genes from Gram-negative bacteria, including *H. pylori* and *H. cetorum*, whereas the second cluster contains both genes from Gram-positive and Gram-negative bacteria, including *H. macacae*. Interestingly, a cognate endonuclease is observed in all organisms from the first cluster with the exception of the two *Helicobacter* spp. (Fig. 7B). Consequently, the loss of the cognate endonuclease in the Hpy99XIX RM system likely happened before the acquisition of M.Hpy99XIX by the common ancestor of *H. pylori* and *H. cetorum*. In the second cluster, a cognate endonuclease is only observed in two species, including *H. macacae*. Altogether, the presence of an unrelated M.Hpy99XIX ortholog and a cognate endonuclease in *H. macacae* indicates that this RM system was likely acquired by *H. macacae* in a separate evolutionary event, unrelated to the acquisition of M.Hpy99XIX by *H. pylori*/*H. cetorum*. Furthermore, the complete absence of remnants of M.Hpy99XIX in the Hardy isolates suggests that this acquisition event in the *H. pylori*/*H. cetorum* lineage likely took place after the divergence of the Hardy ecospecies from the Ubiquitous *H. pylori* isolates.

We have previously identified lineage-specific variations of target motif frequency, suggesting the influence of selective pressure. In this case, we asked whether the acquisition of M.Hpy99XIX in the Ubiquitous branch of *H. pylori* could have influenced the gain and loss of ATTAAT motifs in the genomes, possibly due to their influence on transcription. Genomes of Hardy strains contained on average 25% (~70 motifs/Mb) fewer ATTAAT motifs than genomes of Ubiquitous strains (Fig. 7C). Based on an analysis of compositional bias, the ATTAAT motif is significantly more underrepresented in Hardy

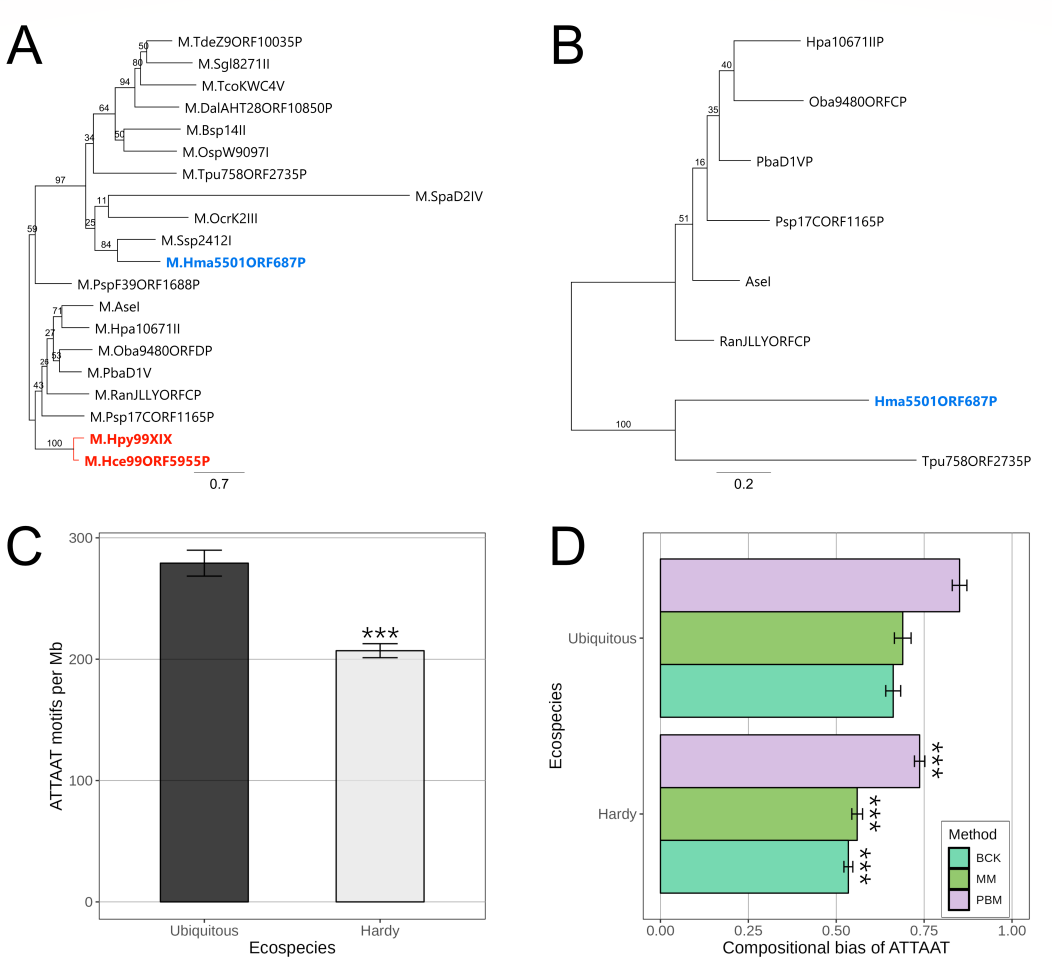


FIG 7 M.Hpy99XIX orthologs in bacteria and distribution of the ATTAAT motifs in Ubiquitous versus Hardy ecospecies. (A) Phylogenetic tree of M.Hpy99XIX orthologs. (B) Phylogenetic tree of endonucleases associated with M.Hpy99XIX orthologs. Genes from *H. pylori*/*H. cetorum* and *H. macacae* are highlighted in red and blue, respectively. The scale bar indicates the number of amino acid substitutions per site. (C) Frequency of the ATTAAT motif in the Hardy and Ubiquitous ecospecies. The number of ATTAAT motifs was counted in each strain and normalized by genome size to obtain the number of motifs per million base pairs. Student's *t*-test: ***$P < 0.001$. (D) Compositional bias of the ATTAAT motif in the Hardy and Ubiquitous ecospecies. The compositional bias represents the ratio of observed to expected number of motifs in the genome. The expected number was calculated using three methods: a maximum-order Markov chain model-based method (MM), the Pevzner and co-author method (PBM), and the Burge and co-author method (BCK).

than in Ubiquitous genomes, suggesting that the presence of the M.Hpy99XIX MTase is associated with an increased density of its target motif (Fig. 7D).

Next, we examined genes related to iron homeostasis whose expression was affected, directly or indirectly, by the M.Hpy99XIX MTase. While the *frpB1* gene was found in all Hardy genomes, the ATTAAT motif located within its promoter was present only in 8% of the strains (4 out of 48). Instead, the sequences ATCAGC and ATTAGC were observed in 90% (43 out of 48) and 2% (1 out of 48) of the genomes. Strikingly, the *fecA1* gene was missing in 85% of the Hardy genomes (41 out of 48) without any ORF remnants or traces of mobile genetic elements. In the remaining 15% of the Hardy genomes (7 out of 48), an intact *fecA1* could be observed, including the promoter regions containing a single or both ATTAAT motifs in five out of seven and two out of seven cases, respectively. Other iron homeostasis genes considered in this study (*fur*, *ribBA*, *pdxJ*, *hydA*, and *pfr*) were present in all Hardy genomes. These results indicate that the absence of the ATTAAT MTase is a conserved feature of Hardy strains, consistent with its role in iron homeostasis.

DISCUSSION

The role of bacterial orphan MTases has long remained enigmatic since, due to the lack of a cognate restriction endonuclease, they cannot serve in the defense against invading DNA. Until now, they have been typically associated with transcriptional regulation of genes involved in important cellular processes (12, 13, 45). Deletion of specific type II MTases in *H. pylori* can lead to transcriptional changes affecting large groups of genes (12, 13). DNA methylation has also been linked to a multitude of phenotypes, although the functional connections between (loss of) methylation and specific phenotypes have not been elucidated in most cases. For the GCGC-specific MTase M.Hpy99III, effects on specific genes have been linked to methylation of GCGC motifs within their promoter regions. However, GCGC motifs in promoter regions were found only in a small subset of genes differentially regulated by inactivation of GCGC methylation, suggesting an important role of indirect regulation to the overall transcriptional response following disruption of GCGC methylation (12, 24).

In this study, we characterized the transcriptional response and phenotypes associated with the ATTAAT-specific M.Hpy99XIX MTase (9) and demonstrated how it is intertwined with the iron homeostasis regulatory pathway. We determined that M.Hpy99XIX is the only *H. pylori* MTase that is never associated with a cognate endonuclease and therefore the only strictly orphan MTase in the species. Other MTases have also been shown to be variably orphan in *H. pylori*. In particular, in the RM.Hpy99III RM system, targeting GCGC, the endonuclease is often truncated in *H. pylori* geographical populations other than hpAfrica2 (12).

In *H. pylori*, the M.Hpy99XIX gene is present and predicted to be active in 99% of the strains. The rarity of predicted loss-of-function mutations suggests that this MTase likely serves an important function for *H. pylori*. Furthermore, strains with an inactivated M.Hpy99XIX gene might have either evolved compensatory mutations or could represent evolutionary dead ends unable to compete with strains carrying an active M.Hpy99XIX. Strains in which the M.Hpy99XIX gene is completely missing are restricted to the recently defined Hardy ecospecies of *H. pylori* (44), which is distinct from the Ubiquitous ecospecies, the latter representing the diverse *H. pylori* strains that have been studied up to now. The absence of related M.Hpy99XIX orthologs in other *Helicobacter* spp., other than *H. cetorum*, and the lack of any ORF remnants or insertion sites in Hardy genomes suggest that M.Hpy99XIX was never present in this group. The common ancestor of the Hardy and Ubiquitous ecospecies is thought to be extremely ancient and to have diverged before the split between the two ancestral super-lineages (the first leading to hpAfrica2 and the second to hpAfrica1 and all other populations) of *H*. pylori that happened at least 100 kya ago (46). Therefore, it seems most likely that M.Hpy99XIX was acquired, possibly via horizontal gene transfer, by *H. pylori* Ubiquitous after the subdivision between Hardy and Ubiquitous ecospecies but prior to the split into the two *H. pylori* super-lineages (47). The absence of remnants of a potential cognate endonuclease and the existence of orphan ATTAAT MTases in other bacterial species indicate that M.Hpy99XIX may have been acquired directly as an orphan MTase. Interestingly, the Hardy ecospecies is characterized by an array of fixed mutations (44) in the Fur and ArsRS transcription factors and iron acquisition genes (*frpB4*, *tonB1*, *exbB*, and *exbD*) and a specific iron-dependent urease enzyme in addition to the nickel-dependent urease already present in all *H. pylori* strains and strictly required for pathogenesis. In particular, these genetic traits are thought to represent an evolutionary strategy adapted to a carnivorous diet. On the opposite, we showed that (i) the promoter-bound ATTAAT motif modulating the expression of *frpB1* and (ii) the entire *fecA1* gene (including its promoter with two ATTAAT motifs) are highly conserved and specific to Ubiquitous strains, compared to Hardy strains. Collectively, the acquisition of M.Hpy99XIX and the evolution of ATTA^{m6}AT methylation-dependent iron-uptake genes in Ubiquitous strains are likely the result of synergistic evolution in response to the iron status encountered by *H. pylori* and contributed to the delineation of distinct ecospecies.

Inactivation of the M.Hpy99XIX gene led to the differential regulation of many genes. More genes were downregulated than upregulated in the absence of ATTAAT methylation, suggesting that methylation has the tendency to increase gene expression, which is in line with previous observations with the M.Hpy99III MTase, for instance (12). In a different strain, *H. pylori* P12, deletion of the homologous ATTAAT MTase gene (termed M.HpyPVII in this particular strain) also led to the differential regulation of 102 genes (13). Around 16% of the differentially expressed genes matched between *H. pylori* J99 and P12, which is similar to the extent of strain-specific variation observed with the GCGC-specific MTase M.Hpy99III in *H. pylori* strains J99 and BCM300 (12). Furthermore, RNA-Seq experiments were performed at different optical densities (optical density at 600 nm [OD_{600}] = 0.8 for J99 and 0.4 for P12) and analyzed using different methods (DESeq2 for J99 and TCC for P12), likely introducing some additional bias into the comparison of both strains.

To understand the basis of the transcriptional effect observed in the M.Hpy99XIX KO strain, we compared all the significantly regulated genes to known *H. pylori* regulons (12, 28–33). Interestingly, there was a substantial overlap between ATTAAT methylation-regulated genes and iron-regulated genes. In *H. pylori* and other bacteria, iron homeostasis is maintained by Fur (ferric uptake regulator) and a network of differentially regulated genes according to extracellular iron availability (29, 40, 43). In particular, the allosteric behavior of Fur produces a conformational switch when bound to a Fe^{2+} co-factor, which modulates its binding to regulatory elements (27, 48, 49). Within our data set, we identified two genes, *fecA1* and *frpB1*, that are both involved in iron acquisition (35, 50, 51), and both carry an ATTAAT motif as well as a FUR box in their promoter regions (39). By specifically mutating these ATTAAT motifs, we could demonstrate that ATTAAT methylation exerts a direct transcriptional effect on these two genes. The precise mechanism behind direct methylation-based transcriptional regulation remains to be characterized. In *H. pylori*, a similar phenomenon was also demonstrated by targeted mutagenesis of a GCGC motif located in the promoter region of the toxin-antitoxin *jhp0832* gene and was later validated using a reporter assay for three additional promoters (12, 24). However, GCGC motifs within promoters tend to be located at the −13 position (relative to the transcription start site) (24), whereas ATTAAT did not appear to have any bias toward a specific position, suggesting that there might not be a common mechanism underlying methylation-based transcriptional regulation between the two MTases and that only a subset of ATTAAT motifs within promoters may have a role in gene expression, depending on uncharacterized factors. Interestingly, bacterial promoters are typically AT rich (52), and the ATTAAT motif is the only known 100% AT type II motif in *H. pylori,* as well as the most overrepresented in promoter regions, suggesting that the AT content of motifs likely influences their frequencies in this context.

The promoter region of the *fur* gene itself contains a FUR box (39) but not an ATTAAT motif and thus is unlikely to be directly regulated by M.Hpy99XIX. Nevertheless, we could show that the tolerance of the M.Hpy99XIX KO mutant to both high and low extracellular iron conditions was significantly decreased compared to the corresponding wild type in two *H. pylori* strains, J99 as well as N6, indicating that M.Hpy99XIX can indirectly contribute to the maintenance of iron homeostasis in *H. pylori*. The disruption of the Fur regulatory network was further confirmed by the decreased tolerance of the M.Hpy99XIX mutants to other metal ions, such as nickel or copper, and to oxidative stress. Although *fur* cannot be directly regulated by M.Hpy99XIX, it was nonetheless downregulated in the absence of ATTA^{m6}AT methylation, again indicating an indirect connection between the two systems. We surmise that, because of the allosteric nature of Fur activation (49), modulation of iron content is likely to have a larger effect on the Fur regulon than a potential direct or indirect regulation of the expression of *fur* itself. Strikingly, targeted mutagenesis of the single ATTAAT motif (thus preventing methylation at this site) in the promoter region of the *frpB1* gene yielded very similar phenotypes to the M.Hpy99XIX KO strain in all tested conditions. Furthermore, we could also demonstrate

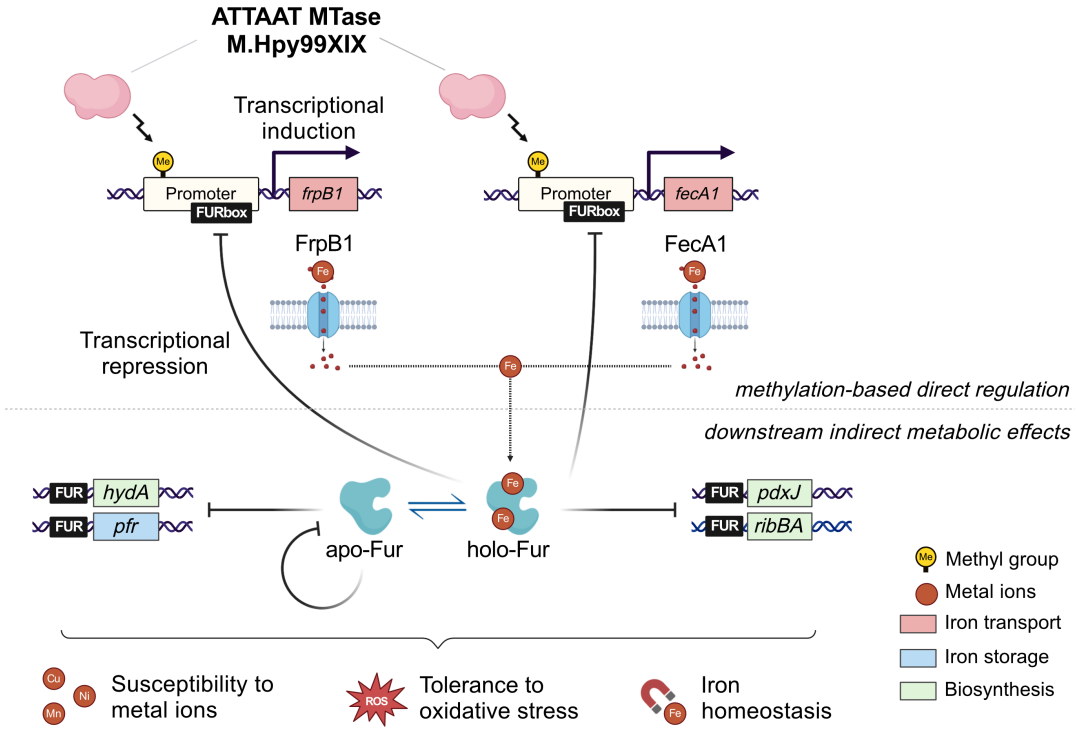


FIG 8 Model of the direct and indirect transcriptional and phenotypic effects of the M.Hpy99XIX methyltransferase. The effects of ATTAAT methylation by M.Hpy99XIX can be described in two steps. The first step consists of direct gene regulation via methylation of an ATTAAT motif within the promoter region, demonstrated by qPCR and RNA-Seq for two genes involved in iron acquisition, *frpB1* and *fecA1*. Changes in iron acquisition and intracellular iron content influence the allosteric regulation of fur, the ferric uptake regulator, preceding the second step. Switches between the apo- and holo-Fur conformations lead to both induction and repression of genes within the Fur regulatory network, including *frpB1* and *fecA1*, as well as genes related to iron biosynthesis and storage, as shown by qPCR. These indirect downstream effects can then generate a feedback loop and cause downstream metabolic effects such as changes in susceptibility to metal ions and tolerance to oxidative stress, as demonstrated by phenotype growth assays in multiple environmental conditions.

similar expression patterns between the *frpB1* point mutant and the M.Hpy99XIX KO strains for six genes associated with iron homeostasis, indicating that the Fur regulatory network is disrupted in both mutant strains.

Altogether, our data suggest that profound changes in the transcriptome in response to loss of a MTase can be mediated by very focused direct transcriptional regulation of a small group of genes. The effects of the M.Hpy99XIX MTase on *H. pylori* observed in this study can be summarized as a two-step model (Fig. 8). The first step consists of direct gene regulation via methylation of ATTAAT motifs located within the promoter region of two genes involved in iron acquisition, *frpB1* and *fecA1*. Changes in iron transport and thus intracellular iron content modulate the allosteric regulation of Fur (29, 40, 43), the ferric uptake regulator, preceding the second step. Switches between the apo- and holo-Fur conformations lead to either induction or repression of genes within the Fur regulatory network, including *frpB1* and *fecA1*, as well as genes related to iron biosynthesis and storage (29, 40, 43, 53). These indirect downstream effects can then generate a feedback loop and cause downstream metabolic effects such as a change in susceptibility to metal ions and tolerance to oxidative stress. Additional mechanisms related to gene regulation, such as non-coding regulatory RNAs or conformational alterations of chromosomal DNA regions (54, 55), may also be involved in the connection between methylation and regulatory networks and will be worthwhile to consider in future studies.

In conclusion, the role of M.Hpy99XIX in gene expression and in iron homeostasis represents compelling evidence of the importance of this orphan MTase outside the confines of a typical RM system in *H. pylori* and suggests that the methylome could

represent a universal and dynamic interface linking genome diversity and transcriptional regulation. Its general absence from the Hardy ecospecies is striking and will likely provide novel clues to inform future studies into the evolution and functional differentiation of the two ecospecies. Furthermore, these findings underscore the potential for bacterial epigenetics to influence not only short-term adaptations of individual bacteria of one species to changes in their ecological niche but also broader ecological interactions which may influence further species evolution.

MATERIALS AND METHODS

Analyses of type II RM system frequencies in *H. pylori* and other bacterial species

The genome collection used to determine the frequency of methyltransferases and endonucleases, normalized across four geographical populations of *H. pylori* (hpAfrica1, hpAfrica2, hpEurope, and hpAsia2), is based on our previous analysis of the methylome diversity in *H. pylori* (8). The genome collection used to characterize the absence and mechanisms of inactivation of the M.Hpy99XIX MTase was built out of 988 complete genomes from the HpGP project (23) as well as 24 additional complete genomes from several *H. pylori* evolutionary studies (56–58) and reference strains (5, 59–62). MTase and endonuclease sequences from *H. pylori* were retrieved from the REBASE database (22), and their presence in the collection was assessed using BLASTn and BLASTp (63). To infer the location of transcription start sites and the sequence of promoter regions, previously published promoter regions of the 50 bp upstream sequence of the TSS obtained using the dRNA-Seq method (64) were mapped to the genome collection using bbmap (65) with the following settings: vslow, k = 8, maxindel = 200, minid = 0.5, secondary = f and ordered = t.

Sequences of orthologs of the M.Hpy99XIX MTase and endonucleases were obtained from the RM system database REBASE (22). Protein sequences were aligned using MAFFT (66), and phylogenetic trees were built using RAxML with the rapid hill-climbing mode, the GAMMA LG model, and 100 bootstrap replicates (67).

Analyses of target motif frequencies and distribution across the *H. pylori* genome

Frequency and position of MTase target motifs were analyzed using the Biostrings version 2.58.0 Bioconductor R package (68). Fold change of the target motifs in promoter regions was calculated as the $\log_2$ of the ratio of motif frequency within promoter over motif frequency in the genome. A positive fold change indicates overrepresentation of the motif in promoters, whereas a negative fold change indicates underrepresentation. The position of the ATTAAT and TATAAT motif across all promoter regions was summarized by computing a probability density function using the R package mclust (69).

Enrichment of the ATTAAT motif in coding sequences was calculated using the distAMo software (70). z-scores of 2 and −2 were considered significant enrichment and depletion, respectively.

Compositional bias of the ATTAAT motif was calculated using the methods of Burge and co-authors (71), Pevzner and co-authors (72), and a maximum-order Markov chain model implemented in the CBcalc software (73).

Bacterial culture and growth curves

H. pylori strains J99 and N6 were cultured from frozen stocks on blood agar plates or in liquid cultures as described previously (12). Liquid cultures of bacteria were grown in brain heart infusion broth (BD Difco, Heidelberg, Germany) with yeast extract (2.5 g L^{-1}), 10% heat-inactivated horse serum, and a mix of antibiotics (vancomycin [10 mg L^{-1}], polymyxin B [3.2 mg L^{-1}], amphotericin B [4 mg L^{-1}], and trimethoprim [5 mg L^{-1}]). Anaerocult C gas-producing bags (Merck, Darmstadt, Germany) were used to generate

microaerobic conditions in airtight jars (Oxoid, Wesel, Germany). Growth curves were generated for the different strains of 10 mL liquid culture starting at an OD_{600} of 0.06 in 50 mL falcons, incubated with shaking (140 rpm, 37°C, microaerobic conditions) and measured after 8, 24, 32, 48, 56, 72, and 80 h of growth.

RNA extraction

For RNA extraction, 5 mL of bacterial culture grown in liquid media to an OD_{600} of 0.8 was pelleted (4°C, 6,000 × *g*, 5 min), snap-frozen in liquid nitrogen, and stored at −80°C. Afterward, bacterial pellets were disrupted with a Precellys 24 Homogenizer (Bertin Technologies, Montigny-le-Bretonneux, France) using Lysing Matrix B 2 mL tubes containing 0.1 mm silica beads (MP Biomedicals, Eschwege, Germany). Isolation of RNA was performed with the RNeasy Mini Kit (Qiagen, Hilden, Germany) with an on-column DNase digestion with DNase I and a second DNase treatment using the TURBO DNA-free Kit (Invitrogen, Carlsbad, USA). Isolated RNA was checked for absence of DNA contamination by PCR of the 16S rRNA gene (12). The RNA concentration was measured using the NanoDrop 2000 spectrophotometer (Peqlab Biotechnologies, Erlangen, Germany). The quality of the RNA was tested with the 4200 TapeStation System using RNA Screen Tapes (Agilent, Waldbronn, Germany). All RINe numbers of prepared RNA were higher than 8.8, confirming high quality and little RNA degradation.

Transcriptomic analysis: RNA sequencing

RNA-Seq was performed on an Illumina sequencing platform (2 × 150 bp, >10 M read pairs) by Eurofins (Ebersberg, Germany). Probe-based ribosomal RNA depletion was performed for each sample prior to mRNA fragmentation, strand-specific cDNA synthesis, and library preparation. Three biological replicates were used for each condition. Data were analyzed with the R package DESeq2 (74). An absolute fold change larger than one and an FDR adjusted *P* value below 0.01 were used as a cutoff.

Construction of *H. pylori* mutant strains

Inactivation of the MTase in *H. pylori* strain J99 was carried out by the Multiplex Genome Editing (MuGent) technique as previously described (47, 75). Briefly, we use natural co-transformation with a chloramphenicol resistance cassette within the non-essential *rdxA* locus as a selective marker in addition to a non-selective PCR product carrying a construct of interest. To generate gene knockouts, 1 kb regions up- and downstream of the MTase were amplified by PCR, digested, and ligated into a digested pUC19 vector. A PCR product of the deletion construct was then used as a non-selective product in the co-transformation reaction. To perform targeted mutagenesis and generate point mutants, a 1 kb fragment of the target sequence was cloned into the pUC19 vector, and an inverse PCR was performed using primers carrying the desired mutation. The resulting PCR product was then digested with DpnI (NEB, Frankfurt am Main, Germany), PCR purified with the QIAquick PCR Purification Kit (Qiagen), and re-ligated using the Quick Ligation Kit (NEB) before transformation into *E. coli* MC1061 (12, 76). PCR products of the target sequence were then used as a non-selective product in the co-transformation reaction (47, 75). Sanger sequencing was used to verify the acquisition of the desired mutations. Inactivation of the MTase in *H. pylori* strain N6 was carried out by insertion of an *aphA-3* cassette conferring kanamycin resistance into the MTase gene. A vector for this was constructed with the NEB Gibson Assembly Master Mix (NEB) using a digested pUC19 vector and PCR products of the MTase gene and the kanamycin resistance gene. The resulting plasmid was recovered after transformation of the reaction into *E. coli* MC1061 and characterized by PCR and Sanger sequencing. One hundred nanograms of purified plasmid was then used for the transformation in *H. pylori*.

Lastly, we generated a functional complementation of the MTase gene in the J99 MTase KO strain using the pADC/CAT suicide plasmid approach, as described previously (12, 77). The MTase gene was inserted into the urease locus under the control of the urease operon promoter.

The *H. pylori* mutants were checked via PCR and selected on antibiotic-containing plates. The absence or recovery of ATTAAT methylation was checked by digestion of gDNA with AseI (NEB). The primers and plasmids used in this study are described in Table S4.

Transcriptomic analysis: quantitative PCR

One microgram of RNA was used for cDNA synthesis using the SuperScript III Reverse Transcriptase (Thermo Fischer Scientific, Darmstadt, Germany) as described previously (78). Quantitative PCR was performed with gene-specific primers (Table S4) and SYBR Green Master Mix (Qiagen). Reactions were run in a BioRad CFX96 system. Standard curves and samples were run in biological quadruplicate and technical duplicate. Samples were normalized to an internal 16S rRNA control for quantitative comparison. The quality standards used correspond to those in our previously published MIQE (12).

Growth assays

Susceptibility to iron [1 mM $(NH4)_2Fe(SO_4)_2$], copper (0.35 mM $CuSO_4$), zinc (0.5 mM $ZnCl_2$), manganese (0.5 mM $MnCl_{2)}$, nickel (20 µM $NiSO_4$), oxidative stress (1 µM paraquat), and iron chelation [75 µM dpp (2,2′-dipyridyl)] was tested by adding the respective compounds to 100 µL liquid cultures grown in 96-well plates with a starting OD_{600} of 0.1. The OD_{600} was measured for the following 18 h with a CLARIOstar microplate reader (BMG Labtech, Ortenberg, Germany) under constant 10% CO_2. The average of the last 10 measurements obtained over a 20 min period was used as the end-point OD_{600} for each measurement. The end-point values obtained for each compound were normalized to the final OD_{600} of an untreated control.

Reproducibility

All data were analyzed using R version 4.2.3, Geneious 11.0.4, or GraphPad Prism version 9.4.1.

ACKNOWLEDGMENTS

We thank Gudrun Pfaffinger and Evelyn Weiss for their excellent technical assistance.

Funding was provided by the Deutsche Forschungsgemeinschaft (project no. 158989968, grants SFB900/A1 and SFB900/Z1 to S.S. and grant SFB900/B6 to C.J.), by the Bavarian Ministry of Science and the Arts through project HelicoPredict in the framework of the research network bayresq.net (to S.S.), and by the Deutsches Zentrum für Infektionsforschung, Partner Site Munich, Germany (no. TTU 06.824 to S.S. and TTU 06.820 to C.J.).

W.G. designed, performed, and analyzed experiments; performed bioinformatic analysis; and contributed to the manuscript writing, figure design, final reading, and editing. F.A. designed and analyzed experiments; designed, performed, and supervised bioinformatic analysis; and contributed to manuscript writing, figure design, final reading, and editing. C.J. designed and analyzed experiments and contributed to manuscript writing, final reading, and editing. S.S. conceptualized the study; acquired funding; designed and analyzed experiments; and contributed to manuscript writing, figure design, final reading, and editing. All authors read and approved the final version.

AUTHOR AFFILIATIONS

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

AUTHOR ORCIDs

Wilhelm Gottschall http://orcid.org/0000-0003-0118-4166
Florent Ailloud http://orcid.org/0000-0001-9063-0839
Christine Josenhans http://orcid.org/0000-0003-4620-2954
Sebastian Suerbaum http://orcid.org/0000-0001-9583-7505

FUNDING

Funder	Grant(s)	Author(s)
Deutsche Forschungsgemeinschaft	158989968	Christine Josenhans Sebastian Suerbaum
Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst (Bavarian State Ministry of Education, Science and the Arts)	Helicopredict	Sebastian Suerbaum
Deutsches Zentrum für Infektionsforschung	06.824	Sebastian Suerbaum
Deutsches Zentrum für Infektionsforschung	06.820	Christine Josenhans

AUTHOR CONTRIBUTIONS

Wilhelm Gottschall, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review and editing | Florent Ailloud, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing | Christine Josenhans, Conceptualization, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review and editing | Sebastian Suerbaum, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review and editing

DIRECT CONTRIBUTION

This article is a direct contribution from Sebastian Suerbaum, a Fellow of the American Academy of Microbiology, who arranged for and secured reviews by Timothy Cover, Vanderbilt University Medical Center, and Seyed Hasnain, Jamia Hamdard.

DATA AVAILABILITY

Raw sequencing data generated in this study were deposited in the National Center for Biotechnology Information Sequence Read Archive database under BioProject ID PRJNA1238819.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Supplemental Figures (mBio01209-25-s0001.pdf). Figures S1 to S9.
Supplemental Tables (mBio01209-25-s0002.xlsx). Tables S1 to S4.

REFERENCES

1. Malfertheiner P, Camargo MC, El-Omar E, Liou JM, Peek R, Schulz C, Smith SI, Suerbaum S. 2023. *Helicobacter pylori* infection. Nat Rev Dis Primers 9:19. https://doi.org/10.1038/s41572-023-00431-8
2. Ailloud F, Estibariz I, Suerbaum S. 2021. Evolved to vary: genome and epigenome variation in the human pathogen *Helicobacter pylori*. FEMS Microbiol Rev 45:fuaa042. https://doi.org/10.1093/femsre/fuaa042
3. Suerbaum S, Michetti P. 2002. *Helicobacter pylori* infection. N Engl J Med 347:1175–1186. https://doi.org/10.1056/NEJMra020542
4. Tomb J-F, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA, et al. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature 388:539–547. https://doi.org/10.1038/41483

5. Alm RA, Trust TJ. 1999. Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes. J Mol Med (Berl) 77:834–846. https://doi.org/10.1007/s001099900067
6. Krebes J, Didelot X, Kennemann L, Suerbaum S. 2014. Bidirectional genomic exchange between *Helicobacter pylori* strains from a family in Coventry, United Kingdom. Int J Med Microbiol 304:1135–1146. https://doi.org/10.1016/j.ijmm.2014.08.007
7. Vasu K, Nagaraja V. 2013. Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol Mol Biol Rev 77:53–72. https://doi.org/10.1128/MMBR.00044-12
8. Ailloud F, Gottschall W, Suerbaum S. 2023. Methylome evolution suggests lineage-dependent selection in the gastric pathogen *Helicobacter pylori*. Commun Biol 6:839. https://doi.org/10.1038/s42003-023-05218-x
9. Krebes J, Morgan RD, Bunk B, Spröer C, Luong K, Parusel R, Anton BP, König C, Josenhans C, Overmann J, Roberts RJ, Korlach J, Suerbaum S. 2014. The complex methylome of the human gastric pathogen *Helicobacter pylori*. Nucleic Acids Res 42:2415–2432. https://doi.org/10.1093/nar/gkt1201
10. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al. 2009. Real-time DNA sequencing from single polymerase molecules. Science 323:133–138. https://doi.org/10.1126/science.1162986
11. Stoddart D, Heron AJ, Mikhailova E, Maglia G, Bayley H. 2009. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. Proc Natl Acad Sci U S A 106:7702–7707. https://doi.org/10.1073/pnas.0901054106
12. Estibariz I, Overmann A, Ailloud F, Krebes J, Josenhans C, Suerbaum S. 2019. The core genome ^{m5}C methyltransferase JHP1050 (M.Hpy99III) plays an important role in orchestrating gene expression in *Helicobacter pylori*. Nucleic Acids Res 47:2336–2348. https://doi.org/10.1093/nar/gky1307
13. Yano H, Alam MZ, Rimbara E, Shibata TF, Fukuyo M, Furuta Y, Nishiyama T, Shigenobu S, Hasebe M, Toyoda A, Suzuki Y, Sugano S, Shibayama K, Kobayashi I. 2020. Networking and specificity-changing DNA methyltransferases in *Helicobacter pylori*. Front Microbiol 11:1628. https://doi.org/10.3389/fmicb.2020.01628
14. Casadesús J, Low D. 2006. Epigenetic gene regulation in the bacterial world. Microbiol Mol Biol Rev 70:830–856. https://doi.org/10.1128/MMBR.00016-06
15. Braaten BA, Nou X, Kaltenbach LS, Low DA. 1994. Methylation patterns in *pap* regulatory DNA control pyelonephritis-associated pili phase variation in *E. coli*. Cell 76:577–588. https://doi.org/10.1016/0092-8674(94)90120-1
16. Camacho EM, Serna A, Madrid C, Marqués S, Fernández R, de la Cruz F, Juárez A, Casadesús J. 2005. Regulation of *finP* transcription by DNA adenine methylation in the virulence plasmid of *Salmonella enterica*. J Bacteriol 187:5691–5699. https://doi.org/10.1128/JB.187.16.5691-5699.2005
17. Kumar S, Karmakar BC, Nagarajan D, Mukhopadhyay AK, Morgan RD, Rao DN. 2018. N^4-cytosine DNA methylation regulates transcription and pathogenesis in *Helicobacter pylori*. Nucleic Acids Res 46:3429–3445. https://doi.org/10.1093/nar/gky126
18. Kumar R, Mukhopadhyay AK, Ghosh P, Rao DN. 2012. Comparative transcriptomics of *H. pylori* strains AM5, SS1 and their *hpyAVIBM* deletion mutants: possible roles of cytosine methylation. PLoS One 7:e42303. https://doi.org/10.1371/journal.pone.0042303
19. Loenen WAM, Dryden DTF, Raleigh EA, Wilson GG, Murray NE. 2014. Highlights of the DNA cutters: a short history of the restriction enzymes. Nucleic Acids Res 42:3–19. https://doi.org/10.1093/nar/gkt990
20. Luria SE, Human ML. 1952. A nonhereditary, host-induced variation of bacterial viruses. J Bacteriol 64:557–569. https://doi.org/10.1128/jb.64.4.557-569.1952
21. Wilson GG. 1991. Organization of restriction-modification systems. Nucleic Acids Res 19:2539–2566. https://doi.org/10.1093/nar/19.10.2539
22. Roberts RJ, Vincze T, Posfai J, Macelis D. 2023. REBASE: a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 51:D629–D630. https://doi.org/10.1093/nar/gkac975
23. Thorell K, Muñoz-Ramírez ZY, Wang D, Sandoval-Motta S, Boscolo Agostini R, Ghirotto S, Torres RC, HpGP Research Network, Falush D, Camargo MC, Rabkin CS. 2023. The *Helicobacter pylori* Genome Project: insights into *H. pylori* population structure from analysis of a worldwide collection of complete genomes. Nat Commun 14:8184. https://doi.org/10.1038/s41467-023-43562-y
24. Meng B, Epp N, Wijaya W, Mrázek J, Hoover TR. 2021. Methylation motifs in promoter sequences may contribute to the maintenance of a conserved (m5)C methyltransferase in *Helicobacter pylori*. Microorganisms 9:12. https://doi.org/10.3390/microorganisms9122474
25. Pribnow D. 1975. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. Proc Natl Acad Sci U S A 72:784–788. https://doi.org/10.1073/pnas.72.3.784
26. Alm RA, Ling LS, Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, deJonge BL, Carmel G, Tummino PJ, Caruso A, Uria-Nickelsen M, Mills DM, Ives C, Gibson R, Merberg D, Mills SD, Jiang Q, Taylor DE, Vovis GF, Trust TJ. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. Nature 397:176–180. https://doi.org/10.1038/16495
27. Vannini A, Roncarati D, D'Agostino F, Antoniciello F, Scarlato V. 2022. Insights into the orchestration of gene transcription regulators in *Helicobacter pylori*. Int J Mol Sci 23:22. https://doi.org/10.3390/ijms232213688
28. Loh JT, Shum MV, Jossart SDR, Campbell AM, Sawhney N, McDonald WH, Scholz MB, McClain MS, Forsyth MH, Cover TL. 2021. Delineation of the pH-responsive regulon controlled by the *Helicobacter pylori* ArsRS two-component system. Infect Immun 89:e00597-20. https://doi.org/10.1128/IAI.00597-20
29. Vannini A, Pinatel E, Costantini PE, Pelliciari S, Roncarati D, Puccio S, De Bellis G, Scarlato V, Peano C, Danielli A. 2024. (Re)-definition of the holo- and *apo*-Fur direct regulons of *Helicobacter pylori*. J Mol Biol 436:168573. https://doi.org/10.1016/j.jmb.2024.168573
30. Roncarati D, Danielli A, Spohn G, Delany I, Scarlato V. 2007. Transcriptional regulation of stress response and motility functions in *Helicobacter pylori* is mediated by HspR and HrcA. J Bacteriol 189:7234–7243. https://doi.org/10.1128/JB.00626-07
31. Vannini A, Pinatel E, Costantini PE, Pelliciari S, Roncarati D, Puccio S, De Bellis G, Peano C, Danielli A. 2017. Comprehensive mapping of the *Helicobacter pylori* NikR regulon provides new insights in bacterial nickel responses. Sci Rep 7:45458. https://doi.org/10.1038/srep45458
32. Pflock M, Bathon M, Schär J, Müller S, Mollenkopf H, Meyer TF, Beier D. 2007. The orphan response regulator HP1021 of *Helicobacter pylori* regulates transcription of a gene cluster presumably involved in acetone metabolism. J Bacteriol 189:2339–2349. https://doi.org/10.1128/JB.01827-06
33. Hung C-L, Cheng H-H, Hsieh W-C, Tsai ZT-Y, Tsai H-K, Chu C-H, Hsieh W-P, Chen Y-F, Tsou Y, Lai C-H, Wang W-C. 2015. The CrdRS two-component system in *Helicobacter pylori* responds to nitrosative stress. Mol Microbiol 97:1128–1141. https://doi.org/10.1111/mmi.13089
34. Hussein S, Hantke K, Braun V. 1981. Citrate-dependent iron transport system in *Escherichia coli* K-12. Eur J Biochem 117:431–437. https://doi.org/10.1111/j.1432-1033.1981.tb06357.x
35. Ernst FD, Bereswill S, Waidner B, Stoof J, Mäder U, Kusters JG, Kuipers EJ, Kist M, van Vliet AHM, Homuth G. 2005. Transcriptional profiling of *Helicobacter pylori* Fur- and iron-regulated gene expression. Microbiology (Reading) 151:533–546. https://doi.org/10.1099/mic.0.27404-0
36. Carrizo-Chávez MA, Cruz-Castañeda A, Olivares-Trejo J de J. 2012. The *frpB1* gene of *Helicobacter pylori* is regulated by iron and encodes a membrane protein capable of binding haem and haemoglobin. FEBS Lett 586:875–879. https://doi.org/10.1016/j.febslet.2012.02.015
37. Hawley DK, McClure WR. 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res 11:2237–2255. https://doi.org/10.1093/nar/11.8.2237
38. Moran CP, Lang N, LeGrice SF, Lee G, Stephens M, Sonenshein AL, Pero J, Losick R. 1982. Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. Mol Gen Genet 186:339–346. https://doi.org/10.1007/BF00729452
39. Pich OQ, Carpenter BM, Gilbreath JJ, Merrell DS. 2012. Detailed analysis of *Helicobacter pylori* Fur-regulated promoters reveals a Fur box core sequence and novel Fur-regulated genes. Mol Microbiol 84:921–941. https://doi.org/10.1111/j.1365-2958.2012.08066.x
40. Bereswill S, Greiner S, van Vliet AHM, Waidner B, Fassbinder F, Schiltz E, Kusters JG, Kist M. 2000. Regulation of ferritin-mediated cytoplasmic iron storage by the ferric uptake regulator homolog (Fur) of *Helicobacter pylori*. J Bacteriol 182:5948–5953. https://doi.org/10.1128/JB.182.21.5948-5953.2000
41. Contreras M, Thiberge J-M, Mandrand-Berthelot M-A, Labigne A. 2003. Characterization of the roles of NikR, A nickel-responsive pleiotropic autoregulator of *Helicobacter pylori*. Mol Microbiol 49:947–963. https://doi.org/10.1046/j.1365-2958.2003.03621.x

42. Worst DJ, Gerrits MM, Vandenbroucke-Grauls CM, Kusters JG. 1998. *Helicobacter pylori ribBA*-mediated riboflavin production is involved in iron acquisition. J Bacteriol 180:1473–1479. https://doi.org/10.1128/JB.180.6.1473-1479.1998
43. Gancz H, Censini S, Merrell DS. 2006. Iron and pH homeostasis intersect at the level of Fur regulation in the gastric pathogen *Helicobacter pylori*. Infect Immun 74:602–614. https://doi.org/10.1128/IAI.74.1.602-614.2006
44. Tourrette E, Torres RC, Svensson SL, Matsumoto T, Miftahussurur M, Fauzia KA, Alfaray RI, Vilaichone R-K, Tuan VP, Wang D, Yadegar A, Olsson LM, Zhou Z, Yamaoka Y, Thorell K, Falush D, Helicobacter Genomics Consortium. 2024. An ancient ecospecies of *Helicobacter pylori*. Nature 635:178–185. https://doi.org/10.1038/s41586-024-07991-z
45. Oliveira PH, Fang G. 2021. Conserved DNA methyltransferases: a window into fundamental mechanisms of epigenetic regulation in bacteria. Trends Microbiol 29:28–40. https://doi.org/10.1016/j.tim.2020.04.007
46. Moodley Y, Linz B, Bond RP, Nieuwoudt M, Soodyall H, Schlebusch CM, Bernhöft S, Hale J, Suerbaum S, Mugisha L, van der Merwe SW, Achtman M. 2012. Age of the association between *Helicobacter pylori* and man. PLoS Pathog 8:e1002693. https://doi.org/10.1371/journal.ppat.1002693
47. Bubendorfer S, Krebes J, Yang I, Hage E, Schulz TF, Bahlawane C, Didelot X, Suerbaum S. 2016. Genome-wide analysis of chromosomal import patterns after natural transformation of *Helicobacter pylori*. Nat Commun 7:11995. https://doi.org/10.1038/ncomms11995
48. Roncarati D, Pelliciari S, Doniselli N, Maggi S, Vannini A, Valzania L, Mazzei L, Zambelli B, Rivetti C, Danielli A. 2016. Metal-responsive promoter DNA compaction by the ferric uptake regulator. Nat Commun 7:12593. https://doi.org/10.1038/ncomms12593
49. Agriesti F, Roncarati D, Musiani F, Del Campo C, Iurlaro M, Sparla F, Ciurli S, Danielli A, Scarlato V. 2014. FeON-FeOFF: the *Helicobacter pylori* Fur regulator commutates iron-responsive transcription by discriminative readout of opposed DNA grooves. Nucleic Acids Res 42:3138–3151. https://doi.org/10.1093/nar/gkt1258
50. Pettersson A, Maas A, van Wassenaar D, van der Ley P, Tommassen J. 1995. Molecular characterization of FrpB, the 70-kilodalton iron-regulated outer membrane protein of *Neisseria meningitidis*. Infect Immun 63:4181–4184. https://doi.org/10.1128/iai.63.10.4181-4184.1995
51. Danielli A, Romagnoli S, Roncarati D, Costantino L, Delany I, Scarlato V. 2009. Growth phase and metal-dependent transcriptional regulation of the *fecA* genes in *Helicobacter pylori*. J Bacteriol 191:3717–3725. https://doi.org/10.1128/JB.01741-08
52. Browning DF, Busby SJ. 2004. The regulation of bacterial transcription initiation. Nat Rev Microbiol 2:57–65. https://doi.org/10.1038/nrmicro787
53. Pich OQ, Merrell DS. 2013. The ferric uptake regulator of *Helicobacter pylori*: a critical player in the battle for iron and colonization of the stomach. Future Microbiol 8:725–738. https://doi.org/10.2217/fmb.13.43
54. Kinoshita-Daitoku R, Kiga K, Miyakoshi M, Otsubo R, Ogura Y, Sanada T, Bo Z, Phuoc TV, Okano T, Iida T, et al. 2021. A bacterial small RNA regulates the adaptation of *Helicobacter pylori* to the host environment. Nat Commun 12:2085. https://doi.org/10.1038/s41467-021-22317-7
55. Pernitzsch SR, Alzheimer M, Bremer BU, Robbe-Saule M, De Reuse H, Sharma CM. 2021. Small RNA mediated gradual control of lipopolysaccharide biosynthesis affects antibiotic resistance in *Helicobacter pylori*. Nat Commun 12:4433. https://doi.org/10.1038/s41467-021-24689-2
56. Nell S, Estibariz I, Krebes J, Bunk B, Graham DY, Overmann J, Song Y, Spröer C, Yang I, Wex T, Korlach J, Malfertheiner P, Suerbaum S. 2018. Genome and methylome variation in *Helicobacter pylori* with a *cag* Pathogenicity Island during early stages of human infection. Gastroenterology 154:612–623. https://doi.org/10.1053/j.gastro.2017.10.014
57. Ailloud F, Didelot X, Woltemate S, Pfaffinger G, Overmann J, Bader RC, Schulz C, Malfertheiner P, Suerbaum S. 2019. Within-host evolution of *Helicobacter pylori* shaped by niche-specific adaptation, intragastric migrations and selective sweeps. Nat Commun 10:2273. https://doi.org/10.1038/s41467-019-10050-1
58. Estibariz I, Ailloud F, Woltemate S, Bunk B, Spröer C, Overmann J, Aebischer T, Meyer TF, Josenhans C, Suerbaum S. 2020. *In vivo* genome and methylome adaptation of *cag*-negative *Helicobacter pylori* during experimental human infection. mBio 11:e01803-20. https://doi.org/10.1128/mBio.01803-20
59. Linz B, Windsor HM, McGraw JJ, Hansen LM, Gajewski JP, Tomsho LP, Hake CM, Solnick JV, Schuster SC, Marshall BJ. 2014. A mutation burst during the acute phase of *Helicobacter pylori* infection in humans and rhesus macaques. Nat Commun 5:4165. https://doi.org/10.1038/ncomms5165
60. Furuta Y, Namba-Fukuyo H, Shibata TF, Nishiyama T, Shigenobu S, Suzuki Y, Sugano S, Hasebe M, Kobayashi I. 2014. Methylome diversification through changes in DNA methyltransferase sequence specificity. PLoS Genet 10:e1004272. https://doi.org/10.1371/journal.pgen.1004272
61. Draper JL, Hansen LM, Bernick DL, Abedrabbo S, Underwood JG, Kong N, Huang BC, Weis AM, Weimer BC, van Vliet AHM, Pourmand N, Solnick JV, Karplus K, Ottemann KM. 2017. Fallacy of the unique genome: sequence diversity within single *Helicobacter pylori* strains. mBio 8:e02321-16. https://doi.org/10.1128/mBio.02321-16
62. Baltrus DA, Amieva MR, Covacci A, Lowe TM, Merrell DS, Ottemann KM, Stein M, Salama NR, Guillemin K. 2009. The complete genome sequence of *Helicobacter pylori* strain G27. J Bacteriol 191:447–448. https://doi.org/10.1128/JB.01416-08
63. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
64. Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R, Stadler PF, Vogel J. 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature 464:250–255. https://doi.org/10.1038/nature08756
65. Bushnell B, Rood J, Singer E. 2017. BBMerge - Accurate paired shotgun read merging via overlap. PLoS One 12:e0185056. https://doi.org/10.1371/journal.pone.0185056
66. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340
67. Stamatakis A. 2015. Using RAxML to infer phylogenies. Curr Protoc Bioinform 51:14. https://doi.org/10.1002/0471250953.bi0614s51
68. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, et al. 2015. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods 12:115–121. https://doi.org/10.1038/nmeth.3252
69. Scrucca L, Fop M, Murphy TB, Raftery AE. 2016. Mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. R J 8:289–317.
70. Sobetzko P, Jelonek L, Strickert M, Han W, Goesmann A, Waldminghaus T. 2016. DistAMo: a web-based tool to characterize DNA-motif distribution on bacterial chromosomes. Front Microbiol 7:283. https://doi.org/10.3389/fmicb.2016.00283
71. Burge C, Campbell AM, Karlin S. 1992. Over- and under-representation of short oligonucleotides in DNA sequences. Proc Natl Acad Sci U S A 89:1358–1362. https://doi.org/10.1073/pnas.89.4.1358
72. Pevzner PA, Borodovsky MA, Mironov AA. 1989. Linguistics of nucleotide sequences. I: the significance of deviations from mean statistical characteristics and prediction of the frequencies of occurrence of words. J Biomol Struct Dyn 6:1013–1026. https://doi.org/10.1080/07391102.1989.10506528
73. Rusinov IS, Ershova AS, Karyagina AS, Spirin SA, Alexeevski AV. 2018. Comparison of methods of detection of exceptional sequences in prokaryotic genomes. Biochemistry (Moscow) 83:129–139. https://doi.org/10.1134/S0006297918020050
74. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8
75. Dalia AB, McDonough E, Camilli A. 2014. Multiplex genome editing by natural transformation. Proc Natl Acad Sci U S A 111:8937–8942. https://doi.org/10.1073/pnas.1406478111
76. Silva D, Santos G, Barroca M, Costa D, Collins T. 2023. Inverse PCR for site-directed mutagenesis. Methods Mol Biol 2967:223–238. https://doi.org/10.1007/978-1-0716-3358-8_18
77. Huang S, Kang J, Blaser MJ. 2006. Antimutator role of the DNA glycosylase *mutY* gene in *Helicobacter pylori*. J Bacteriol 188:6224–6234. https://doi.org/10.1128/JB.00477-06
78. Stein SC, Faber E, Bats SH, Murillo T, Speidel Y, Coombs N, 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. https://doi.org/10.1371/journal.ppat.1006514

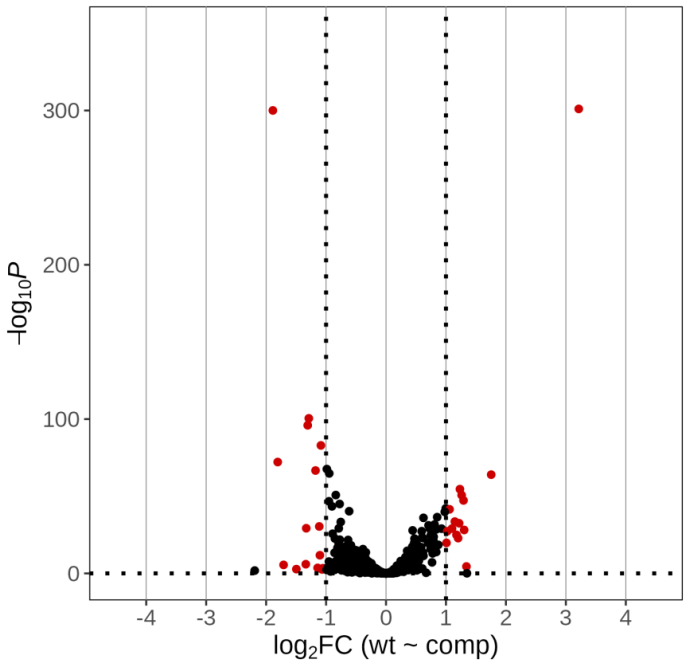


Figure S1. Transcriptomic effect of M.Hpy99XIX complementation in *H. pylori* strain J99. Volcano plot of *H. pylori* J99 wild-type compared to *H. pylori* J99 M.Hpy99XIX complementation strain. A negative and positive Log_2 fold change indicate a gene downregulated or upregulated in the complementation strain compared to the wild-type, respectively. Genes with an adjusted p-value < 0.01 and an absolute log_2 fold change > 1 are shown in red.

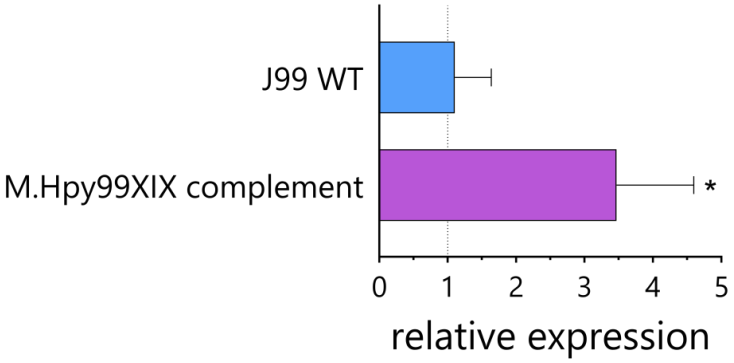


Figure S2. Expression of the M.Hpy99XIX methyltransferase in the complementation mutant. qPCR analysis of the expression of M.Hpy99XIX in the complementation mutant compared to the J99 WT. Student's t-test: * $p<0.05$. Error bars indicate the standard-deviation of 2-3 replicates.

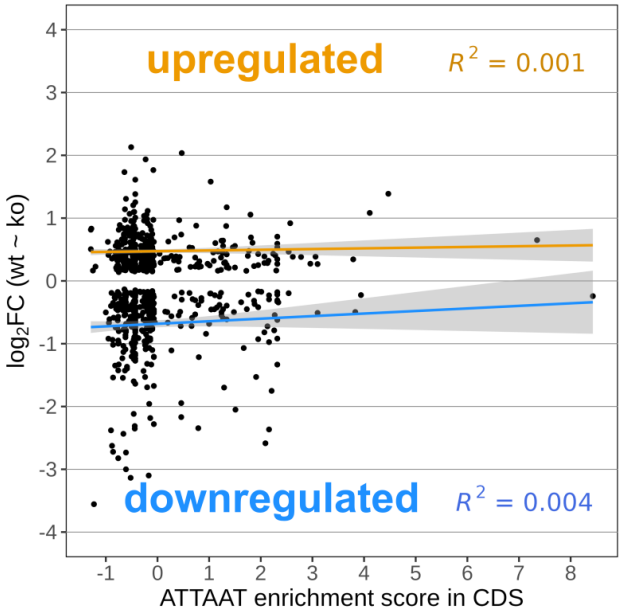


Figure S3. Association of ATTAAT motifs located in coding sequences (CDS) with transcriptomic changes in the M.Hpy99XIX KO strain. Enrichment score of the ATTAAT motif in CDS (calculated using the DistAMo algorithm, see Methods) is plotted on the x-axis again the Log_2FC of differentially expressed genes (J99 WT versus M.Hpy99XIX KO) on the y-axis. A linear regression was performed separately for up-regulated genes (Log_2FC > 0, orange line) and downregulated genes (Log_2FC < 0, blue line). Confidence intervals are shown as a grey area for each regression. The coefficient of determinations (R-squared) are indicated for each regression and, in both cases, suggest an absence of correlation between the enrichment of the ATTAAT motifs in CDS and the transcriptomic changes observed in the M.Hpy99XIX KO strain.

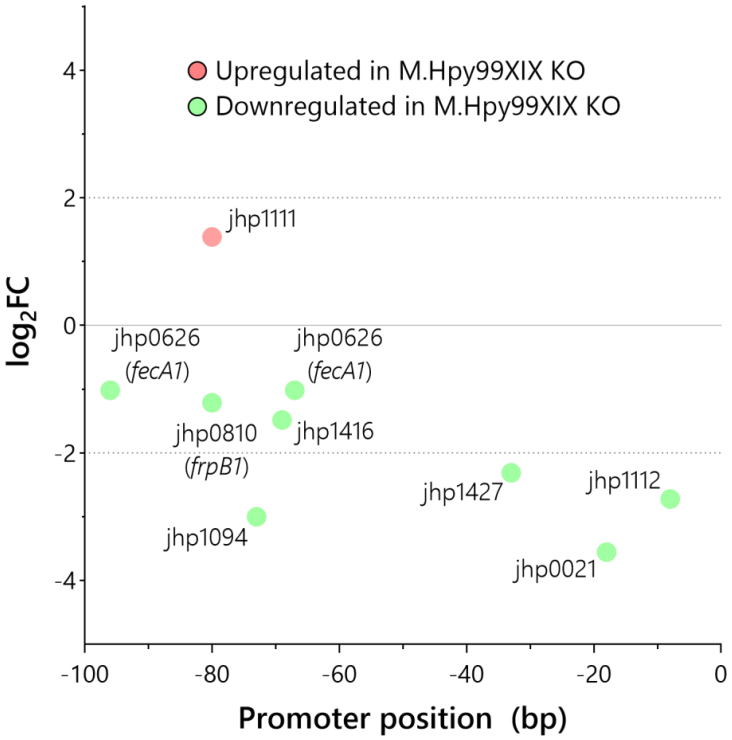


Figure S4. Characterization of the motif position in promoter regions of differentially expressed genes. The position of the first ATTAAT motif in the promoter region defined as 100 bp upstream of the TSS is displayed for the 8 differentially expressed genes according to their $\log_2$ fold change.

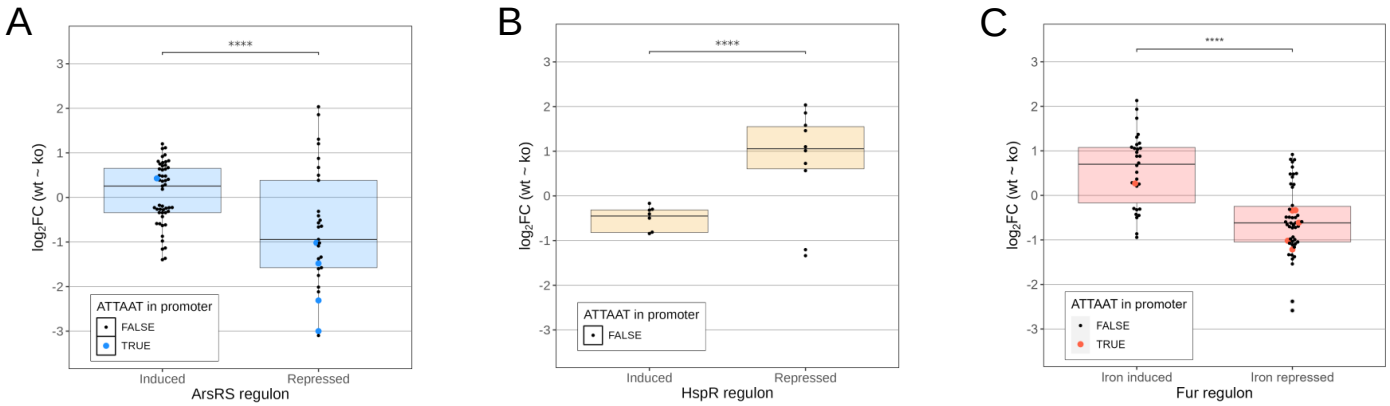


Figure S5. Comparison between M.Hpy99XIX transcriptional effects and environment responsive regulons. A. Comparison of RNA-Seq expression data of M.Hpy99XIX KO against WT with the ArsRS-responsive regulon (Loh et al. 2021) split into induced and repressed genes. Student's t-test: *** p<0.001, **** p<0.00001. **B.** Comparison of RNA-Seq expression data of M.Hpy99XIX KO against WT with the HspR-responsive regulon (Roncarati et al. 2007) split into induced and repressed genes. Student's t-test: *** p<0.001, **** p<0.00001. **C.** Comparison of RNA-Seq expression data of M.Hpy99XIX KO against WT with the Fur-responsive regulon (Vannini et al. 2024) split into iron induced and repressed genes. Student's t-test: *** p<0.001, **** p<0.00001. Log_2 fold-changes of significantly differentially expressed genes (adjusted p-value < 0.01) are represented by individual points and by a boxplot. The presence of an ATTAAT motif within -100 bp of the TSS of each gene is indicated by a larger coloured point.

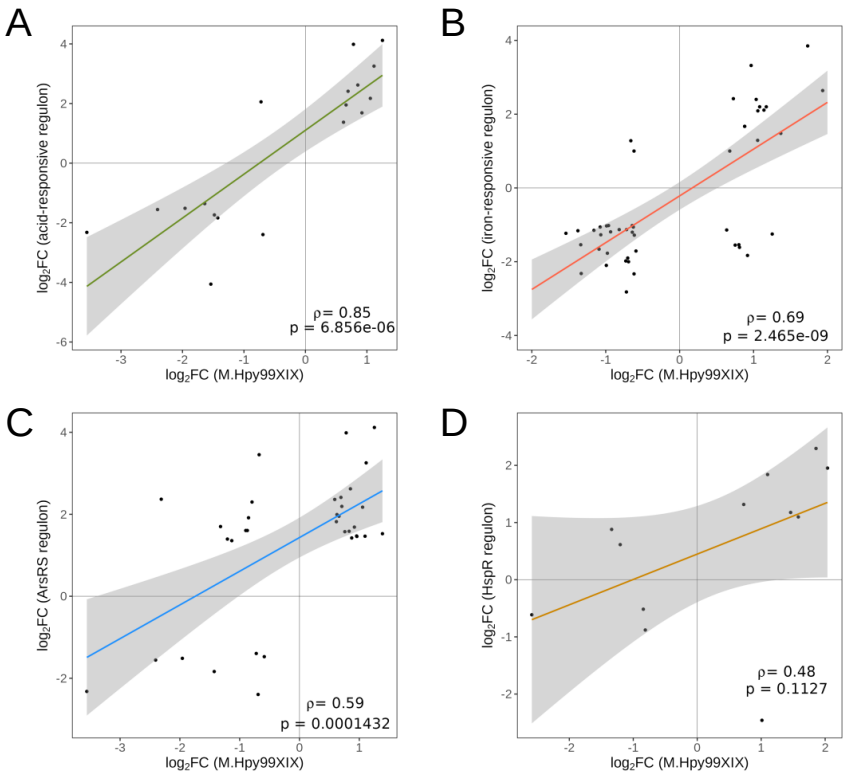


Figure S6. Association between M.Hpy99XIX transcriptional effects and environment responsive regulons. Associations between genes with >|50%| differential expression (|$\log_2$ fold-change| > $\log_2$(1.5) and p.adj < 0.01) observed in the M.Hpy99XIX KO strain and in different regulons were calculated using the Pearson's correlation coefficient (indicated by ρ and p-value indicated below). A linear regression with 95% confidence interval is also shown. **A.** Acid-responsive regulon. **B.** Iron-responsive regulon. **C.** ArsRS regulon. **D.** HspR regulon.

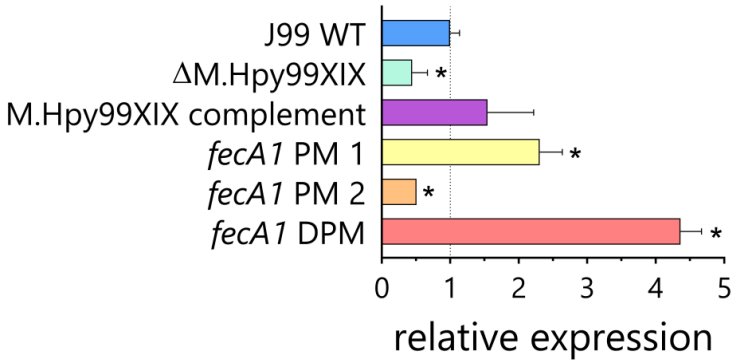


Figure S7. Characterization of the effect of ATTAAT methylation within the promoter region of *fecA1.* Comparison of *fecA1* expression in the M.Hpy99XIX knockout, the point-mutants (PM) and double point-mutant (DPM) to the wild-type strain by qPCR. Student's t-test: * $p<0.05$. Error bars indicate the standard-deviation of 2-3 replicates.

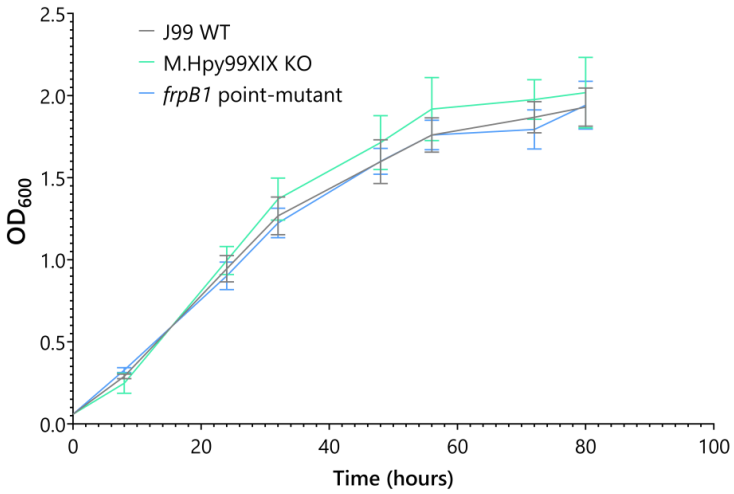


Figure S8. Characterization of the effect of global (M.Hpy99XIX KO) and local (*frpB1* point mutant) ATTAAT methylation on growth of *H. pylori* strain J99. Comparison of growth in liquid culture between the M.Hpy99XIX knockout and the *frpB1* point mutant over a time course of 80 hours. Error bars indicate the standard deviation of 5 replicates.

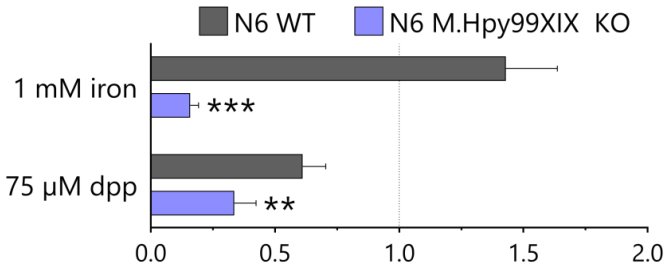


Figure S9. Impact of ATTAAT methylation on growth under environmental conditions related to iron metabolism. Growth of WT and mutants under iron, iron chelation (dpp) conditions was normalised to growth without treatment and compared after 20 hours. Student's t-test: * p<0.05, ** p<0.01, *** p<0.001, **** p<0.0001. Error bars indicate the standard deviation of 5 replicates.

4. Chapter 4 Discussion

H. pylori exhibits a high genetic diversity, both across its phylogeographic population structure and within the gastric niche (Ailloud et al., 2021; Suerbaum & Josenhans, 2007). Although several methylomes and methyltransferases of *H. pylori* had been characterized prior to this thesis, a comprehensive overview of *H. pylori*'s epigenetic landscape was still lacking, making it difficult to model epigenetic diversity across phylogeographic populations. Furthermore, while some methyltransferases (MTases) in *H. pylori* were known to regulate gene expression, the extent and mechanisms of their broader transcriptomic effects – particularly indirect or network-level regulation – remained unclear. Manuscript I examined MTase distribution, the variability of target motifs, and the evolution of motif patterns in *H. pylori*. In manuscript II, we investigated the function of the strictly orphan MTase M.Hpy99XIX and its role in iron homeostasis.

4.1 Methylome evolution and selective pressures

4.1.1 Influence of phylogeography on the methylome

Our analysis of the frequency of MTases and the abundance of their cognate motifs in a large collection of *H. pylori* genomes demonstrated a strong influence of the phylogeographic structure. This pattern is consistent with the highly diverse phylogeographic structure of *H. pylori*, which is a result of the long-term co-evolution with human populations (Falush et al., 2003; Moodley et al., 2012). Motifs with a low density demonstrated an increased degree of genomic instability, suggesting that they are subject to selective pressures that result in their removal. This could be explained by motif avoidance, wherein restriction sites are removed to prevent self-restriction while keeping the restriction endonuclease (REase) active (Loenen et al., 2014; Murray, 2000). The classical function of REases is in the protection against phages, which *H. pylori* has been shown to carry (Munoz et al., 2020). These prophages show a related phylogeography to *H. pylori* population structure (Vale & Lehours, 2018). This suggests that there may be a connection between the prophage phylogeography, the MTase repertoire, and motif density. In type II RM systems, motif avoidance is more pronounced, likely due to their fixed TRDs and greater diversity (Callens et al., 2021). The type II MTases were found to be the most conserved among the different types of MTases. This may be linked to the separation of MTase and REase functions and their genetic structure. The loss of the MTase would lead to post-segregational killing due to the remaining REase activity (Kobayashi, 2001; Naito et al., 1995).

4.1.2 Selective pressures on motifs and regulatory implications

We observed both positive and negative correlations between the MTase presence and the cognate motif. Here, the host immune system may exert an influence on the selection pressure to deplete or enrich motifs by activating immune responses in macrophages and dendritic cells via the cGAS-STING pathway (Balzarolo et al., 2021). We observed in manuscript I that m5C motifs were more likely to mutate due to spontaneous deamination, resulting in non-random patterns of distribution of these motifs. This suggests that some motifs may be selectively maintained due to beneficial roles, whereas others may be avoided or lost due to negative effects.

It has been demonstrated that over time, local environmental pressures and host interactions shape the genome and epigenome (Munoz-Ramirez et al., 2021; Tourrette et al., 2024). It is possible that by gaining or losing motifs, the effects of methylation are fine-tuned to attain a fitness advantage. This is supported by the observation in manuscript II, where individual motifs were shown to contribute to larger downstream effects affecting metal sensitivity and oxidative stress tolerance. In another study, individual GCGC motifs were shown to significantly affect gene expression in *H. pylori* (Estibariz et al., 2019; Meng et al., 2021). Combined, GCGC methylation was shown to affect cell adherence and copper resistance (Estibariz et al., 2019). Phase-variable methylation may also influence immune evasion by regulating the expression of outer membrane proteins and, in turn, cell adhesion and host interaction (Srikhanta et al., 2011).

More broadly, this thesis suggests that DNA methylation patterns in *H. pylori* evolve under selective pressures from both the environment and the host, driving the gain or loss of particular motifs in a lineage-specific manner over time. Exploring this epigenetic plasticity across the full diversity of *H. pylori*, with an analysis of the HpGP dataset (Thorell et al., 2023), could uncover novel motifs and MTases with regulatory or ecological significance.

4.1.3 Methyltransferase evolution, ecology, and environmental integration

While the functional assays in manuscript II showed that methylation altered gene expression and the bacterial phenotype, the population structure was reflected in the MTase composition and motif abundance in manuscript I. The ACGT motif showed the highest variation in motif density between populations but showed no correlation to the frequency of the MTase. Demographic bottlenecks and the rapid evolution of *H. pylori* following repeated human migration events may have led to this phenomenon by limiting *H. pylori*

diversity and driving population-specific changes in motifs (Thorpe et al., 2022). The recently described Hardy ecospecies, which has been suggested to predate all previously known *H. pylori* populations, exemplifies how environmental factors, such as human diet, can drive evolution (Tourrette et al., 2024). Hardy strains lack the *cagPAI*, and many genes involved in iron uptake and regulation are mutated in Hardy strains compared to Ubiquitous strains (Tourrette et al., 2024). We found that the M.Hpy99XIX MTase was consistently absent from all Hardy strains. This suggests a connection between the acquisition of the M.Hpy99XIX MTase and the evolution of the Ubiquitous ecospecies in regard to ecology, for instance, diet and iron homeostasis. The ecological role of MTases is further supported by the example of the type I MTase HsdM1, whose expression is regulated by the acid-responsive ArsRS system (Zimmerman et al., 2024), highlighting how MTases can integrate environmental signals and influence niche adaptation.

4.2 Orphan methyltransferases and regulation in *H. pylori*

4.2.1 Orphan methyltransferases and the complexity of RM systems in *H. pylori*

H. pylori is a diverse organism, and this diversity is also visible in its large and variable set of RM systems (Ailloud et al., 2021). Previous studies have focused on specific RM systems in *H. pylori* and described additional functions of the MTases aside from methylating DNA. M2.HpyAII was shown to influence multiple factors linked to *H. pylori* pathogenicity (Kumar et al., 2018), and M.Hpy99III regulated the transcription of various genes involved in metabolic pathways, competence, and cell adherence (Estibariz et al., 2019). A recent study found that *H. pylori* MTases are organized in a regulatory network and can modulate the transcription of each other (Yano et al., 2020). Estibariz et al. (2019) observed that for the GCGC-specific Hpy99III RM system, the MTase was highly conserved, while the cognate REase was only found in specific populations, suggesting this may be a result of selective pressures. The REase may have been spontaneously inactivated in other RM systems as well. Therefore, it would be interesting to determine how often RM systems described as complete are actually functionally orphan. However, the technical complexity involved in answering this question presents a considerable challenge. The activity of the methyltransferase can be studied by 3rd generation sequencing methods to reveal the methylome, but the REase would need to be individually analyzed for its activity. Consequently, it is impossible to determine how frequently REases might be inactivated by missense mutations. To study the function of an MTase it is therefore

best to use strictly orphan MTases as a model. Orphan MTases have been shown to regulate different processes in other bacteria, such as virulence, colonization, and sporulation (Oliveira & Fang, 2021), and might play a larger role in *H. pylori* than previously thought.

4.2.2 Transcriptomic effect of the orphan methyltransferase M.Hpy99XIX in *H. pylori*

It was demonstrated in *H. pylori* that MTases can have a larger transcriptomic effect than might be expected from the presence of corresponding target sequences in promoter regions (Estibariz et al., 2019; Yano et al., 2020). Estibariz et al. (2019) showed that only 13 of the 225 differentially expressed genes in J99 had a GCGC motif in the 50 bp region upstream of the TSS; for BCM-300, this was the case for 11 out of 29 of the differentially expressed genes. In J99, the cognate REase was predicted to be inactive and missing in BCM-300 (Estibariz et al., 2019), making the MTase in these cases functionally orphan. Only transcription factors have been previously shown to regulate as many genes in *H. pylori* (Vannini et al., 2022). Deletion of the M.Hpy99XIX MTase caused the differential expression of over 100 genes, despite the limited amount of motifs in promoter regions. This suggested an indirect or network-level effect, as seen for transcriptional regulators (Vannini et al., 2022). Several of the genes affected could be traced back to genes of the Fur (Vannini et al., 2024) and ArsRS regulons (Loh et al., 2021). Transcriptomic data is making the delineation of genes responsive to certain conditions easier, as seen in the example of the pH-responsive regulon controlled by ArsRS in *H. pylori* (Loh et al., 2021). Manuscript II suggested a role of the orphan MTase M.Hpy99XIX in iron homeostasis and fine-tuning regulatory networks, so other orphan MTases should be studied to find other potential roles.

4.2.3 Mechanisms of methylation-driven regulation and phenotypic effects in *H. pylori*

GCGC methylation had a large impact on transcription, and a single motif within the promoter region showed a direct effect (Estibariz et al., 2019). The GCGC motif also showed a strong bias for the -13 position relative to the transcriptional start site (TSS) in a later study (Meng et al., 2021). Methylation of ATTAAT motifs in the promoter regions of *frpB1* and *fecA1* had a direct effect on transcription as well. The change in methylation status of the promoter regions of *frpB1* and *fecA1* also led to phenotypic changes in metal

ion sensitivity and resistance to oxidative stress. These downstream effects could potentially affect other regulatory networks, possibly small RNAs and transcription factors (Kinoshita-Daitoku et al., 2021; Pernitzsch et al., 2021).

The ATTAAT motifs in the promoter regions of *frpB1* and *fecA1* are highly conserved, suggesting an important role in this regulatory process. The recently developed method of EM-seq, together with direct high-throughput sequencing-based methods (Patel et al., 2024), could be used to study whether these motifs are always methylated. Methylation patterns also vary between individual cells, creating different subpopulations, each with a distinct phenotypic profile, which may enhance the survival of the population by this bet-hedging strategy (Veening et al., 2008). Single-cell sequencing has been used to identify rare populations in *E. coli* (Wang et al., 2023) and could be used to provide new insights into how methylation and transcription vary between subpopulations in *H. pylori*. Methylation also alters the structure of the DNA, which in turn affects which regions of the DNA are accessible for the transcriptional machinery (Beaulaurier et al., 2019) and may also result in phenotypic differences. Hi-C is a new method being used to describe the chromosomal structure for multiple bacteria (Bohm et al., 2020; Le et al., 2013; Lioy et al., 2018; Marbouty et al., 2015) and could also provide insights into how methylation affects the DNA structure in *H. pylori*.

4.3 Future Directions

4.3.1 Biological and evolutionary implications of M.Hpy99XIX

The findings of this thesis confirm the double role of MTases in H. pylori, both as components of restriction-modification systems and as dynamic and multifunctional regulators of gene expression and genome evolution. The orphan MTase M.Hpy99XIX appears to have a unique role, with significant influence in both *H. pylori* infection and adaptation to different environments. Manuscript II showed that the presence of M.Hpy99XIX is sufficient to distinguish Ubiquitous from Hardy strains. Its complete absence in the Hardy ecospecies raises the possibility that M.Hpy99XIX could serve as a lineage-specific biomarker or even a virulence-modulating factor, depending on how the pathotype status of Hardy strains is further defined, since they were isolated from populations with a significant gastric disease risk (Tourrette et al., 2024). Exploring whether the acquisition of M.Hpy99XIX is connected to the acquisition of the *cagPAI*, which is missing in the Hardy strains as well as in the hpAfrica2 population, may elucidate the evolution of pathogenic

traits in Ubiquitous lineages (Tourrette et al., 2024). The *cagPAI* encodes a type IV secretion system that translocates the CagA protein (Odenbreit et al., 2000), and iron deficiency was shown to enhance CagA translocation (Noto & Peek, 2015), leading to accelerated carcinogenesis (Noto et al., 2013). Therefore, it would be particularly interesting to determine how iron homeostasis was affected in the common ancestor of Hardy and Ubiquitous strains after acquiring the M.Hpy99XIX MTase.

4.3.2 Mechanistic hypotheses and epigenetic regulation in *H. pylori*

Methylation has been shown to have a direct effect on upregulating genes in *H. pylori* (Estibariz et al., 2019; Srikhanta et al., 2017). However, how this is mechanistically possible has not yet been described. In *E. coli*, Fur has been shown to compete with Dam for binding to the promoter of *sci1* (Brunet et al., 2020). This could also be the case in *H. pylori*. To study whether there is competition between Fur and M.Hpy99XIX at the promoter regions of *frpB1* and *fecA1*, as described in manuscript II, could reveal methylation-dependent regulatory mechanisms that extend beyond the canonical promoter repression or activation. Additionally, other promoters may be subject to the same mechanism of regulation in *H. pylori* that are yet to be found. In *H. pylori*, the DNA conformation has been shown to play a regulatory role (Ye et al., 2007), which may also be influenced by methylation, as discussed previously. Finally, it would be interesting to identify methylation-sensitive regulators in H. *pylori* to determine if this could be a potential mechanism.

4.3.3 Translational potential and therapeutic applications

Current treatment for a *H. pylori* infection typically consists of a combination of antibiotics with bismuth (Malfertheiner et al., 2023). However, the increase in antibiotic resistance necessitates testing for antibiotic susceptibilities (Salahi-Niri et al., 2024). As of today, there is no vaccine that has been approved for human use, although many studies have been conducted (Gong et al., 2025). From a translational perspective, MTases are increasingly recognized as targets for antimicrobial development, especially given the limitations of bactericidal therapies. Recent advances in therapeutics, such as nanoparticle-based targeted delivery systems (Chitas et al., 2024) and probiotic adjunct therapies (Luo et al., 2023), highlight promising alternative approaches. Additionally, the growing interest in transcriptional regulators as antibiotic targets (Vannini et al., 2022) suggests that MTases like M.Hpy99XIX, due to its high conservation, could be leveraged as epigenetic targets in future therapeutic strategies.

Chapter 5 References

Ailloud, F., Estibariz, I., & Suerbaum, S. (2021). Evolved to vary: genome and epigenome variation in the human pathogen *Helicobacter pylori*. *FEMS Microbiol Rev*, *45*(1).

Balzarolo, M., Engels, S., de Jong, A. J., Franke, K., van den Berg, T. K., Gulen, M. F., Ablasser, A., Janssen, E. M., van Steensel, B., & Wolkers, M. C. (2021). m6A methylation potentiates cytosolic dsDNA recognition in a sequence-specific manner. *Open Biol*, *11*(3), 210030.

Beaulaurier, J., Schadt, E. E., & Fang, G. (2019). Deciphering bacterial epigenomes using modern sequencing technologies. *Nat. Rev. Genet*, *20*, 157-172.

Bjork, G. R., Wikstrom, P. M., & Bystrom, A. S. (1989). Prevention of translational frameshifting by the modified nucleoside 1-methylguanosine. *Science*, *244*(4907), 986-989.

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

Bohm, K., Giacomelli, G., Schmidt, A., Imhof, A., Koszul, R., Marbouty, M., & Bramkamp, M. (2020). Chromosome organization by a conserved condensin-ParB system in the actinobacterium *Corynebacterium glutamicum*. *Nat Commun*, *11*(1), 1485.

Boye, E., & Lobner-Olesen, A. (1990). The role of dam methyltransferase in the control of DNA replication in *E. coli*. *Cell*, *62*(5), 981-989.

Braaten, B. A., Nou, X., Kaltenbach, L. S., & Low, D. A. (1994). Methylation patterns in pap regulatory DNA control pyelonephritis-associated pili phase variation in *E. coli*. *Cell*, *76*(3), 577-588.

Brunet, Y. R., Bernard, C. S., & Cascales, E. (2020). Fur-Dam Regulatory Interplay at an Internal Promoter of the Enteroaggregative *Escherichia coli* Type VI Secretion sci1 Gene Cluster. *J Bacteriol*, *202*(10).

Callens, M., Pradier, L., Finnegan, M., Rose, C., & Bedhomme, S. (2021). Read between the lines: diversity of nontranslational selection pressures on local codon usage. *Genome Biol Evol*, *13*(9).

Chitas, R., Fonseca, D. R., Parreira, P., & Martins, M. C. L. (2024). Targeted nanotherapeutics for the treatment of *Helicobacter pylori* infection. *J Biomed Sci*, *31*(1), 78.

Denic, M., Turlin, E., Michel, V., Fischer, F., Khorasani-Motlagh, M., Zamble, D., Vinella, D., & de Reuse, H. (2021). A novel mode of control of nickel uptake by a multifunctional metallochaperone. *PLoS Pathog*, *17*(1), e1009193.

Denic, M., Turlin, E., Zamble, D. B., Betton, J. M., Vinella, D., & De Reuse, H. (2023). The SlyD metallochaperone targets iron-sulfur biogenesis pathways and the TCA cycle. *mBio*, *14*(5), e0096723.

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

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

Ernst, F. D., Bereswill, S., Waidner, B., Stoof, J., Mader, U., Kusters, J. G., Kuipers, E. J., Kist, M., van Vliet, A. H. M., & Homuth, G. (2005). Transcriptional profiling of *Helicobacter pylori* Fur- and iron-regulated gene expression. *Microbiology (Reading)*, *151*(Pt 2), 533-546.

Estibariz, I., Overmann, A., Ailloud, F., Krebes, J., Josenhans, C., & Suerbaum, S. (2019). The core genome ^{m5}C methyltransferase JHP1050 (M.Hpy99III) plays an important role in orchestrating gene expression in *Helicobacter pylori*. *Nucleic Acids Res*, *47*(5), 2336–2348.

Faass, L., Hauke, M., Stein, S. C., & Josenhans, C. (2023). Innate immune activation and modulatory factors of *Helicobacter pylori* towards phagocytic and nonphagocytic cells. *Curr Opin Immunol*, *82*, 102301.

Falush, D., Kraft, C., Taylor, N. S., Correa, P., Fox, J. G., Achtman, M., & Suerbaum, S. (2001). Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. *Proc Natl Acad Sci USA*, *98*(26), 15056-15061.

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., Yamaoka, Y., Megraud, F., Otto, K., Reichard, U., Katzowitsch, E., Wang, X., Achtman, M., & Suerbaum, S. (2003). Traces of human migrations in *Helicobacter pylori* populations. *Science*, *299*(5612), 1582-1585.

Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., Korlach, J., & Turner, S. W. (2010). Direct detection of DNA methylation during Single-Molecule, Real-Time sequencing. *Nat. Methods*, *7*(6), 461-465.

Fox, K. L., Dowideit, S. J., Erwin, A. L., Srikhanta, Y. N., Smith, A. L., & 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*(15), 5242-5252.

Gancz, H., Censini, S., & Merrell, D. S. (2006). Iron and pH homeostasis intersect at the level of Fur regulation in the gastric pathogen *Helicobacter pylori*. *Infect Immun*, *74*(1), 602-614.

Gong, J., Wang, Q., Chen, X., & Lu, J. (2025). *Helicobacter pylori* Vaccine: Mechanism of Pathogenesis, Immune Evasion and Analysis of Vaccine Types. *Vaccines (Basel)*, *13*(5).

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

Han, S., Liu, J., Li, M., Zhang, Y., Duan, X., Zhang, Y., Chen, H., Cai, Z., Yang, L., & Liu, Y. (2022). DNA Methyltransferase Regulates Nitric Oxide Homeostasis and Virulence in a Chronically Adapted *Pseudomonas aeruginosa* Strain. *mSystems*, *7*(5), e0043422.

Heithoff, D. M., Sinsheimer, R. L., Low, D. A., & Mahan, M. J. (1999). An essential role for DNA adenine methylation in bacterial virulence. *Science*, *284*(5416), 967-970.

Hoernes, T. P., Clementi, N., Faserl, K., Glasner, H., Breuker, K., Lindner, H., Huttenhofer, A., & Erlacher, M. D. (2016). Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. *Nucleic Acids Res*, *44*(2), 852-862.

Husson, M. O., Legrand, D., Spik, G., & Leclerc, H. (1993). Iron acquisition by *Helicobacter pylori*: importance of human lactoferrin. *Infect Immun*, *61*(6), 2694-2697.

Kim, H., Kim, J. H., Cho, H., & Ko, K. S. (2022). Overexpression of a DNA Methyltransferase Increases Persister Cell Formation in *Acinetobacter baumannii*. *Microbiol Spectr*, *10*(6), e0265522.

Kinoshita-Daitoku, R., Kiga, K., Miyakoshi, M., Otsubo, R., Ogura, Y., Sanada, T., Bo, Z., Phuoc, T. V., Okano, T., Iida, T., Yokomori, R., Kuroda, E., Hirukawa, S., Tanaka, M., Sood, A., Subsomwong, P., Ashida, H., Binh, T. T., Nguyen, L. T., Van, K. V., Ho, D. Q. D., Nakai, K., Suzuki, T., Yamaoka, Y., Hayashi, T., & Mimuro, H. (2021). A bacterial small RNA regulates the adaptation of *Helicobacter pylori* to the host environment. *Nat Commun*, *12*(1), 2085.

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.

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

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

Le, T. B., Imakaev, M. V., Mirny, L. A., & Laub, M. T. (2013). High-resolution mapping of the spatial organization of a bacterial chromosome. *Science*, *342*(6159), 731-734.

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

Lioy, V. S., Cournac, A., Marbouty, M., Duigou, S., Mozziconacci, J., Espeli, O., Boccard, F., & Koszul, R. (2018). Multiscale Structuring of the *E. coli* Chromosome by Nucleoid-Associated and Condensin Proteins. *Cell*, *172*(4), 771-783 e718.

Loenen, W. A., Dryden, D. T., Raleigh, E. A., Wilson, G. G., & Murray, N. E. (2014). Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res*, *42*(1), 3-19.

Loh, J. T., Shum, M. V., Jossart, S. D. R., Campbell, A. M., Sawhney, N., McDonald, W. H., Scholz, M. B., McClain, M. S., Forsyth, M. H., & Cover, T. L. (2021). Delineation of the pH-Responsive Regulon Controlled by the *Helicobacter pylori* ArsRS Two-Component System. *Infect Immun*, *89*(4).

Luo, Q., Liu, N., Pu, S., Zhuang, Z., Gong, H., & Zhang, D. (2023). A review on the research progress on non-pharmacological therapy of *Helicobacter pylori*. *Front Microbiol*, *14*, 1134254.

Malfertheiner, P., Camargo, M. C., El-Omar, E., Liou, J. M., Peek, R., Schulz, C., Smith, S. I., & Suerbaum, S. (2023). *Helicobacter pylori* infection. *Nat Rev Dis Primers*, *9*(1), 19.

Marbouty, M., Le Gall, A., Cattoni, D. I., Cournac, A., Koh, A., Fiche, J. B., Mozziconacci, J., Murray, H., Koszul, R., & Nollmann, M. (2015). Condensin- and Replication-Mediated Bacterial Chromosome Folding and Origin Condensation Revealed by Hi-C and Super-resolution Imaging. *Mol Cell*, *59*(4), 588-602.

Meng, B., Epp, N., Wijaya, W., Mrazek, J., & Hoover, T. R. (2021). Methylation motifs in promoter sequences may contribute to the maintenance of a conserved (m5)C methyltransferase in *Helicobacter pylori*. *Microorganisms*, *9*(12).

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

Munoz-Ramirez, Z. Y., Pascoe, B., Mendez-Tenorio, A., Mourkas, E., Sandoval-Motta, S., Perez-Perez, G., Morgan, D. R., Dominguez, R. L., Ortiz-Princz, D., Cavazza, M. E., Rocha, G., Queiroz, D. M. M., Catalano, M., Palma, G. Z., Goldman, C. G., Venegas, A., Alarcon, T., Oleastro, M., Vale, F. F., Goodman, K. J., Torres, R. C., Berthenet, E., Hitchings, M. D., Blaser, M. J., Sheppard, S. K., Thorell, K., & Torres, J. (2021). A 500-year tale of co-evolution, adaptation, and virulence: *Helicobacter pylori* in the Americas. *ISME J*, *15*(1), 78-92.

Munoz, A. B., Stepanian, J., Trespalacios, A. A., & Vale, F. F. (2020). Bacteriophages of *Helicobacter pylori*. *Front Microbiol*, *11*, 549084.

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.

Naito, T., Kusano, K., & Kobayashi, I. (1995). Selfish behavior of Restriction-Modification systems. *Science*, *267*(5199), 897-899.

Noto, J. M., Gaddy, J. A., Lee, J. Y., Piazuelo, M. B., Friedman, D. B., Colvin, D. C., Romero-Gallo, J., Suarez, G., Loh, J., Slaughter, J. C., Tan, S., Morgan, D. R., Wilson, K. T., Bravo, L. E., Correa, P., Cover, T. L., Amieva, M. R., & Peek, R. M., Jr. (2013). Iron deficiency accelerates *Helicobacter pylori*-induced carcinogenesis in rodents and humans. *J Clin Invest*, *123*(1), 479-492.

Noto, J. M., & Peek, R. M., Jr. (2015). *Helicobacter pylori* and CagA under conditions of iron deficiency. *Gut Microbes*, *6*(6), 377-381.

Odenbreit, S., Puls, J., Sedlmaier, B., Gerland, E., Fischer, W., & Haas, R. (2000). Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. *Science*, *287*(5457), 1497-1500. (Not in File)

Oliveira, P. H., & Fang, G. (2021). Conserved DNA Methyltransferases: A Window into Fundamental Mechanisms of Epigenetic Regulation in Bacteria. *Trends Microbiol*, *29*(1), 28-40.

Oliveira, P. H., Ribis, J. W., Garrett, E. M., Trzilova, D., Kim, A., Sekulovic, O., Mead, E. A., Pak, T., Zhu, S., Deikus, G., Touchon, M., Lewis-Sandari, M., Beckford, C., Zeitouni, N. E., Altman, D. R., Webster, E., Oussenko, I., Bunyavanich, S., Aggarwal, A. K., Bashir, A., Patel, G., Wallach, F., Hamula, C., Huprikar, S., Schadt, E. E., Sebra, R., van Bakel, H., Kasarskis, A., Tamayo, R., Shen, A., & Fang, G. (2020). Epigenomic characterization of *Clostridioides difficile* finds a conserved DNA methyltransferase that mediates sporulation and pathogenesis. *Nat Microbiol*, *5*(1), 166-180.

Patel, L., Ailloud, F., Suerbaum, S., & Josenhans, C. (2024). Single-base resolution quantitative genome methylation analysis in the model bacterium *Helicobacter pylori* by enzymatic methyl sequencing (EM-Seq) reveals influence of strain, growth phase, and methyl homeostasis. *BMC Biol*, *22*(1), 125.

Pernitzsch, S. R., Alzheimer, M., Bremer, B. U., Robbe-Saule, M., De Reuse, H., & Sharma, C. M. (2021). Small RNA mediated gradual control of lipopolysaccharide biosynthesis affects antibiotic resistance in *Helicobacter pylori*. *Nat Commun*, *12*(1), 4433.

Pich, O. Q., Carpenter, B. M., Gilbreath, J. J., & Merrell, D. S. (2012). Detailed analysis of *Helicobacter pylori* Fur-regulated promoters reveals a Fur box core sequence and novel Fur-regulated genes. *Mol Microbiol*, *84*(5), 921-941.

Pingoud, A., Fuxreiter, M., Pingoud, V., & Wende, W. (2005). Type II restriction endonucleases: structure and mechanism. *Cell Mol Life Sci*, *62*(6), 685-707.

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

Roberts, R. J., Vincze, T., Posfai, J., & Macelis, D. (2023). REBASE: a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res*, *51*(D1), D629-D630.

Salahi-Niri, A., Nabavi-Rad, A., Monaghan, T. M., Rokkas, T., Doulberis, M., Sadeghi, A., Zali, M. R., Yamaoka, Y., Tacconelli, E., & Yadegar, A. (2024). Global prevalence of *Helicobacter pylori* antibiotic resistance among children in the world health organization regions between 2000 and 2023: a systematic review and meta-analysis. *BMC Med*, *22*(1), 598.

Sanchez-Romero, M. A., & Casadesus, J. (2020). The bacterial epigenome. *Nat Rev Microbiol*, *18*(1), 7-20.

Sergiev, P. V., Serebryakova, M. V., Bogdanov, A. A., & Dontsova, O. A. (2008). The *ybiN* gene of *Escherichia coli* encodes adenine-N6 methyltransferase specific for modification of A1618 of 23

S ribosomal RNA, a methylated residue located close to the ribosomal exit tunnel. *J Mol Biol*, *375*(1), 291-300.

Sharma, C. M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiß, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., Stadler, P. F., & Vogel, J. (2010). The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature*, *464*, 250-255.

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

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

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

Stewart, F. J., Panne, D., Bickle, T. A., & Raleigh, E. A. (2000). Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *J Mol Biol*, *298*(4), 611-622.

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

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

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

Tan, S., Noto, J. M., Romero-Gallo, J., Peek Jr., R. M., & Amieva, M. R. (2011). *Helicobacter pylori* perturbs iron trafficking in the epithelium to grow on the cell surface. *PLoS Pathog*, *7*((5): e1002050).

Thorell, K., Munoz-Ramirez, Z. Y., Wang, D., Sandoval-Motta, S., Boscolo Agostini, R., Ghirotto, S., Torres, R. C., Hp, G. P. R. N., Falush, D., Camargo, M. C., & Rabkin, C. S. (2023). The *Helicobacter pylori* Genome Project: insights into *H. pylori* population structure from analysis of a worldwide collection of complete genomes. *Nat Commun*, *14*(1), 8184.

Thorpe, H. A., Tourrette, E., Yahara, K., Vale, F. F., Liu, S., Oleastro, M., Alarcon, T., Perets, T. T., Latifi-Navid, S., Yamaoka, Y., Martinez-Gonzalez, B., Karayiannis, I., Karamitros, T., Sgouras, D. N., Elamin, W., Pascoe, B., Sheppard, S. K., Ronkainen, J., Aro, P., Engstrand, L., Agreus, L., Suerbaum, S., Thorell, K., & Falush, D. (2022). Repeated out-of-Africa expansions of *Helicobacter pylori* driven by replacement of deleterious mutations. *Nat Commun*, *13*(1), 6842.

Tourrette, E., Torres, R. C., Svensson, S. L., Matsumoto, T., Miftahussurur, M., Fauzia, K. A., Alfaray, R. I., Vilaichone, R. K., Tuan, V. P., Helicobacter Genomics, C., Wang, D., Yadegar, A., Olsson, L. M., Zhou, Z., Yamaoka, Y., Thorell, K., & Falush, D. (2024). An ancient ecospecies of *Helicobacter pylori*. *Nature*.

Vale, F. F., & Lehours, P. (2018). Relating Phage Genomes to *Helicobacter pylori* Population Structure: General Steps Using Whole-Genome Sequencing Data. *Int J Mol Sci*, *19*(7).

Vannini, A., Pinatel, E., Costantini, P. E., Pelliciari, S., Roncarati, D., Puccio, S., De Bellis, G., Scarlato, V., Peano, C., & Danielli, A. (2024). (Re)-definition of the holo- and apo-Fur direct regulons of *Helicobacter pylori*. *J Mol Biol*, *436*(10), 168573.

Vannini, A., Roncarati, D., D'Agostino, F., Antoniciello, F., & Scarlato, V. (2022). Insights into the Orchestration of Gene Transcription Regulators in *Helicobacter pylori*. *Int J Mol Sci*, *23*(22).

Veening, J. W., Smits, W. K., & Kuipers, O. P. (2008). Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol*, *62*, 193-210.

Waidner, B., Greiner, S., Odenbreit, S., Kavermann, H., Velayudhan, J., Stahler, F., Guhl, J., Bisse, E., van Vliet, A. H., Andrews, S. C., Kusters, J. G., Kelly, D. J., Haas, R., Kist, M., & Bereswill, S. (2002). Essential role of ferritin Pfr in *Helicobacter pylori* iron metabolism and gastric colonization. *Infect Immun*, *70*(7), 3923-3929.

Wang, B., Lin, A. E., Yuan, J., Novak, K. E., Koch, M. D., Wingreen, N. S., Adamson, B., & Gitai, Z. (2023). Single-cell massively-parallel multiplexed microbial sequencing (M3-seq) identifies rare bacterial populations and profiles phage infection. *Nat Microbiol*, *8*(10), 1846-1862.

Yamaoka, Y. (2010). Mechanisms of disease: *Helicobacter pylori* virulence factors. *Nat Rev Gastroenterol Hepatol*, *7*(11), 629-641.

Yano, H., Alam, M. Z., Rimbara, E., Shibata, T. F., Fukuyo, M., Furuta, Y., Nishiyama, T., Shigenobu, S., Hasebe, M., Toyoda, A., Suzuki, Y., Sugano, S., Shibayama, K., & Kobayashi, I. (2020). Networking and specificity-changing DNA methyltransferases in *Helicobacter pylori*. *Front Microbiol*, *11*, 1628.

Ye, F., Brauer, T., Niehus, E., Drlica, K., Josenhans, C., & Suerbaum, S. (2007). Flagellar and global gene regulation in *Helicobacter pylori* modulated by changes in DNA supercoiling. *Int. J. Med. Microbiol*, *297*, 65–81.

Zimmerman, E. H., Ramsey, E. L., Hunter, K. E., Villadelgado, S. M., Phillips, C. M., Shipman, R. T., & Forsyth, M. H. (2024). The *Helicobacter pylori* methylome is acid-responsive due to regulation by the two-component system ArsRS and the type I DNA methyltransferase HsdM1 (HP0463). *J Bacteriol*, *206*(1), e0030923.

Chapter 6 Acknowledgements

I would like to express my sincere gratitude to my mentor, **Professor Dr. med. Sebastian Suerbaum**, for giving me the opportunity to begin my journey in his research group and for his continuous guidance and support throughout. His scientific insight, encouragement, and leadership have been a great source of inspiration and growth during my time in the lab.

A heartfelt thank you goes to **Dr. Florent Ailloud**, for his unwavering support, valuable mentorship, and encouragement over the course of this work.

I would also like to thank **Professor Dr. Christine Josenhans**, a member of my Thesis Advisory Committee, for her insightful feedback and support, both academically and through the graduate school.

I am equally grateful to **PD Dr. Christian Schulz**, also a member of my Thesis Advisory Committee, for his constructive feedback and helpful discussions.

I am also deeply grateful to **Gudrun Pfaffinger** and **Evelyn Weiss** for their constant assistance, technical expertise, and patience, which were essential to my progress.

Special thanks to **Annemarie Overmann**, **Lorenzo Apolloni**, and **Fabian Neukirchinger** for sharing part of this journey with me—it was a privilege to work alongside you.

I would also like to thank the members of the **Josenhans group** for the stimulating discussions and memorable celebrations that enriched my experience.

My appreciation extends to the facilities and staff of the **Max von Pettenkofer-Institut** and **Ludwig Maximilians University Munich** for providing a supportive research environment.

Finally, I am incredibly thankful for the love and support of my **family and friends**, whose encouragement and belief in me kept me going through all challenges.

Thank you all!