Symbiosis

ISSN Online: 2372-0956
www.symbiosisonlinepublishing.com

Research Article | SOJ Microbiology & Infectious Diseases | Open Access

Evolutionary dynamics of the *vapD* gene in *Helicobacter pylori* and its wide distribution among bacterial phyla

Gabriela Delgado-Sapién[1], Rene Cerritos-Flores[2], Alejandro Flores-Alanis[1], José L Méndez[1], Alejandro Cravioto[1], Rosario Morales-Espinosa[1*]

*[*1]Laboratorio de Genómica Bacteriana, Departamento de Microbiología y Parasitología, Facultad de Medicina, Universidad Nacional Autónoma de México, Mexico City, México 04510.*

[2]Centro de Investigación en Políticas, Población y Salud, Facultad de Medicina, Universidad Nacional Autónoma de México, Mexico City, México 04510.

Received: 12th August , 2020; Accepted: 15th November 2020 ; Published: 03rd December, 2020

***Corresponding author:** RosarioMorales-Espinosa, PhD, MD, Laboratorio de Genómica Bacteriana, Departamento de Microbiología y Parasitología. Universidad Nacional Autónoma de México. Avenida Universidad 3000, Colonia Ciudad Universitaria, Delegación Coyoacán, C.P. 04510, México City, México.Tel.: +525 5523 2135; Fax: +525 5623 2114 E-mail: marosari@unam.mx*

Abstract

The *vapD* gene is present in microorganisms from different phyla and encodes for the virulence-associated protein D *(*VapD*).* In some microorganisms, it has been suggested that *vapD* participates in either protecting the bacteria from respiratory burst within the macrophage or in facilitating the persistence of the microorganism within the respiratory epithelial cell. The aim of this study was to define the phylogenetic relationship of the *Helicobacter pylori vapD* gene with other *vapD* genes of different bacterial species from different phyla and to estimate the genealogy of *vapD* gene within *H. pylori* species. Sixteen sequences of *Helicobacter pylori vapD* gene obtained from Mexican patients and 211 *vapD* sequences from 72 species of six bacterial phyla were analysed. Our results showed that the *vapD* region is a hot spot that presents a greater diversity in the Mexican strains of *H. pylori* to that previously reported. Rearrangements in the *vapD* region led to the formation of new ORFs in Mexican strains, which were not seen as being fortuitous, suggesting that these chromosomal rearrangements might provide some type of advantage to the bacteria. Phylogenetic analysis, codon usage bias and GC content indicated that *vapD* was acquired by horizontal gene transfer. Then, in some *H. pylori* strains it was incorporated and fixed into the bacterial chromosome and maintained in these strains in a similar fashion as an essential gene. Genealogical analysis of the *H. pylori vapD* gene showed two divisions: one that grouped most of the strains from different parts of the world and the other that grouped only Mexican strains together with the 60190 and 26695 reference strains.

Keywords: *vapD* gene*; Helicobacter pylori;* Genealogy; Phylogeny; Horizontal Gene Transfer.

Introduction

As has been widely documented, *Helicobacter pylori* is a genetically diverse pathogen that colonizes human gastric epithelial cells leading to a variety of gastric diseases in susceptible patients. The clinical outcomes of *H. pylori* infection are diverse, so it is important to understand the genomic variability that enables the microorganism to adapt to the host, whilst at the same time exhibiting a range of potential virulence factors [1,2,3]. Genetic diversity is seen among *H. pylori* strains from different origins and ethnic populations, as well as within *H. pylori* populations within a single stomach. It is well known that *H. pylori* is a highly recombinant microorganism [4-8] and a natural transformant, which explains its genomic variability and diversity that favour a better adaptive capacity and its permanence on the gastric mucosa for decades. The important events that contribute to its adaptive evolution are horizontal gene transfer (HGT) [4,9] and nucleotide insertion-deletion or substitution [10], which result in the polymorphism of individual genes [11]. Additionally, recombination between homologous chromosomal DNA fragments of different *H. pylori* strains [10] and inversion or translocation of large DNA fragments within the same genome lead to chromosomal rearrangements [12,13]. Another distinctive hallmark is the presence of strain-specific genes [14,15], which are found within the plasticity zone where the highest diversity among *H. pylori* strains can be seen [16].

Several virulence factors of *H. pylori* have been well described and in specific populations (Anglo-Saxon) have been shown to have a clear association with gastric pathologies [17-20]. However, in Latin American populations, *H. pylori* infection is often characterized by mixed genotypes without any clear association between a specific genotype and gastric pathology [11,21,22]. It has been shown that the CagA antigen and specific *vacA* (s1a/m1) genotype are the accepted virulence markers for gastric and peptic ulcer disease, as well as for gastric cancer, and the chronicity of infection suggests that some *H. pylori* strains have strategies allowing it to persist inside the cells for decades or even longer [3,23,24]. Various studies have reported that some *H. pylori* strains can invade the gastric cells and remain inside for an indefinite time [25,26], although the molecular mechanism related to this phenomenon has not been fully understood. Recently, Morales-Espinosa *et al.* demonstrated that *vapD* is expressed in the intracellular environment of adenocarcinoma gastric (AGS) cells and high levels of *vapD* expression were

detected in gastric biopsies of patients with severe gastric pathologies. This suggests that VapD is necessary to support the long-term persistence of *H. pylori* in gastric cells and maintain the chronicity of the infection [3].

The *vapD* gene encodes for a virulence-associated protein D (VapD) that is found in various microorganisms from different phyla. Reports suggest that in some microorganisms, VapD participates in protecting the bacteria from respiratory burst with in the macrophage or in facilitating the persistence of the microorganism within the respiratory epithelial cell [27-29]. Virulence-associated protein (*vap*) genes were first identified in pathogenic strains of *Dichelobacter nodosus* [30,31] as part of a pathogenicity island of a plasmid. Subsequently, *vap* genes have been recognized in other microorganisms, such as *Haemophilus influenza* [29], where *vapD* is present in the bacterial chromosome as part of a Toxin-Antitoxin (TA) module with ribonuclease activity that promotes bacterial persistence inside respiratory cells. The action of Toxin and Antitoxin (TA) module systems favour bacterial persistence within epithelial cells enabling them to adapt to environmental conditions and antibiotic treatment [32-35]. Since persistence is a major factor contributing to the chronic state of infections and tolerance to antibiotic treatment, it was proposed that one of the roles of TA was to contribute to dormancy, i.e., making the cells metabolically inactive [36,37].

In *Rhodococcus equi*, *vapA* and *vapD* genes are present in a virulence plasmid and are highly induced in an acid tolerance response within the macrophage. The ability to withstand a stressful environment is an important factor in the virulence of an intracellular bacterium [27,28]. In *Actinobacillus actinomycetemcomitans* [38], the product of the *orf2* gene is known to have 78.9% amino acid identity with VapD of *D. nodosus*. The *vapD* gene is a strain-specific gene that contributes to the higher genetic diversity of *H. pylori.* It was first described in the 60190 strain by Cao and Cover in 1997, in a variable chromosomal region of 3.8 kb downstream from *vacA* gene. In the 26695 strain, it is found in the HP0315 locus while in the J99 strain, *vapD* is located in the JHP0829 locus, but the ORF in this case is truncated encodes for a non-functional protein. Subsequent data suggests that *vapD* is present in 36% to 61% of *H. pylori* strains [38,39]. In 2012, Kwon *et al* [40]determined the structural and biochemical characteristics of VapD and found that this protein displeyed a purine-specific endoribonuclease activity, which was later shown to be structurally related to the Cas2 proteins [41].

Although an endoribonuclease function has been attributed to VapD in *H. pylori*, it is not clear what the mechanism of action is, or how important the *VapD* gene is in this specific bacterium, since it is widely distributed among microorganisms of different phyla [28,42,43]. Furthermore, previous studies in some microorganisms of different genera have suggested that the *vapD* gene, originates either from phages through integration events, or from a plasmid containing bacteriophage-related *int* genes [44]. Therefore, the aim of this study was to determine the phylogenetic relationship of the *Helicobacter pylori vapD* gene with the *vapD* gene present in other bacterial species from different phyla in order to infer its possible horizontal transfer, and to estimate the genealogy from *vapD* gene within the *H. pylori* species.

Material and Methods

Background of *H. pylori vapD* isolated from Mexican patients

In a previous study, we characterized the *vapD* gene from a group of Mexican *H. pylori* strains (MxHp) isolated from adults and children [39]. In that study, a set of primers (D1 and D2), were used to obtain a 485 bp PCR product (expected size), but we also obtained *vapD* gene amplicons (from 800 bp to 1300 bp) that were larger than had been expected. In terms of the chromosomal *vapD* region of the 60190 strain, our D1 and D2 primer sequences were located at 100 bp upstream from the *vapD* ORF 5' end to 100 bp downstream from the *vapD* ORF 3' end. According to previous results, we considered that there could be a greater variability in the *vapD* gene from Mexican strains, as well as in the *vapD* region, than had been described previously [38,39].

vapD gene sequencing

From the previously mentioned study [39], we randomly selected 16 amplicons of *vapD* gene, 7 of which had an expected size of 498 bp and nine had a larger than expected PCR products. All the products were cloned into a pCR 2.1 vector (Invitrogen, Carlsbad, CA) and sequenced on both strands using the M13 sequencing primers and the Sanger method. The SeqMan program of the DNASTAR Lasergene 7 package (DNASTAR, Inc., Madison, WI) was used to edit, trim and assemble each sequence.

The GenBank ID for the *vapD* gene sequences obtained from our laboratory [39] were as follows: MxHp72a6 strain (AY781665), MxHp72a10 strain (AY781666), MxHp252a2 strain (AY781671), MxHp254a8 strain (AY781670), MxHp254c9 strain (AY781669), MxHp128 strain (EU822947), MxHp563 strain (EU822948), MxHp262c11 strain (AY781668), MxHp262c13 strain (AY781667), MxHp21.23a strain (EU818714), MxHp21.23c strain (EU822949), MxHp248 strain (EU826974), MxHp249 strain (EU826975), MxHp54 strain (EU826976), MxHp84 strain (EU826977) and MxHp118 strain (EU818713). The *vapD* sequences of the *H. pylori* 26695, 60190 and J99 strains were used as reference sequences. For *Helicobacter* species and the other bacterial species included in the study, the sequences were downloaded from the NCBI (Table S1).

Genealogical analyses of *vapD* gene among *H. pylori* strains

In order to assess the genealogical relationship of the *vapD* gene within *H. pylori* species, we selected *vapD* sequences from strains that had been isolated from different parts of the world, including our sequences reported in the GenBank database (Table S1), to create a network of haplotypes. This assessment used the most conserved region of the *vapD* gene (158 bp) sequence. Genealogy was reconstructed using TCS 1.21 software [45] and based on individual allele (haplotype) sequences. The haplotype network shows the connection between each allele, inferring the number of mutational steps between the connecting axes of each. When there is more than one mutational step between one haplotype

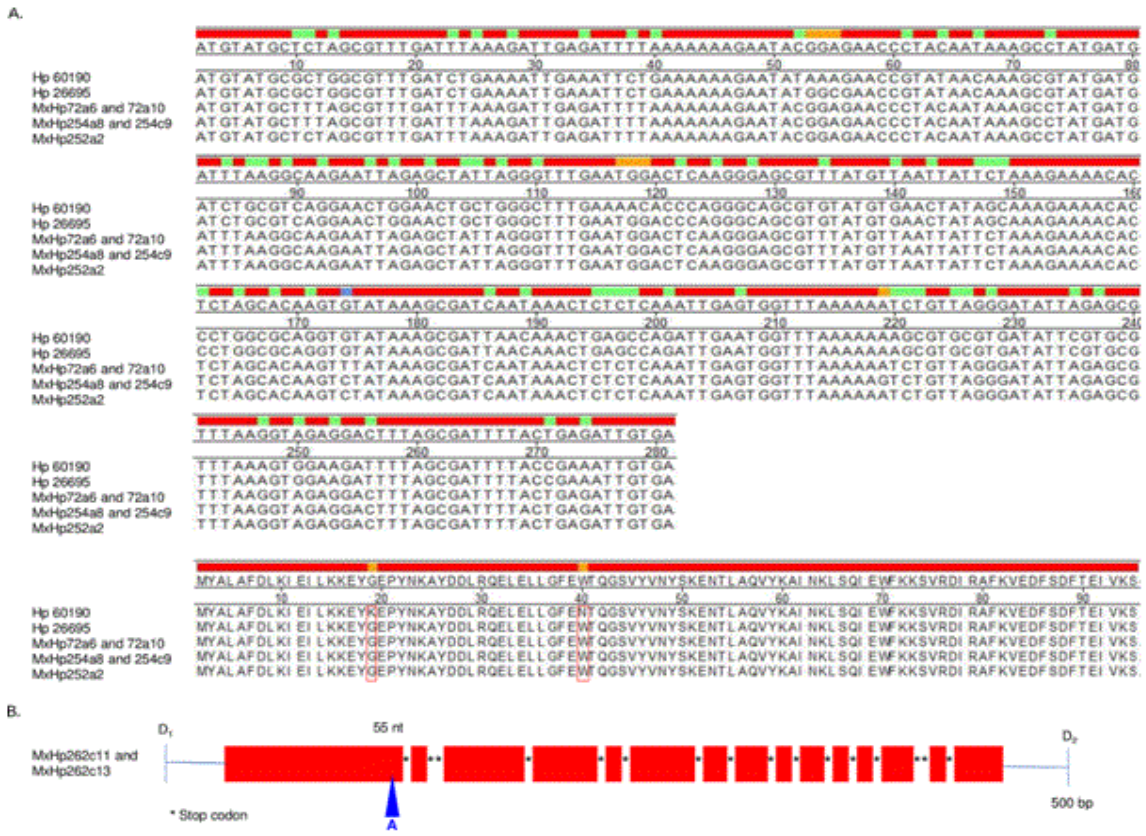

Figure 1: Alignment of *Helicobacter pylori vapD* gene from Mexican strains. (A) Alignment of nucleotide sequences of *vapD* ORF from Mexican (MxHp) strains and 60190 strain. Sequence analysis showed a large nucleotide polymorphism between the sequences. However, the amino acid alignment of the deduced *vapD* product showed an identical amino acid sequence among MxHp strains and 26695 strain, and only two aa changes with respect to the 60190 strain. (B) The red cases indicate that the *vapD* sequences for the MxHp strains were identical but in the MxHp 262c11 and MxHp 22c13 strains, there was an adenine (A) insertion around nucleotide 55, which caused changes in the reading frame and with the subsequent formation of stop codons (*). All *vapD* genes were amplified with D1 and D2 primer set.

Table 1: Traces of IS elements present in the *vapD* region of *H. pylori* strains from Mexican populations

Strain	IS family	Group	IS	Origin	*E. value	Sequence
21.23a	IS *5*	IS *5*	IS *Pto9*	*Psycroflexus torquis*	0.54	GTTTAGCCCTATCTTA
54	IS *1595*	ISPna *2*	IS *Aur1*	*Actinobacillus ureae*	0.8	AACAAAAAAATTAG
54	IS *3*	IS *3*	IS *Cb3*	*Clostridium beijerinncki*	0.8	TAACAAAAAAATTAGAAAT
54	IS *4*	IS *231*	IS *Cb2*	*Clostridium beijerinncki*	0.8	AAAAGATCAATAACAAAAA
84	IS *1595*	ISPna *2*	IS *Aur1*	*Actinobacillus ureae*	0.8	TAACAAAAAAATTAGG
84	IS *3*	IS *3*	IS *Cb3*	*Clostridium beijerinncki*	0.79	TAACAAAAAAATTAGAAAT
84	IS *4*	IS *231*	IS *Cb2*	*Clostridium beijerinncki*	0.79	AAAAGATCAATAACAAAAA
248	IS *607*	-	IS *607*	*Helicobacter pylori*	1	TAAAAAACTTATTAAA
249	IS *607*	-	IS *607*	*Helicobacter pylori*	1	TAAAAAACTTATTAAA
21.23c	IS *200*/IS *605*	IS *1341*	IS *Hasp2*	*Halobacterium sp.*	0.11	CGGTGAATGGTTCGCT
21.23c	IS *1595*	ISPna *2*	IS *Aur1*	*Actinobacillus ureae*	0.8	TAACAAAAAAATTAG
21.23c	IS *3*	IS *3*	IS *Cb3*	*Clostridium beijerinncki*	0.8	TAACAAAAAAATTAGAAAT
21.23c	IS *4*	IS *231*	IS *Cb2*	*Clostridium beijerinncki*	0.8	AAAAGATCAATAACAAAAA

118	IS *1595*	IS *Pna2*	IS *Caje5*	*Campylobacter jejuni*	0.09	TTTAAAAAAGGAATAA
118	IS *3*	IS *51*	IS *Spr1*	*Serratia proteamaculans*	0.09	GATATTTTAAAAAAGG
118	IS *1595*	IS *Pna2*	IS *Pto1*	*Psycroflexus torquis*	0.39	CAAACCCTTTATAAC
*E. value: expected value						

and another, the existence of an intermediate haplotype can be assumed, which is represented by a small black circle or a number that represent the mutational steps.

Horizontal transfer inference of *vapD* gene

Currently, the methods for inferring horizontal transfer events are based on the analysis of gene sequences or complete genomes. In this current study, we analysed the *vapD* gene sequences using a phylogenetic method based on the reconstruction and comparison of phylogenetic trees. Two phylogenetic trees (16S rRNA and *vapD*) from 6 bacterial phyla and several *H. pylori* strains isolated from different parts of the world were constructed. This was followed by two parametric methods that use specific characteristics of the gene sequence, such as GC content and codon usage [46,47].

16S rRNA phylogenetic reconstruction of bacterial phyla

The phylogeny of different bacterial phyla was inferred from 16S rRNA sequence comparisons. Sequences of 152 strains from six bacterial phyla and 72 species were analysed (Table S1). These species were selected because they contained the *vapD* gene. The sequences were retrieved from the GenBank database and were aligned with Muscle software v.3.8.3 [48]. jModelTest [49] was used to select the optimal evolutionary model by evaluating the selected parameters using a corrected version of the Akaike Information Criteria (AIC). This approach suggested the HKY + I + G substitution model. Phylogeny was performed using MrBayes software v 3.0 [50] with five Markov chain Monte Carlo (MCMC) being run for five million generations. After discarding the first 20% iterations, the phylogenetic tree was constructed using Figtree v1.4.3 [51].

Phylogenetic analysis of the *vapD* gene from different bacterial species

In order to explore the phylogenetic relationships of the *vapD* gene among different bacterial phyla, as well as among *H. pylori* strains, we selected 211 *vapD* sequences from a wide variety of bacterial species found in GenBank. All the sequences were aligned using the Muscle software v.3.8.3 [48]. The substitution model was the same as for 16S rRNA. *vapD* phylogeny was constructed using MrBayes software v 3.0 [50]. An independent run of MrBayes was performed and consisted of five MCMC for ten million generations, the first 20% iterations being discarded as burn-in. Finally, the phylogenetic tree was drawn using Figtree v1.4.3 [51].

Codon usage bias and GC content

Using the online Codon Usage Database program [52], we obtained the Relative Synonymous Codon Usage (RSCU) of the genome Coding Regions (CDS's) for the *H. pylori* 26695 strain. For the *vapD* gene, the RSCU values were obtained by DnaSP v5.1 software [53]. Methionine (Met) and tryptophan (Trp) that have unique codons, and the stop codons UAG, UAA and UGA, were all excluded from further analyses.

For the *H. pylori* genomic GC content, we collected information for 125 complete genome assemblies from the NCBI database. For the *vapD* gene, we used three sequences obtained during the previously mentioned study [39], and 26 sequences from the NCBI database. The GC content for each sequence was calculated using DnaSP v5.1 software [53].

X-squared was used to examine the significance ($P<0.05$) of codon usage differences and GC content between genomes and the *vapD* gene. Rstudio v3.2.2 [54] was employed for all statistical analyses.

Determination of the insertion sequences present in the *vapD* region

Mobile genetic elements, such as Insertion Sequences (IS's), have played an important role in moving genetic material between organisms, including those during the early stages of evolution. To explore the presence of IS in the *vapD* region, we used the database from the ISfinder website [55].

Results

Based on the PCR product size, 16 *vapD* sequences from *H. pylori* Mexican strains were chosen to be cloned, sequenced, and analysed. Among the 16 analysed sequences, we found genetic variations and rearrangements that affected the structure of *vapD* gene. In 9 strains, the *vapD* gene was complete, while in 6 strains only a fragment of the gene was found, and in one strain the entire gene was deleted.

Genetic diversity associated with the *H. pylori vapD* gene from Mexican strains

The strains MxHp72a6, MxHp72a10, MxHp254a8, MxHp254c9, MxHp252a2, MxHp262c11, and MxHp262c13 contained the complete *vapD* ORF, and among these sequences, three polymorphic sites were found. Strains MxHp72a6, MxHp72a10, MxHp254a8 and MxHp254c9 showed nucleotide substitutions in only two positions, at nucleotides 171 and 216 of the *vapD* ORF. Meanwhile, the MxHp252a2 strain presented a third substitution at position 10. When the nucleotide sequences of these Mexican strains were compared against the nucleotide sequences of the reference strains, 26695 and 60190, a higher polymorphism was observed throughout all the *vapD* sequences

(Fig. 1A). However, VapD amino acid sequence analysis (Mexican and reference strains) showed high homology among all proteins with only two amino acid changes (Fig. 1A), which reflected more synonymous than non-synonymous substitution events (dN/dS= -1.403), thereby suggesting that these genes were under purifying selection. On the other hand, the nucleotide sequence analysis of the MxHp262c11 and MxHp262c13 strains showed a nucleotide (adenine) insertion around position 55. This nucleotide insertion caused a frame shift with the formation of several internal stop codons yielding a truncated protein (Fig. 1B) [39].

Genetic rearrangements associated to *H. pylori vapD* gene and *vapD* region from Mexican strains

Sequence analysis of MxHp21.23a, MxHp54, MxHp84, MxHp248, MxHp249, MxHp21.23c, MxHp563 and MxHp118 strains, all of which yielded a larger than the expected PCR product (from 860 bp to 1259 bp), showed high nucleotide identity with two discontinued chromosomal regions of the J99 strain, from nucleotide 911502 to 910648 and from nucleotide 908908 to 908521. In our strains, complete DNA fragment analysis showed a large deletion of 1738 bp with respect to the J99 strain (from 910647 to 908909), which corresponds to jhp0827 (*tnpA*) and jhp0826 (*tnpB*) loci present in the J99 strain, but not in our strains (Fig. 2). The deletion in our strains led to this chromosomal portion being rearranged in their ORFs, generating up to four new ORFs. Particularly ORF2, which was present in all of our strains, encodes for a hypothetical protein comprising of 93 amino acids, and exhibits high homology (93%) with proteins encoded for ORFs of different plasmids, such as ORF1 of pHac1 of the *Helicobacter acinonychis* strain Sheeba, *H. pylori* ORF12 of pAL226; ORF5 of pHe15; HPAG1poo6; pHP69; pHe14-07 and R4 pHP666 (Fig. 2). ORF1 and ORF4 were also formed and found in MxHp21.23a, MxHp54, MxHp84 and MxHp21.23c strains. ORF1 has a 134 bp size and encodes for a hypothetical protein of 44 amino acids. ORF1 exhibits no homology with any protein reported in the database, although it overlaps with the amino-terminal region (20 amino acids) of the hypothetical protein JHP0829, as well as with the *vapD* gene (from 5 to 14 amino acids) of *H. pylori.* ORF4 (present in the complementary strand) is between 203 bp and 233bp in size and is similar (94%) to the hypothetical protein PMM1631 of *Prochlorococcus marinus*. The fourth ORF present was ORF3, which was detected in a complete form (98 amino acids) in the MxHp 21.23a strain. ORF3 is similar to several other genes or ORFs, including *H. pylori* JHP0825, ORF2 of pHAC1 of *H.*

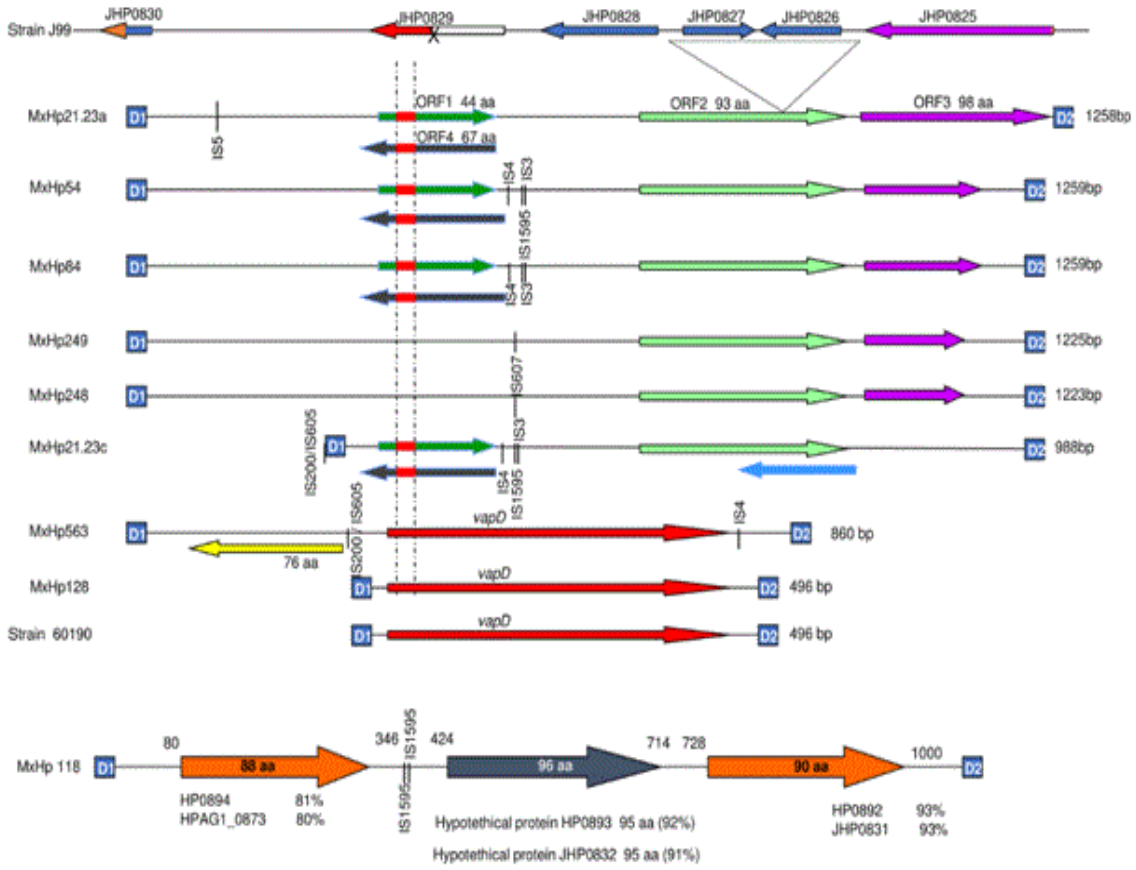

Figure 2: Chromosomal arrangements of the *vapD* region present in Mexican strains. Using D1 /D2 primers set, the different chromosomal arrangements of the *vapD* region in Mexican (MxHp) strains, presented PCR products that were larger than expected. Comparison and analysis were made against the J99 type strain. The position of the "scars" present in each sequence for different IS´s is also shown.

acinonychis strain Sheeba, *H. pylori* ORF13, addiction module toxin, RelE/StbE family and the plasmid stabilization system of *Xylella fastidiosa*. In MxHp54, MxHp84, MxHp249 and MxHp248 strains, ORF3 is truncated and overlaps with the N-terminal region of the hypothetical protein ORF13, JHP0825 of *H. pylori,* and pHAC1_2 of *H. acinonychis.* ORF1 and ORF4 were not present in strains MxHp248 and MxHp249. Arrangements that have commonality with the strains described above are those of ORF2, that are also found in a complete form, the truncated ORF3, and 30 nucleotides that correspond to 10 conserved amino acids (AFDLKIEILK) of the VapD protein. Sequence analysis of the MxHp563 strain showed both a complete *vapD* ORF and a different ORF that was not previously formed in our strains, although it was described in NCBI databases in other *H. pylori* strains isolated from Latin American countries. This ORF encodes for a hypothetical protein of 76 amino acids, which to date has not been assigned a function. The MxHp128 strain contained the complete *vapD* ORF, which was identical to the *vapD* gene of the 60190 strain. Nonetheless, seven polymorphic sites were found with respect to the *H. pylori* 26695 strain. Finally, sequence analysis of the MxHp118 strain showed a greater chromosomal variability in the *vapD* region with the formation of 3 ORFs: the first corresponded to a hypothetical protein of 88 amino acids, which identifies with HP0894, HPAG1_0873, type II toxin-antitoxin system mRNA interferase toxin, RelE/StbE family of *H. pylori* and type II toxin-antitoxin system YafQ family toxin; the second ORF encodes a hypothetical protein of 96 amino acids; and the third ORF encodes a protein of 90 amino acids, which identifies with the type II toxin-antitoxin system mRNA interferase toxin, RelE/StbE family [*Helicobacter pylori*] or the type II toxin-antitoxin system YafQ family toxin [*Helicobacter pylori*].

Genealogical analyses of *Helicobacter pylori vapD* gene

The genealogical network estimation with statistical parsimony showed a haplotype network with the formation of two main clades, namely clade A and B (Fig. 3). Clade A formed a homogenous and compact cluster that diverged from clade B at 31 mutational steps, and comprised most of the analysed sequences, which were isolated from different geographical regions, and from different genetic populations (hpAfrica2 and hpAsia2) and subpopulations (hspAmerind, hspAfrica1NAmerica, hspEuropeColombia, hspEuropeN, hspEuropeS, hspSAfrica and hspWAfrica) [7,56,57]. Nevertheless, several samples of this group (50%) could not be assigned to any genetic population or subpopulation. The presence of loops into clades A and B suggests signals of recombination among members of each clade, but not between clades. A diversification process was observed in clade A. The Mexican strains with the complete *vapD* ORF and the *vapD* gene haplotypes from 60190 and 26695 (hspEuropeN) strains were grouped into clade B. Further investigation into clade B showed that two divergent clades, namely C and D were present. Clade C was positioned at approximately 44 mutational steps from clade B and was formed by 3 strains from Europe, which belonged to the hspEuropeN, hspEuropeS and hspEasia subpopulations (Fig. 3). Meanwhile, clade D diverged at more than 60 mutational steps from clade B and was formed by 6 Mexican strains that presented structural reorganization in the *vapD* region. Strain J99 (hspWAfrica), which has a truncated *vapD*, was also found in this clade D group (Fig. 3).

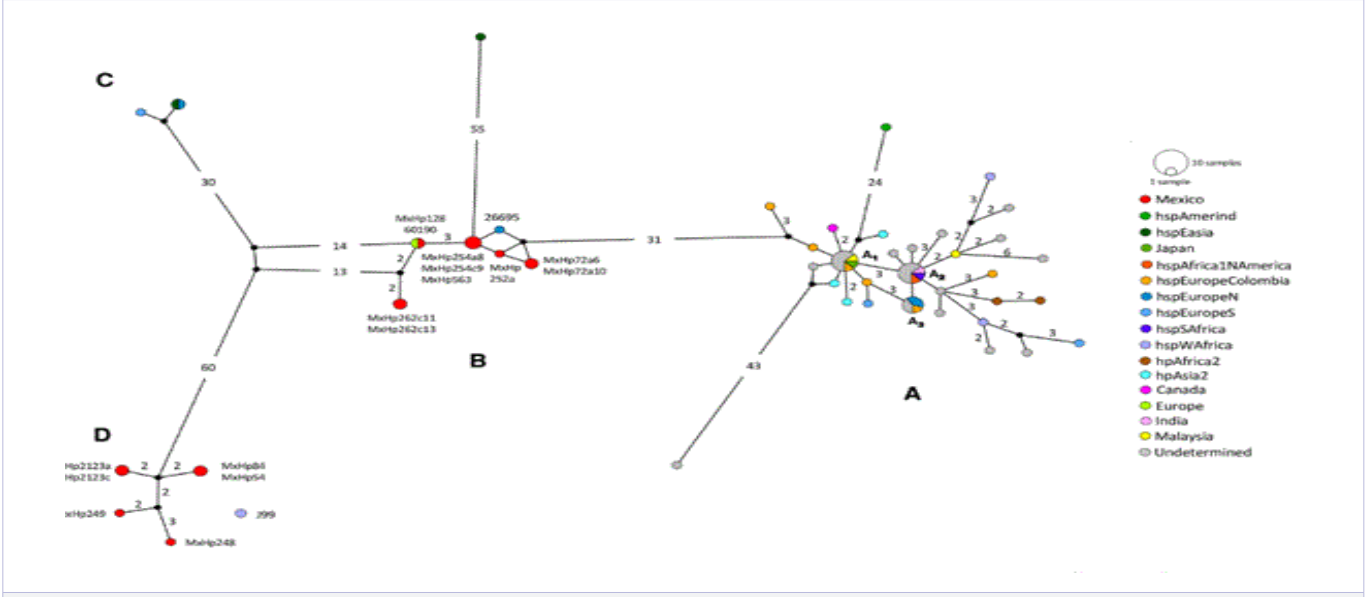

Figure 3: Genealogical relationship of the *vapD* gene of *Helicobacter pylori* strains isolated from different parts of world. This figure shows the network of haplotypes that form two well-defined subpopulations, A and B, and two smaller groups, C and D. The numbers between bars show the mutational steps that diverge between one subpopulation and another; the small black circles indicate a hypothetical intermediate haplotype and the circle size refers to the number of sequences that have that haplotype. Subpopulation A clusters most of the sequences arising from strains from different parts of the world; the cluster of subpopulation B corresponds to strains isolated from Mexico and the collection strains, 26695 and 60190. Subpopulations C and D are very divergent groups. More specifically, cluster D enclosed all those *vapD* sequences that are truncated, such as J99, and those that presumably suffered different genetic events that resulted in the fragmentation of the gene.

vapD gene phylogenetic relationship among different bacterial phyla, codon usage and GC content

Analysis of the 16S rRNA phylogenetic tree showed that the cluster corresponding to the Proteobacteria phylum was grouped into four well defined and separate Proteobacteria Classes: alpha, beta, gamma and epsilon (α, β, γ and ε) (Fig. 4A). In this same cluster, there were also two other phyla: Firmicutes and Actinomycetales. In the second and third clusters, we found the Cyanobacteria and Fusobacterium phyla. The last cluster (bottom of the tree) was formed by the Bacteroidetes phylum that included 3 families (Bacteroidetes, Sphingobacteria and Flavobacteria) [58-60].

The phylogeny of the 16S rRNA gene reflects the relationships (ancestor-offspring) of the species at the level of large taxonomic groups. When this phylogeny was compared against the phylogeny of the *vapD* gene, it showed a great phylogenetical incongruence among bacteria of both the same phylum and different phyla, suggesting that the *vapD* gene could have participated in horizontal gene transfer. This idea is supported by the fact that this gene is present in different species that belong

Figure 4: Markov Chain Monte Carlo (MCMC) phylogenetic trees of 16S rRNA and *vapD* genes. (A) 16S rRNA MCMC tree. This phylogeny reflects the ancestor-offspring relationships of the species at the large, taxonomic group level. The branch colours depict the taxonomy of the 16S rRNA based lineages: Proteobacteria (green), Firmicutes (dark green), Actinomycetales (red), Cyanobacteria (blue), Fusobacterium (brown) and Bacteroidetes (lilac). (B) *vap* DMCMC tree. The *vapD* phylogenetic tree shows the relationship of this gene present in species of divergent phyla. Significant conflict between the 16S rRNA and *vapD* trees is shown by the alternative diagonal lines. These phylogenetic incongruities could be explained by subsequent ancient horizontal transfer events..

to very divergent phyla.

On the other hand, the phylogeny of the *vapD* gene showed a large dispersion without a genetic relationship between species nor grouping within a defined phylum, as was seen in the case of the 16S rRNA gene. In general, *vapD* phylogeny showed a great intragenic diversity within the same species that can be extrapolated to describe the diversity of the gene within the same phylum. Even though most of the sequences of the *Helicobacter vapD* gene were in the same branch, they did not form a compact cluster. Some *Helicobacter* species were observed closer to bacteria from the Fusobacterium phylum, Firmicutes phylum or to Actinomycetales. This issue was not unique to the *Helicobacter* genus but also observed in members of other phyla. In the 16S rRNA phylogeny of the Bacteroidetes phylum, the species were grouped into a single cluster, while in the *vapD* phylogeny they were divided into five clusters. The first three *vapD* clusters were located near to the α and γ-proteobacteria, while the other two clusters were in a separate and independent branch close to both the ε-proteobacteria group and members of the Synergistetes phylum. Meanwhile, strains of the *Xylella fastidiosa* and *Aggregatibacter* sp. form different and distant clusters in the 16S rRNA phylogeny but in the *vapD* phylogeny, they were grouped into the same cluster (Fig. 4).

Relative Synonymous Codon Usage (RSCU) values and GC content were calculated for the complete genome of *H. pylori* and for the *vapD* gene (Table S2). RSCU values greater than 1, reflected a preference for the use of a specific codon to determine amino acids. RSCU values for the *H. pylori* complete genome were as follows: 26 codons had a value greater than 1 and 29 had a value less than 1. When the codon usage of the *vapD* gene was obtained, 16 had a value greater than 1, while none were detected with a value less than 1. The following codons were shared between the *H. pylori* chromosome and *vapD* gene: UUU (F), AAU (N), GUG (V), AGC (S), ACC (T), UAU (Y), AAA (K), GAU (D) and GAA (E); only leucine (L) showed a significant difference (p=0.0487). Histidine (His) and cysteine (Cys) were removed from the table because there were no RSCU values for the *vapD* gene (Table S2).

The average GC content in the whole *H. pylori* genome was 38.84 ± 0.19. This was lower than for the *vapD* gene, which had a content of 41.23 ± 0.92 (Fig. S1). These results suggest that *vapD* could be foreign DNA acquired from an external source other than *H. pylori.*

Presence of IS in *H. pylori vapD* region

Insertion sequence (IS) analysis in the *vapD* region of our strains showed traces of different IS sequences from different families. In the *vapD* region of the MxHp21.23a strain, we found remnants of a 16 bp of IS *Pto9* (transposase) from the IS *5* family (Table 1).

MxHp54 and MxHp84 strains presented traces of different ISs: 15 bp of IS *Aurf* from the IS *1594* family, 19 bp of IS *Cb3* from the IS *3* family, and 19 bp of IS *Cb2* from the IS4 family (Table 1). With respect to MxHp248 and MxHp249 strains, we found scars of IS *607* (16 bp), which is specific to *H. pylori.* The *vapD* region of the MxHp21.23c strain was the region with most traces of ISs, where we identified IS *3* and IS *4*, which are families of *Clostridium beijerinncki*. This strain also carried 15 bp of IS *Aur1* belonging to the IS *1595* family from *Actinobacillus ureae* and 16 bp belonging to the IS *200*/IS *605* family. Finally, we identified traces of the IS *1595* (16 bp) family from *Campylobacter jejuni* and *Psycroflexus torquis* and 16 bp from the IS *3* family of *Serratia proteomaculans* in strain MxHp118 (Table 1).

Table S1: GenBank accession number for the sequences used in this study.

GenBank accession number	Isolate	Locus
NP_207113	*H. pylori* 26695	*vapD*
ASYV01000064	*H. pylori* PZ5080	*vapD*
AKHP02000000	*H. pylori* FD535	*vapD*
EQL49894	*H. pylori* FD430	*vapD*
OOP87307	*H. pylori* CA22327	*vapD*
PDW75985	*H. pylori* 22311	*vapD*
EKE81371	*H. pylori* R030b	*vapD*
PUD52554	*H. pylori* GC65HL	*vapD*
AEN17075	*H. pylori* SNT49	*vapD*
KNX43647	*H. pylori* UM300	*vapD*
EJC13024	*H. pylori* HP23	*vapD*
ANH42653	*H. pylori* L7	*vapD*
PUD05235	*H pylori* 38:2	*vapD*
EKE84337	*H. pylori* R32b	*vapD*
PUB97335	*H. pylori* 55:2	*vapD*
OOP75995	*H. pylori* CA2231	*vapD*
ANH45646	*H pylori* CC33C	*vapD*
PUD18190	*H. pylori* 56599	*vapD*
AAC45241	*H. pylori* 60190	*vapD*
PUD52921	*H. pylori* GC23HL	*vapD*
EJB78246	*H. pylori* HpA27	*vapD*
PUD75866	*H. pylori* B31	*vapD*
PUD39986	*H. pylori* 3755	*vapD*
PUB99539	*H. pylori* 38:5	*vapD*
PUD86084	*H. pylori* B25	*vapD*
PDX33282	*H. pylori* 2061	*vapD*
PDW33764	*H. pylori* 3053	*vapD*
KNX48507	*H. pylori* UM408	*vapD*
EMG86822	*H. pylori* GAM114Ai	*vapD*
AUZ23548	*H. pylori* dRdM2addM2	*vapD*
BAW56543	*H. pylori* F55	*vapD*
CP002953	*H. pylori* ELS37	*vapD*

CP003419	*H. pylori* XZ274	*vapD*
CP000241	*H. pylori* HPAG1	*vapD*
EQD88536	*H. pylori* SouthAfrica50	*vapD*
CP000012	*H. pylori* 51	*vapD*
PDX44683	*H. pylori* 2006	*vapD*
OOQ10045	*H. pylori* CC26100	*vapD*
PUD90214	*H. pylori* B23+S27R2:R48	*vapD*
CP002073	*H. pylori* SJM180	*vapD*
AFX90414	*H. pylori* Aklavik86	*vapD*
OPG59870	*H. pylori* 2036	*vapD*
EMH03913	*H. pylori* GAM245Ai	*vapD*
PDW32774	*H. pylori* 3056	*vapD*
EQD94110	*H. pylori* PZ5026	*vapD*
AP014523	*H. pylori* NY40	*vapD*
PDW41337	*H. pylori* 22316	*vapD*
PDW34671	*H. pylori* 3046	*vapD*
CP003486	*H. pylori* HUPB14	*vapD*
AGT74177	*H. pylori* SouthAfrica20	*vapD*
WP_100949690	*H. pylori*	*vapD*
WP_033596798	*H. pylori*	*vapD*
WP_050545892	*H. pylori*	*vapD*
WP_001988454	*H. pylori*	*vapD*
WP_080025382	*H. pylori*	*vapD*
AFI03858	*Helicobacter cetorum* MIT 00-7128	*vapD*
WP_104748587	*Helicobacter cetorum*	*vapD*
EEO24159	*Helicobacter bilis* ATCC 43879	*vapD*
KGL25492	*Helicobacter bilis* ATCC 49314	*vapD*
WP_104747237	*Helicobacter bilis*	*vapD*
WP_104696282	*Helicobacter salomonis*	*vapD*
WP_027327662	*Helicobacter pametensis*	*vapD*
WP_018450011	*Leptotrichia shahii*	*vapD*
WP_071124930	*Leptotrichia massiliensis*	*vapD*
WP_021744953	*Leptotrichia sp. oral taxon 879*	*vapD*
WP_018497980	*Leptotrichia wadei*	*vapD*
WP_026747354	*Leptotrichia trevisanii*	*vapD*
WP_064615231	*Streptobacillus*	*vapD*
CP001779	*Streptobacillus moniliformis* DSM 12112	*vapD*
CP027400	*Streptobacillus moniliformis strain* FDAARGOS_310	*vapD*
WP_064608975	*Streptobacillus moniliformis*	*vapD*
WP_012858896	*Streptobacillus moniliformis*	*vapD*
SDF71462	*Sporolituus thermophilus* DSM 23256	*vapD*
WP_093691362	*Sporolituus thermophilus*	*vapD*
KGQ40318	*Gallibacterium anatis*	*vapD*
KGQ53121	*Gallibacterium anatis*	*vapD*
OKZ70694	*Clostridiales bacterium* 41_12_ two_minus	*vapD*
CDE43797	*Clostridium* sp. CAG:411	*vapD*
DQ517426	*Actinobacillus pleuropneumoniae* 12494 plasmid p12494	*vapD*
AEJN02000103	*Aggregatibacter actinomycetemcomitans*	*vapD*
CP007502	*Aggregatibacter actinomycetemcomitans* HK1651	*vapD*
CP001733	*Aggregatibacter actinomycetemcomitans* D11S-1	*vapD*
CP012958	*Aggregatibacter actinomycetemcomitans strain* VT1169	*vapD*
CP003099	*Aggregatibacter actinomycetemcomitans* ANH9381	*vapD*
CP016553	*Aggregatibacter actinomycetemcomitans strain* IDH781	*vapD*
CP003496	*Aggregatibacter actinomycetemcomitans* D7S-1	*vapD*
CP012959	*Aggregatibacter actinomycetemcomitans strain* 624	*vapD*
AP014520	*Aggregatibacter actinomycetemcomitans* NUM4039	*vapD*
PCGW01000011	*Aggregatibacter actinomycetemcomitans* 310b	*vapD*
CP001607	*Aggregatibacter aphrophilus*	*vapD*
AKS65690	*Aggregatibacter aphrophilus* NJ8700	*vapD*
AAC37126	*Aggregatibacter actinomycetemcomitans Plasmid* pVT7361	*vapD*
WP_025141941	*Pedobacter jeongneungensis*	*vapD*
KIA92090	*Pedobacter kyungheensis*	*vapD*
CP018790	*Campylobacter sp. plasmid* pSUIS6137	*vapD*
ARR01447	*Campylobacter* sp. RM6137	*vapD*
CP004067	*Campylobacter coli* CVM N29710	*vapD*
KGI22836	*Prevotella timonensis*	*vapD*

CCY64135	*Prevotella* sp. CAG:1124	*vapD*
CDA67310	*Prevotella copri* CAG:164	*vapD*
CP002006	*Prevotella ruminicola* 23	*vapD*
AP018051	*Prevotella melaninogenica* GAI 07411 plasmid pPME0001	*vapD*
WP_005845454	*Prevotella dentalis*	*vapD*
CP003368	*Prevotella dentalis* DSM 3688	*vapD*
CDC29181	*Prevotella* sp. CAG:386	*vapD*
KGI22838	*Prevotella timonensis*	*vapD*
PVX43535	*Prevotella colorans*	*vapD*
OAV75707	*Bacteroidales bacterium* Barb7	*vapD*
OAV68226	*Bacteroidales bacterium* Barb6XT	*vapD*
OAV70465	*Bacteroidales bacterium* Barb4	*vapD*
WP_013619538	*Bacteroides salanitronis*	*vapD*
CP002531	*Bacteroides salanitronis* DSM 18170 plasmid pBACSA01	*vapD*
CP018939	*Bacteroides fragilis*	*vapD*
L22307	*Dichelobacter nodosus*	*vapD*
M74565	*Dichelobacter nodosus*	*vapD*
L22308	*Dichelobacter nodosus*	*vapD*
Q46565	*Dichelobacter nodosus virulence*	*vapD*
L31763	*Dichelobacter nodosus* tRNASer	*vapD*
CP000513	*Dichelobacter nodosus*	*vapD*
CP031475	*Dichelobacter nodosus* VCS1703A	*vapD*
CP022124	*Fusobacterium nucleatum* subsp. animalis strain ChDC F332	*vapD*
CP007062	*Fusobacterium nucleatum* subsp. animalis 7_1	*vapD*
RRD31709	*Fusobacterium nucleatum*	*vapD*
EGQ80846	*Fusobacterium nucleatum* subsp. animalis ATCC 51191	*vapD*
FP929056	*Fretibacterium fastidiosum* draft genome	*vapD*
CBL28245	*Fretibacterium fastidiosum*	*vapD*
AF364087	*Riemerella anatipestifer* plasmid pRA34/901	*vapD*
AF048718	*Riemerella anatipestifer* plasmid pCFC1	*vapD*
AF082180	*Riemerella anatipestifer* plasmid pCFC2	*vapD*
CP018939	*Bacteroides fragilis* strain Q1F2 plasmid Q1F2p2	*vapD*
CP012938	*Bacteroides ovatus* ATCC 8483	*vapD*
LT622246	*Bacteroides ovatus* V975	*vapD*
CP022384	*Capnocytophaga leadbetteri* H6253	*vapD*
CP022384	*Capnocytophaga leadbetteri* H6253	*vapD*
CP022022	*Capnocytophaga* sp. ChDC OS43	*vapD*
CP012589	*Capnocytophaga* sp. oral strain F0383	*vapD*
CP027232	*Capnocytophaga* oral F0512	*vapD*
CP001632	*Capnocytophaga ochracea* DSM 7271	*vapD*
YP007366462	*Citrobacter freundii* CFSTE plasmid pMobC	*vapD*
LK985408	*Escherichia coli* FHI100	*vapD*
LM996818	*Escherichia coli* FHI71	*vapD*
LM995883	*Escherichia coli* FHI30	*vapD*
CP006992	*Methylobacterium* sp. AMS5	*vapD*
AP014809	*Methylobacterium populi*	*vapD*
CP001298	*Methylobacterium extorquens* CM4	*vapD*
FP103042	*Methylobacterium extorquens* DM4	*vapD*
CP021054	*Methylorubrum zatmanii* PSBB041	*vapD*
LT962688	*Methylobacterium extorquens* TK 0001	*vapD*
FN995097	*Neisseria lactamica* 02006	*vapD*
CP019894	*Neisseria lactamica* Y921009	*vapD*
CP015886	*Neisseria meningitidis* strain 38277	*vapD*
SBO77403	*Neisseria gonorrhoeae* strain WHO Y	*vapD*
SBO57104	*Neisseria gonorrhoeae* strain WHO K	*vapD*
SBQ20831	*Neisseria gonorrhoeae* strain WHO F	*vapD*
SBO57445	*Neisseria gonorrhoeae* strain WHO M	*vapD*
AGU85211	*Neisseria gonorrhoeae* MS11 plasmid	*vapD*
NC_011034	*Neisseria gonorrhoeae* NCCP11945 plasmid Pngk	*vapD*
SBO74338	*Neisseria gonorrhoeae* strain WHO P	*vapD*
SBO58433	*Neisseria gonorrhoeae* strain WHO O	*vapD*
SBO57901	*Neisseria gonorrhoeae* strain WHO N	*vapD*

SBO57823	*Neisseria gonorrhoeae* strain WHO L	*vapD*
SBO57377	*Neisseria gonorrhoeae* strain WHO G	*vapD*
NP040415	*Neisseria gonorrhoeae* strain UM01 plasmid pJD1	*vapD*
CP034034	*Neisseria gonorrhoeae* FQ01 plasmid	*vapD*
CP022278	*Neisseria* sp. 10023	*vapD*
OFN85042	*Neisseria* sp. HMSC064E01	*vapD*
AWP54479	*H. influenzae* strain 10P129H1	*vapD*
CP005967	*H. influenzae* KR494	*vapD*
CP007472	*H. influenzae* 723	*vapD*
ARB90438	*H. influenzae* FDAARGOS_199	*vapD*
AIT67183	*H. influenzae* Hi375	*vapD*
CKG85575	*H. influenzae* NCTC8143	*vapD*
PRK52554	*H. influenzae* 84P36H1	*vapD*
NP438611	*H. influenzae* Rd KW20	*vapD*
BR00016	*H. influenzae* PittGG	*vapD*
CBW28768	*H. influenzae* 10810	*vapD*
AKA46443	*H. influenzae* 2019	*vapD*
AAX87502	*H. influenzae* 86-028NP	*vapD*
AJO88533	*H. influenzae* 477	*vapD*
CP005384	*H. parasuis* ZJ0906	*vapD*
CP001321	*H. parasuis* SH0165	*vapD*
CP009158	*H. parasuis* SH03	*vapD*
CP015099	*H. parasuis* SC1401	*vapD*
CP009237	*H. parasuis* KL0318	*vapD*
EU714231	*Xylella fastidiosa isolate* AZ04	*vapD*
EU714230	*Xylella fastidiosa isolate* NM02	*vapD*
AP018005	*Candidatus Rickettsiella viridis* ApRA04	*vapD*
ABB28825	*Chlorobium chlorochromatii* CaD3	*vapD*
PIE55217	*Dethiosulfovibrio peptidovorans*	*vapD*
OOO00658	*Epulopiscium* sp. Nele67Bin004	*vapD*
CP006942	*Mannheimia* sp. 1261	*vapD*
OUP06775	*Mediterranea* sp. An20	*vapD*
CP024450	*Moraxella osloensis* NP7	*vapD*
LADV01000113	*Peptococcaceae bacterium* BRH_c23	*vapD*
WP_108831492	*Peptoniphilus* sp. MarseilleP3761	*vapD*
CP012903	*Providencia rettgeri* strain N1501091 pNDM15109	*vapD*
EJW09502	*Rhodovulum* sp. PH10	*vapD*
LK931667	*Thiomonas* sp. CB2	*vapD*
CP000542	*Verminephrobacter eiseniae* EF012	*vapD*
AP018248	*Tolypothrix tenuis* PCC 7101	*vapD*
AP018307	*Aulosira laxa* NIES-50	*vapD*
AGC23530	*plasmid Citrobacter freundii*	*vapD*
WP_015353868	*Citrobacter freundii*	*vapD*
STM47029	*Escherichia coli* NCTC10757	*vapD*
KSW14140	*Proteus mirabilis* PM655	*vapD*
KSW14422	*Proteus mirabilis* PM593	*vapD*
EII22447	*Escherichia coli* 9.0111	*vapD*
NC_004854	*Rhodococcus equi* ATCC33701	*vapD*
JN990998	*Rhodococcus equi* BBG163	*vapD*
JN990997	*Rhodococcus equi* SNP89	*vapD*
P018172	*Calothrix* sp. NIES-2098	*vapD*
EU127530	*Campylobacter coli* 8693/04	16S rRNA
DQ174144	*Campylobacter jejuni* LMG 8843	16S rRNA
AF372091	*Campylobacter jejuni* NCTC 11351	16S rRNA
AF372092	*Campylobacter coli* LMG 6440	16S rRNA
L04312	*Campylobacter coli*	16S rRNA
AY554142	*Campylobacter* sp. BTP1Tcr	16S rRNA
AY554143	*Campylobacter* sp. EQ1	16S rRNA
AY554144	*Campylobacter* sp. WB1	16S rRNA
EU781617	*Campylobacter ureolyticus* 4	16S rRNA
GQ167665	*Campylobacter ureolyticus* UNSWR	16S rRNA
L04321	*Campylobacter ureolyticus* ATCC 33387	16S rRNA
GQ167666	*Campylobacter ureolyticus* UNSWCD	16S rRNA
DQ174173	*Campylobacter mucosalis* ATCC 43264	16S rRNA
M35016	*Dichelobacter nodosus* 198A ATCC 27521	16S rRNA
DQ016290	*Dichelobacter nodosus* AN363/05	16S rRNA
DQ016291	*Dichelobacter nodosus* AN484/05	16S rRNA
CP000513	*Dichelobacter nodosus* VCS1703A4	16S rRNA
AB078974	*Haemophilus parasuis* 322	16S rRNA
AB078973	*Haemophilus parasuis* 319	16S rRNA
CP000672	*Haemophilus influenzae* PittGG	16S rRNA
L42023	*Haemophilus influenzae* Rd KW20	16S rRNA
M88148	*Helicobacter acinonychis* 901193	16S rRNA

U18766	*Helicobacter bilis* Hb1 MIT 931909	16S rRNA
AF292378	*Helicobacter cetorum* 9956	16S rRNA
M88147	*Helicobacter pametensis* B9A	16S rRNA
U89351	*Helicobacter salomonis*	16S rRNA
U01330	*Helicobacter pylori* ATCC 43504	16S rRNA
Z25744	*Helicobacter pylori* NCTC 11916	16S rRNA
KF297892	*Helicobacter pylori* Hp1	16S rRNA
AF277832	*Helicobacter* sp. LU1	16S rRNA
JX001468	*Citrobacter freundii* 7/A10	16S rRNA
LN589732	*Citrobacter freundii* MF3631	16S rRNA
FJ971857	*Citrobacter freundii* ATCC 8090	16S rRNA
AF025363	*Citrobacter rodentium* CDC 184373	16S rRNA
AF025364	*Citrobacter sedlakii* CDC 469686	16S rRNA
AF025367	*Citrobacter gillenii* CDC 469386	16S rRNA
EU888872	*Citrobacter gillenii* A8P19	16S rRNA
KR088351	*Citrobacter gillenii* BK16	16S rRNA
AF025368	*Citrobacter braakii* CDC 8058	16S rRNA
JN118501	*Citrobacter braakii* C274	16S rRNA
JX518488	*Citrobacter braakii* UAAD6	16S rRNA
AF025369	*Citrobacter murliniae* CDC 297059	16S rRNA
KC017346	*Citrobacter murliniae* amBHI1	16S rRNA
AF025371	*Citrobacter farmeri* CDC 299181	16S rRNA
KC429579	*Citrobacter farmeri* TERIYE	16S rRNA
KP036922	*Citrobacter farmeri* TCS20	16S rRNA
AF025373	*Citrobacter werkmanii* CDC 087658	16S rRNA
KM268970	*Citrobacter werkmanii* C18	16S rRNA
HQ238425	*Citrobacter youngae* S521B50	16S rRNA
JF939011	Citrobacter youngae	16S rRNA
AB273741	*Citrobacter youngae* GTC 1314	16S rRNA
HF558364	*Citrobacter koseri* CDC813286	16S rRNA
JF935072	*Citrobacter koseri* CtST1	16S rRNA
HQ992945	*Citrobacter koseri* LMG 5519	16S rRNA
JX297468	*Citrobacter amalonaticus* TTK014	16S rRNA
KC689291	*Citrobacter amalonaticus* DD2	16S rRNA
KT027832	*Citrobacter amalonaticus* E91	16S rRNA
AM947041	*Citrobacter hormaechei* 615	16S rRNA
JQ390124	*Citrobacter* sp. F11	16S rRNA
AY513502	*Escherichia coli* O157	16S rRNA
AE014075	*Escherichia coli* CFT073	16S rRNA
U00006	*Escherichia coli* K12 MG1655	16S rRNA
JN654455	*Escherichia coli* NCTC 50365	16S rRNA
JF508184	*Escherichia coli* ATCC 11303	16S rRNA
KP005067	*Escherichia coli* EC01	16S rRNA
AJ810279	*Fusobacterium nucleatum animalis* OMZ 990	16S rRNA
GQ301042	*Fusobacterium nucleatum animalis* ATCC 51191	16S rRNA
AB573068	*Fusobacterium nucleatum* JCM 8532	16S rRNA
AJ133496	*Fusobacterium nucleatum* ATCC 25586	16S rRNA
AJ006964	*Fusobacterium nucleatum vincentii* ATCC 49256	16S rRNA
X55404	*Fusobacterium nucleatum* NCTC 12276T	16S rRNA
X55403	*Fusobacterium nucleatum* NCTC 11326T	16S rRNA
FJ717336	*Leptotrichia wadei* F0279	16S rRNA
AY029802	*Leptotrichia wadei* LB16	16S rRNA
KP192296	*Leptotrichia wadei* KA00185	16S rRNA
GU561360	*Leptotrichia trevisanii* TG9	16S rRNA
AY029805	*Leptotrichia trevisanii* LB06	16S rRNA
AF206305	*Leptotrichia trevisanii*	16S rRNA
CP012410	*Leptotrichia* oral W10393	16S rRNA
AF287813	*Leptotrichia* oral FAC5	16S rRNA
GU086183	Leptotrichia shahii PW1036	16S rRNA
AY029806	*Leptotrichia shahii* LB37	16S rRNA
AJ247245	*Neisseria meningitidis* ATCC 35559	16S rRNA
AY187940	*Neisseria meningitidis*	16S rRNA
X74900	*Neisseria meningitidis* NCTC 10025	16S rRNA
X74901	*Neisseria lactamica* NCTC 10617	16S rRNA
FN995097	*Neisseria lactamica* 02006	16S rRNA
AJ247241	*Neisseria lactamica* DSM 4691	16S rRNA
AJ247239	*Neisseria gonorrhoeae* DSM 9189	16S rRNA
EU233796	*Neisseria gonorrhoeae* NG19	16S rRNA
KF410894	*Neisseria gonorrhoeae* NF131677	16S rRNA
AY612187	*Riemerella anatipestifer* TW96015	16S rRNA
KT449829	*Riemerella anatipestifer* EF3	16S rRNA
AY612184	*Riemerella anatipestifer* TRa9	16S rRNA
JQ810973	*Riemerella columbina* M351	16S rRNA

AJ400913	*Methylobacterium extorquens* IAM 1081	16S rRNA
AJ400917	*Methylobacterium extorquens* IAM 12639	16S rRNA
AJ400914	*Methylobacterium extorquens* IAM 12630	16S rRNA
DQ346736	*Methylobacterium populi* 'PapViBa7'	16S rRNA
AJ549956	*Methylobacterium populi* BJ001	16S rRNA
AY248705	*Methylobacterium* sp. Mb 49	16S rRNA
AJ400938	*Methylobacterium* sp. NI2	16S rRNA
AF192343	*Xylella fastidiosa* ATCC35879	16S rRNA
CP006696	*Xylella fastidiosa* sandyi Ann1	16S rRNA
CP002165	*Xylella fastidiosa* GB514	16S rRNA
FJ755928	*Xylella fastidiosa* isolate AZ04	16S rRNA
AE009442	*Xylella fastidiosa* Temecula1	16S rRNA
FJ755926	*Xylella fastidiosa* isolate NM02	16S rRNA
AF228001	*Gallibacterium anatis* F 149	16S rRNA
AF228013	*Gallibacterium anatis* BJ3453	16S rRNA
AF302255	*Actinobacillus pleuropneumoniae* N273	16S rRNA
AY017472	*Actinobacillus pleuropneumoniae* HS143	16S rRNA
Y09654	*Actinobacillus* sp. M1933/96/1	16S rRNA
CP001733	*Aggregatibacter actinomycetemcomitans* D11S-1	16S rRNA
AB512012	*Aggregatibacter actinomycetemcomitans* IDH781	16S rRNA
AY362906	*Aggregatibacter aphrophilus* CCUG 3715	16S rRNA
CP001607	*Aggregatibacter aphrophilus* NJ8700	16S rRNA
KC866151	*Aggregatibacter* sp. Melo83	16S rRNA
AY425295	*Mannheimia* sp. BJ3956	16S rRNA
AF053898	*Mannheimia* sp. HPA121 CCUG 38468	16S rRNA
X81876	*Prevotella dentalis* DSM 3688	16S rRNA
AF218618	*Prevotella ruminicola* 223/M2/7	16S rRNA
AY331415	*Prevotella genomo* sp. P6 P4PB_24	16S rRNA
CP002530	*Bacteroides salanitronis* DSM 18170	16S rRNA
AB618792	*Bacteroides fragilis* JCM 17586	16S rRNA
AB599947	*Bacteroides intestinalis* SLC8-5	16S rRNA
X67610	*Capnocytophaga ochracea* ATCC 33596	16S rRNA
DQ012356	*Capnocytophaga leadbetteri* AHN8708	16S rRNA
AY005077	*Capnocytophaga* sp. A47ROY	16S rRNA
EF660750	*Pedobacter daejeonensis* PB46	16S rRNA
JN196132	*Pedobacter kyungheensis* THG-T17	16S rRNA
AB666454	*Pedobacter* sp. MaI11-5	16S rRNA
LC062896	*Streptobacillus moniliformis* NMS	16S rRNA
CP001779	*Streptobacillus moniliformis* DSM 12112	16S rRNA
Z35305	*Streptobacillus moniliformis* ATCC 14647	16S rRNA
AY337519	*Clostridium* sp. L15	16S rRNA
Y15984	*Clostridium* sp. RXyl1	16S rRNA
AJ229244	*Clostridium* sp. VeCb10	16S rRNA
GQ287651	*Tolypothrix distorta* SAG 93.79	16S rRNA
AB093486	*Tolypothrix* sp. IAM M-259	16S rRNA
HG970653	*Tolypothrix fasciculata* ACOI 3104	16S rRNA
AB325535	*Tolypothrix tenuis* PCC 7101	16S rRNA
KJ920353	*Aulosira laxa* NIES-50	16S rRNA
AP018172	*Calothrix* sp. NIES-2098	16S rRNA
D37876	*Rhodococcus equi* ATCC33701	16S rRNA
AF273613	*Rhodococcus equi* isolate DY1	16S rRNA
MH299445	*Rhodococcus equi* SN26	16S rRNA
KF059848	*Rhodococcus* sp. 209/s	16S rRNA
AB720119	*Proteus mirabilis* BSN1	16S rRNA
KJ578727	*Proteus mirabilis* str. D	16S rRNA
LR134205	*Proteus mirabilis* NCTC4199	16S rRNA
FJ169187	*Sporolituus thermophilus* AeG	16S rRNA

Discussion

One of the most intriguing aspects of *H. pylori* is its genetic diversity at both the genomic level, as well as in terms of homologous genes from different strains, but its biological significance is still not well understood. *H. pylori* is a highly variable microorganism in its genetic content and its influence on gene genesis, horizontal gene transfer and gene loss. Gene variability is observed more frequently in specific regions called plasticity zones, which are distributed along the chromosome [18].

vapD chromosomal region of *H. pylori* is considered a variable zone. This region was first described in 1997 by Cao and Cover [38], who reported that the *vapD* gene was located five ORFs downstream from the *vacA* gene in strain 60190 and was detected in approximately 60% of their strains. Additionally, they reported a high-level of genetic diversity in the corresponding region of *vapD*-negative strains. In the 26695 strain, the chromosomal locus of *vapD* (HP0315) is different from that of 60190, while in the J99 strain, the *vapD* gene is truncated, suggesting that the *vapD* region is prone to genetic changes. It is probable that the *vapD* gene was moved through different mobile genetic elements and inserted at different loci into the *H. pylori* chromosome after being acquired by horizontal gene transfer.

Previous characterization of the *vapD* gene in our Mexican *H. pylori* strains showed a lesser frequency (38%) [39], when compared to other studies [38]. The results of the present study showed a greater variability in the *vapD* region of some of our strains, observing new chromosomal arrangements and new ORFs in this region.

The chromosomal arrangements observed in the *vapD* region in some of our *H. pylori* strains suggested horizontal gene transfer with insertions or deletions of genetic material from plasmids, as well as from other mobile element present in different *H. pylori* strains or other *Helicobacter* species, such as *H. acinonychis* strain Sheeba, as shown in the MxHp21.23a, MxHp84, MxHp54, MxHp249, MxHp248, MxHp2123c strains. Additionally, recombination events were observed between loci from different chromosomes resulting in the inversion of DNA segments, as shown in the strain MxHp118, the nucleotide composition of which showed a high degree of similarity to chromosomal regions of the J99 and 26695 strains. Analysis of the complete, continuous sequence of the *vapD* region in our strains (MxHp21.23a, MxHp84, MxHp54, MxHp249, MxHp248, MxHp21.23c) presented a high degree of nucleotide similarity with two different chromosomal regions of the J99 strain, in which these two regions are joined or separated by the transposase (IS606 and IS605) genes. The role of transposons in the mobility of genetic material is well known; in the J99 strain, the *vapD* ORF is truncated (JHP0829) and encodes for a hypothetical protein of unknown function. In the Mexican strains (Fig. 2), there are only vestiges of *vapD* (30 nucleotides), corresponding to a highly conserved region of the *vapD* gene present in all strains. The extent of variability seen in the *vapD* region suggests that this region is a hotspot in the *H. pylori* chromosome, allowing it to become more dynamic and plastic by incorporating or losing genetic material, depending on what is more conducive to the survival of the bacterium.

The *vapD* gene encodes for the virulence-associated protein D (VapD), which presents homology with other VapD proteins described in other microorganisms, such as *Dichelobacter nodosus* [31,61], *Rhodococcus equi* [28], *Haemophilus influenza* [29], as well as in many other microorganisms of different genera and phyla [42,43,62,63]. In *A. actinomycetencomitans, R. anatipestifer, R. equi, N. gonorrhoeae* and *X. fastidiosa*, *vapD* has been found in plasmids, while in *D. nodosus* it has been found in both chromosomal and plasmid locations. Strains of pathogenic *Rhodococcus equi* contain a virulence plasmid that encodes for VapD and for other virulence-associated proteins. Although its function is unknown, it has been suggested that VapD participates in the survival of *R. equi* within the macrophage, playing a role in acidic tolerance, while *R. equi* plasmid-cured mutants for VapD fail to induce pneumonia in foals [27,28].

Three copies of the *vapD* gene have been identified in *Dichelobacter nodosus*, the essential causative agent of ovine foot rot. Two copies are part of two larger *vap* regions designated vapABCD, while the third *vap* region contains only *vapD*. One of these genes is similar in both length and amino acid sequence to ORF5 found in the *N. gonorrhoeae* cryptic plasmid, and in the *A. actinomycetencomitans* plasmid pVT736-1. The similarity between *vap* encoded-proteins and plasmid-encoded proteins suggests that the *vap* sequences may have evolved from the site-specific insertion of an integrative plasmid [44,64].

In the avian pathogen *Riemerella anatipestifer,* the *vapD* gene was found in two plasmids, pCFC1 and pCFC2. In pCFC2, the *vapD* gene was associated with the insertion sequence ISRa1 [42]. In the case of the phytopathogenic bacteria *Xylella fastidiosa*, the *vapD* gene (XFa0052) is a strain-specific gene that is located in the plasmid pXF51, which does not have IS elements or transposons [65,66]. It has a great similarity to *the* VapD protein of *R. anatipestifer* and *A. actinomycetemcomitans*. Furthermore, the expression of *vapD* is induced when the bacteria is under heat stress [43,62]. However, in both microorganisms its function is still unknown.

In *H. influenzae*, *vapD* forms part of a toxin-antitoxin(TA) module, where *vapD* is known as a toxin that on the one hand, helps to promote non-typeable *H. influenzae* survival within human respiratory cells, while on the other, it enhances virulence during infection using a mechanism of mRNA cleavage [29]. Although its function is still not clear, *H. pylori vapD* has been reported to have endoribonuclease activity [40], and it may be an evolutionary intermediate of the Cas2 protein in the evolution of the CRISPR-Cas system [67]. Despite several studies indicating the importance of VapD in pathogenesis of different bacterial species, there is a lack of molecular, biochemical and functional data describing the biological role of VapD in the literature.

Analysis of nucleotide sequences of the *vapD* region of the Mexican strains also showed remnants of different Insertion Sequence (IS) families (IS4, IS3, IS1595, IS607, IS200, IS605). This suggests that these IS types are reminiscent of mobile elements

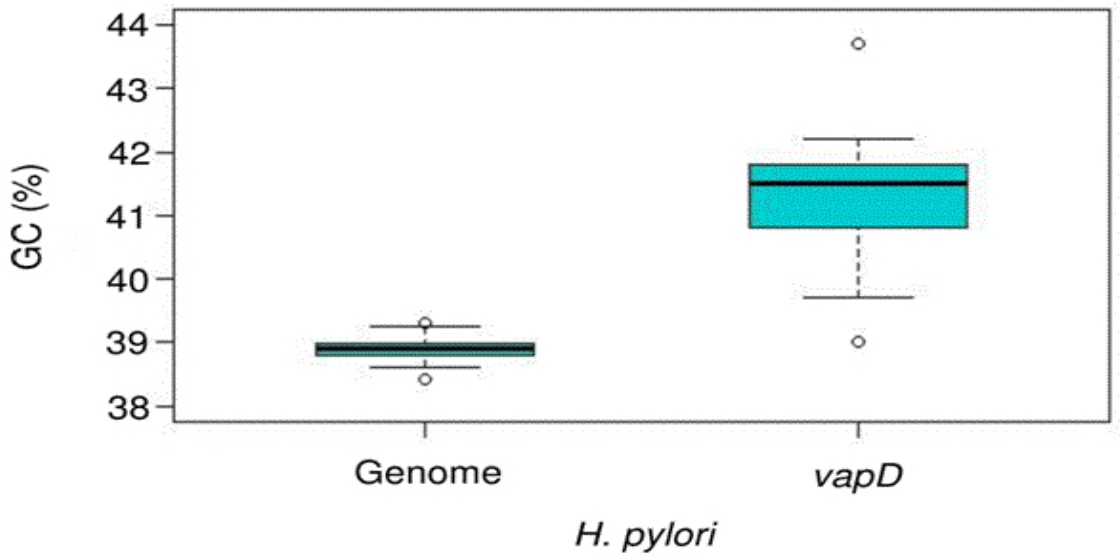

Figure S1: Average GC content of the whole genome of *Helicobacter pylori* strains compared with the average GC content of *H. pylori vapD* gene. There is a marked variation in GC content between the whole genome and *vapD* gene despite the amelioration process to which the gene has been subjected.

Table S2: Codon usage bias comparisons in whole genome and *vapD* gene of *Helicobacter pylori*

		RSCU*		X2
Aminoacid	Codon	Genome	VapD	P value
Phe	UUU	1.58	2	1
	UUC	0.42	0	
Leu	UUA	2.34	0	0.0487
	UUG	1.62	0	
	CCU	0.9	0	
	CUC	0.54	0	
	CUA	0.42	0	
	CUG	0.24	6	
Ile	AUU	1.47	3	0.3581
	AUC	1.14	0	
	AUA	0.39	0	
Val	GUU	1.08	0	0.4338
	GUC	0.56	0	
	GUA	0.4	0	
	GUG	1.96	4	
Ser	UCU	1.38	0	0.3987
	UCC	0.48	0	
	UCA	0.54	0	
	UCG	0.36	0	
	AGU	0.84	0	
	AGC	2.4	6	
Pro	CCU	1.96	0	0.0879
	CCC	1.04	0	
	CCA	0.6	0	
	CCG	0.4	4	
Thr	ACU	1.24	0	0.2446
	ACC	1.28	4	
	ACA	0.64	0	
	ACG	0.88	0	
Ala	GCU	1.56	0	0.2301
	GCC	0.8	0	
	GCA	0.44	0	
	GCG	1.2	4	
Tyr	UAU	1.4	2	1
	UAC	0.6	0	
Gln	CAA	1.7	0	1
	CAG	0.3	2	
Asn	AAU	1.14	0	1
	AAC	0.86	2	
Lys	AAA	1.54	2	1
	AAG	0.46	0	
Asp	GAU	1.46	2	1
	GAC	0.54	0	
Glu	GAA	1.46	2	1
	GAG	0.54	0	

Arg	CGU	0.84	6	0.107
	CGC	1.44	0	
	CGA	0.42	0	
	CGG	0.18	0	
	AGA	1.62	0	
	AGG	1.5	0	
Gly	GGU	0.68	0	0.2779
	GGC	1.4	4	
	GGA	0.44	0	
	GGG	1.48	0	
* Relative Synonymous Codon Usage				

that participated in the transfer of genetic material, yielding an important impact on genome architecture and function. The ISs play an important role in prokaryote chromosomes, promoting gene inactivation or modulation of neighbouring gene expression, as well as promoting foreign DNA insertion, thereby increasing genome diversity and plasticity [55]. The presence of IS remnants suggests ancestral horizontal gene transfer in our strains. In the J99 strain, we observed the presence of the complete IS606 (jhp0827 and jhp0826 loci) with its *tnpA* and *tnpB* genes, and the formation of different ORFs (JHP0829, JHP0828, JHP0827, JHP0826, JHP0825) when compared with those found in the *vapD* region of our strains. However, there were strains, including the 60190 strain, that presented a complete *vapD* ORF and did not contain any ISs or traces of IS in the *vapD* region (flanked by D1 and D2 primers set). Although the MxHp563 strain presented a complete *vapD* ORF, this strain presented traces of ISs (IS *200*/IS *605* and IS *4*) in its *vapD* region and a DNA fragment towards its 5´ end with high homology to a DNA fragment of the J99 strain. It is likely that the MxHp563 strain is an intermediate strain between those strains with the complete *vapD* gene and those strains that are losing it. Its presence in these strains could aid in the survival of *H. pylori* inside gastric cells, as seen previously [3], which indicates that *vapD* is transcribed only when there is *H. pylori*-epithelial cell interaction. Its transcription is elevated in severe gastric pathologies, reaching the highest expression levels in patients above 57 years of age, demonstrating that *vapD* is overexpressed during chronic infection. This may suggest that *vapD* gene detection in *Helicobacter pylori* strains could be associated with chronic infection.

The genealogical relationship of the *vapD* sequences from 72 *H. pylori* strains isolated in Mexico and other parts of the world showed 2 main clades (A and B), where the majority of the strains presented little mutational changes, some signals of recombination between them, and slight alteration in the composition of amino acids of the VapD protein (data not shown), as such changes are seen as synonymous substitutions. Synonymous substitutions do not alter the encoded amino acid sequences, since they are almost neutral with respect to fitness and they are not affected by natural selection [68], suggesting that this gene is subject to purifying selection. This makes *vapD* an interesting gene, since it behaves in a similar fashion to an essential gene, probably because it is necessary for intracellular strains. The designation of the term "essential" is related to functional significance [68]. It is striking that the *vapD* gene of the Mexican strains, which apparently encode for a functional protein, were grouped together into a single group (B) with 60190 and 26695 (hspEuropeN) strains, whereas most strains isolated from other parts of the world and belonging to different genetic subpopulations clustered in group A. We identified two further clusters, C (includes isolates from hspEuropeN, hspEuropeS and hspEasia subpopulations) and D (integrated by MxHp strains and the J99 strain [hspWAfrica]), which apparently diverged from cluster B over 44 and 73 mutational steps, respectively (Fig. 3). Analysis of the VapD protein from the strains that were grouped into clade C showed a protein without function due to the insertion of various stop codons throughout the open reading frame. Moreover, the haplotype network showed that clade B is more stable, recombination is low between the strains, and there are few diversification traces. Unfortunately, we cannot verify if the strains belonging to clade A had diverged from the strains of clade B or vice versa. However, clade B strains position themselves in the network of haplotypes suggesting that clade A had diverged from these strains (clade B). Contrary to clade B, the strains of clade A presented greater diversification and signals of high recombination rates of the *vapD* gene leading to many haplotypes. The low diversity and purifying selection detected in the *vapD* gene among the MxHp strains, and their divergence from the rest of the isolates, could be attributed to two genetic events, namely recombination and genetic drift. Several studies have suggested that *H. pylori* was introduced to the Americas from Europe and Africa [9,69]. Since then, *H. pylori* may have evolved alongside its host, generating new independent evolutionary lineages in Latin America [8,69,70]. Finally, there was greater diversification through deletions, insertions, and recombination, giving rise to clades C and D. Clade D strains had important genetic events in the *vapD* region, as seen in its chromosomal reorganization and formation of new ORFs, and through the elimination of the *vapD* gene or of truncated ORF, likely resulting in a different phenotype.

The relationship between the phylogenies of 16S rRNA and *vapD* genes showed substantially different topologies. While the 16S rRNA phylogenetic tree is coherent for species, genus and phylum, the *vapD* tree reveals a conflicting topology or phylogenetic inconsistency revealing different evolutionary histories. The 16S rRNA gene has enormous repercussions for inferences about phylogenetic relationships among bacteria and in their taxonomy. This gene has relevant characteristics that make it a widely used reference tool in this kind of studies. Some of these unique characteristics include the fact that bacterial genomes contain several copies and more than 99.5% identity between them; it is a very old molecule, present in all current bacteria; and the changes in the sequence occur very slowly, thereby providing information about all prokaryotes along the evolutionary scale. While 16S rRNA genes have high identity between them within the same species, they have enough variability to differentiate between different species and genera [58,71-73], so they have always been used for reference.

However, this harmonious phylogeny is not always reproduced, as has been seen in phylogenies based on many other genes, resulting in conflicting topologies or phylogenetic inconsistencies [74,75]. There are biological factors that may cause these phylogenetic incongruities of a specific gene, especially among different bacteria domains, which may be the result of different homologous recombination or horizontal gene transfer events (HGT), or incomplete lineage sorting through species, genera or phyla. Regarding 16S rRNA and *vapD* phylogeny of *H. pylori*, our results showed substantially different topologies between both trees, which revealed different evolutionary histories for both genes. Focusing more closely on the *vapD* tree, we found that the *vapD* sequences from *Helicobacter* species are more closely related to the *vapD* gene of different phyla, such as Fusobacterium, Actinomycetales or Firmicutes.

GC content analysis of *vapD* showed a relatively small difference in the content of GC between the whole genome and the *vapD* sequences. This could be due to the amelioration of the *vapD* gene over time, making the ancient HGT event even more difficult to detect [47,76]. Our results showed that this gene was inserted and fixed in the host genome over time. Analysis of codon-usage provides insights into the history of genes in a genome. These genes can differ drastically from native genes. Our results showed that while codon usage is compatible between the genome and the *vapD* gene, only the codon for leucine had a different preferential codon usage, indicating that *vapD* translates efficiently with the machinery of the host genome. This event occurs throughout successful generations with the *vapD* gene and when the host genome is placed under the same selection and mutational pressures, resulting in the homogenization of nucleotide composition and codon usage due to amelioration.

Conclusion

This study documents the evolutionary dynamics of the *vapD gene of H. pylori* which in other microorganisms has been shown to participate in the virulence of infectious diseases with a high rate of horizontal transfer through mobile elements. In *H. pylori*, the *vapD* gene not only fixes on the host chromosome, but it is also functional, allowing the bacterium to acquire adaptive characteristics from other microorganisms, independently of its phylogenetic distance. It is not surprising to find alleles of a gene or specific genes from strains associated with virulence, such as the s1-m1 genotype of *vacA* and the *cag*-PAI pathogenicity island, where a clear association has been found between the *vacA* genotype and the presence of *cag*-PAI with the development of gastric pathology. The presence of *vapD* in different strains of *H. pylori* and its expression in gastric biopsies [3] suggest that *vapD* is a strain-specific gene, that could confer particular characteristics on survival in a stressful environment, as has been described in *R. equi* and *H. influenzae.*

Funding and Acknowledgments

This work was supported by DGAPA-PAPIIT grant IN213816 and CONACYT grant CB-255574. We thank Dr Carlos Alberto Santiago Olivares and Dr Luisa Sandner-Miranda for helpful discussions.

Declarations

Conflict of interest statement: The authors declare that they have no competing interests

Ethical approval: NA

References

1. Kao CY, Sheu BS, Wu JJ. *Helicobacter pylori* infection: An overview of bacterial virulence factors and pathogenesis. Biomed J. 2016; 39(1):14–23.doi:10.1016/j.bj.2015.06.002.
2. Šterbenc A, Jarc E, Poljak M, Homan M. *Helicobacter pylori* virulence genes. World J Gastroenterol. 2019; 25(33):4870–4884.doi: 10.3748/wjg.v25.i33.4870.
3. Morales-Espinosa R, Delgado G, Serrano LR, Castillo E, Santiago CA, Hernández-Castro R, et al. High expression of *Helicobacter pylori* VapD in both the intracellular environment and biopsies from gastric patients with severity. PLoS One. 2020;15(3):1–16. doi: 10.1371/journal.pone.0230220.
4. Suerbaum S, Smith JM, Bapumia K, Morelli G, Smith NH, Kunstmann E, et al. Free recombination within *Helicobacter pylori*. Proc Natl Acad Sci U S A. 1998;95(21):12619–12624.doi:10.1073/pnas.95.21.12619.
5. Israel DA, Lou AS, Blaser MJ. Characteristics of *Helicobacter pylor*i natural transformation. FEMS Microbiol Lett. 2000;186(2): 275–280. doi:10.1016/S0378-1097(00)00156-7.
6. Yahara K, Kawai M, Furuta Y, Takahashi N, Handa N, Tsuru T, et al. Genome-wide survey of mutual homologous recombination in a highly sexual bacterial species. Genome Biol Evol. 2012;4(5):628–640. doi:10.1093/gbe/evs043.
7. Yahara K, Furuta Y, Oshima K, Yoshida M, Azuma T, Hattori M, et al. Chromosome painting in silico in a bacterial species reveals fine population structure. Mol Biol Evol. 2013;30(6):1454–1464. doi:10.1093/molbev/mst055.
8. Thorell K, Yahara K, Berthenet E, Lawson DJ, Mikhail J, Kato I, et al. Rapid evolution of distinct *Helicobacter pylori* subpopulations in the Americas. PLoS Genet. 2017;13(2):e1006546.doi:10.1371/journal.pgen.1006546.
9. Falush D, Kraft C, Taylor NS, Correa P, Fox JG, Achtman M, et al. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. Proc Natl Acad Sci U S A. 2001;98(26):15056–15061. doi:10.1073/pnas.251396098.
10. Dorer M S, Sessler Tate H, SNR. Recombination and DNA Repair in *Helicobacter pylori.* Annu Rev Microbiol. 2011;65(1):329–348. doi:10.1146/annurev-micro-090110-102931.
11. Delgado-Sapién G, Rene Cerritos, Gerardo González-Valencia, Mendez Jose L, Cravioto, Alejandro Morales-Espinosa R. Novel *VacA* s1d signal sequence found in *Helicobacter pylori* from Mexican Children with Recurrent Abdominal Pain. SOJ Microbiol Infect Dis. 2017;5(1):1–7. doi:10.15226/sojmid/5/1/00168.
12. Furuta Y, Kawai M, Yahara K, Takahashi N, Handa N, Tsuru T, et al. Birth and death of genes linked to chromosomal inversion. Proc Natl Acad Sci U S A. 2011;108(4):1501–1506.doi:10.1073/pnas.1012579108.
13. Tada I, Tanizawa Y, Arita M. Visualization of consensus genome structure without using a reference genome. BMC Genomics. 2017; 18(Suppl 2):1–9.doi10.1186/s12864-017-3499-7.

14. Janssen PJ, Audit B, Ouzounis CA. Strain-specific genes of *Helicobacter pylori*: distribution, function and dynamics. Nucleic Acids Res. 2001;29(21):4395–4404.doi:10.1093/nar/29.21.4395.
15. Uchiyama I, Albritton J, Fukuyo M, Kojima KK, Yahara K, Kobayashi I. A Novel Approach to *Helicobacter pylori* Pan-Genome analysis for Identification of genomic Islands. PLoS One. 2016;11(8):1–27.doi: 10.1371/journal.pone.0159419.
16. Fischer W, Breithaupt U, Kern B, Smith SI, Spicher C, Haas R. A comprehensive analysis of *Helicobacter pylori* plasticity zones reveals that they are integrating conjugative elements with intermediate integration specificity. BMC Genomics. 2014;15: 310. doi: 10.1186/1471-2164-15-310.
17. Miehlke S, Kibler K, Kim JG, Figura N, Small SM, Graham DY, et al. Allelic variation in the *cagA* gene of *Helicobacter pylori* obtained from Korea compared to the United States. Am J Gastroenterol. 1996; 91(7):1322-1325.
18. Marshall DG, Dundon WG, Beesley SM, Smyth CJ. H*elicobacter pylori* - A conundrum of genetic diversity. Microbiology. 1998;144(11):2925–2939.doi: 10.1099/00221287-144-11-2925.
19. Berthenet E, Yahara K, Thorell K, Pascoe B, Meric G, Mikhail JM, et al. A GWAS on *Helicobacter pylori* strains points to genetic variants associated with gastric cancer risk. BMC Biol. 2018;16(1):84.doi: 10.1186/s12915-018-0550-3.
20. Chang WL, Yeh YC, Sheu BS. The impacts of *H. pylori* virulence factors on the development of gastroduodenal diseases. J Biomed Sci. 2018;25:68.doi: 10.1186/s12929-018-0466-9.
21. Morales-Espinosa R, Castillo-Rojas G, Gonzalez-Valencia G, Ponce de León S, Cravioto A, Atherton JC, et al. Colonization of Mexican patients by multiple *Helicobacter pylori* strains with different *vacA* and *cagA* genotypes. J Clin Microbiol. 1999;37(9):3001–3004. doi:10.1128/JCM.37.9.3001-3004.1999
22. Sugimoto M and Yamaoka Y. The association of *vacA* genotype and *Helicobacter pylori*- related disease in Latin American and African populations. Clin Microbiol Infect. 2009;15(9):835–842.doi:10.1111/j.1469-0691.2009.02769.x.
23. Salama NR, Hartung ML, Müller A. Life in the human stomach: persistence strategies of the bacterial pathogen *Helicobacter pylori.* Nat Rev Microbiol. 2013;11(6):385–399.doi: 10.1038/nrmicro3016.
24. Gupta N, Maurya S, Verma H, Verma VK. Unraveling the factors and mechanism involved in persistence: Host-pathogen interactions in *Helicobacter pylori.* J Cell Biochem. 2019;120(11):18572–18587. doi:10.1002/jcb.29201.
25. Ko GH, Kang SM, Kim YK, Lee JH, Park CK, Youn HS, et al. Invasiveness of *Helicobacter pylori* into human gastric mucosa. *Helicobacter*. 1999;4(2):77–81.doi:10.1046/j.1523-5378.1999.98690.x.
26. Dubois A and Borén T. *Helicobacter pylori* is invasive and it may be a facultative intracellular organism. Cell Microbiol. 2007;9(5):1108–1116.doi:10.1111/j.1462-5822.2007.00921.x.
27. Benoit S, Benachour A, Taouji S, Auffray Y, Hartke A. Induction of *vap* genes encoded by the virulence plasmid of *Rhodococcus equi* during acid tolerance response. Res Microbiol. 2001;152(5):439–449. doi:10.1016/S0923-2508(01)01217-7.
28. Benoit S, Benachour A, Taouji S, Auffray Y, Hartke A. $H_{202,}$ which causes macrophage-related stress, triggers induction of expression of virulence-associated plasmid determinants in *Rhodococcus equi.* Infect Immun.2002;70(7):3768–3776.doi: 10.1128/IAI.70.7.3768-3776.2002.
29. Ren D, Walker AN, Daines DA. Toxin-antitoxin loci *vapBC*-1 and *vapXD* contribute to survival and virulence in nontypeable *Haemophilus influenzae*. BMC Microbiol. 2012;12:263.doi:10.1186/1471-2180-12-263.
30. Katz ME, Strugnell RA, Rood JI. Molecular characterization of a genomic region associated with virulence in *Dichelobacter nodosus*. Infect Immun. 1992;60(11):4586–4592.
31. Katz ME, Wright CL, Gartside TS, Cheetham BF, Doidge C V, Moses EK, et al. Genetic organization of the duplicated *vap* region of the *Dichelobacter nodosus* genome. J Bacteriol. 1994;176(9):2663–2669. doi: 10.1128/jb.176.9.2663-2669.1994.
32. Gerdes K and Maisonneuve E. Bacterial Persistence and Toxin-Antitoxin Loci. Annu Rev Microbiol.2012;66(1):103–123.doi:10.1146/annurev-micro-092611-150159.
33. Mruk I and Kobayashi I. To be or not to be: Regulation of restriction-modification systems and other toxin-antitoxin systems. Nucleic Acids Res. 2014;42(1):70–86.doi: 10.1093/nar/gkt711.
34. Rocker A and Meinhart A. Type II toxin: antitoxin systems. More than small selfish entities? Curr Genet. 2016;62(2):287–290. doi:10.1007/s00294-015-0541-7.
35. Burbank LP and Stenger DC. The DinJ/RelE Toxin-Antitoxin System Suppresses Bacterial Proliferation and Virulence of *Xylella fastidiosa* in Grapevine. Phytopathology. 2017;107(4):388–394.doi:10.1094/PHYTO-10-16-0374-R.
36. Michiels JE, Van den Bergh B, Verstraeten N, Michiels J. Molecular mechanisms and clinical implications of bacterial persistence. Drug Resist Updat. 2016; 29:76–89.doi: 10.1016/j.drup.2016.10.002.
37. Christensen-Dalsgaard M, Jørgensen MG, Gerdes K. Three new RelE-homologous mRNA interferases of *Escherichia coli* differ- entially induced by environmental stresses. Mol Microbiol. 2010; 75(2):333–348. doi: 10.1111/j.1365-2958.2009.06969.x.
38. Cao P and Cover TL. High-level genetic diversity in the *vapD* chromosomal region of *Helicobacter pylori.* J Bacteriol. 1997;179(9):2852–2856.doi:10.1128/jb.179.9.2852-2856.1997.
39. Morales-espinosa R, González-valencia G, Delgado G, Luis J, Torres J, Cravioto A. Frequency and characterization of *vapD* gene in *Helicobacter pylori* strains of different *vacA* and cag-PAI genotype. Bioquimia. 2008;33(2):43–50.
40. Kwon AR, Kim JH, Park SJ, Lee KY, Min YH, Im H, et al. Structural and biochemical characterization of HP0315 from *Helicobacter pylori* as a *VapD* protein with an endoribonuclease activity. Nucleic Acids Res. 2012;40(9):4216–4228.doi: 10.1093/ nar/ gkr1 305.
41. Bangpanwimon K, Sottisuporn J, Mittraparp-Arthorn P, Ueaphatthanaphanich W, Rattanasupar A, Pourcel C, et al. CRISPR-like sequences in *Helicobacter pylori* and application in genotyping. Gut Pathog. 2017;9:65.doi:10.1186/s13099-017-0215-8.
42. Weng SC, Lin WH, Chang YF, Chang CF. Identification of a virulence-associated protein homolog gene and ISRa1 in a plasmid of *Riemerella anatipestifer.* FEMS Microbiol Lett. 1999;179(1):11–19. doi:10.1111/j.1574-6968.1999.tb08701.x.
43. Catani CF, Azzoni AR, Paula DP, Tada SFS, Rosselli LK, De Souza AP, et

al. Cloning, expression, and purification of the virulence-associated protein D from *Xylella fastidiosa*. Protein Expr Purif. 2004;37(2):320–326.doi:10.1016/j.pep.2004.07.002.

44. Cheetham BF, Tattersall DB, Bloomfield GA, Rood JI, Katz ME. Identification of a gene encoding a bacteriophage-related integrase in a *vap* region of the *Dichelobacter nodosus* genome. Gene. 1995;162(1):53–58.doi:10.1016/0378-1119(95)00315-W.
45. Clement M, Posada D, Crandall KA. TCS: a computer program to estimate gene genealogies. Mol Ecol. 2000;9(10):1657–1659.doi: 10.1046/j.1365-294x.2000.01020.x.
46. Fournier GP, Andam CP, Gogarten JP. Ancient horizontal gene transfer and the last common ancestors Phylogenetics and phylogeography. BMC Evol Biol. 2015;15:70.doi:10.1186/s12862-015-0350-0.
47. Ravenhall M, Škunca N, Lassalle F, Dessimoz C. Inferring Horizontal Gene Transfer. PLoS Comput Biol. 2015;11(5): e1004095.doi: 10.1371/journal.pcbi.1004095.
48. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–1797.doi: 10.1093/nar/gkh340.
49. Darriba D, Taboada GL, Doallo RPD. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 2012;9(8):772. doi: 10.1038/nmeth.2109.
50. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, et al. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539–542.doi:10.1093/sysbio/sys029.
51. Rambaut, A. FigTree v1.4.3.
52. Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res. 2000;28(1):292.doi:10.1093/nar/28.1.292.
53. Librado P and Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11):1451–1452. doi:10.1093/bioinformatics/btp187.
54. J.J. Allaire. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA.
55. P Siguier, J Perochon, L Lestrade, J Mahillon, M Chandler. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res [Internet]. 2006;34:D32–D36.doi:10.1093/nar/gkj014.
56. Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, et al. Traces of human migrations in *Helicobacter pylori* populations. Science. 2003;299(5612):1582–1585.doi:10.1126/science.1080857.
57. Waskito LA and Yamaoka Y. The story of *Helicobacter pylori*: depicting human migration from the phylogeography. Adv Exp Med Biol. 2019;1149:1-16.doi: 10.1007/5584_2019_356.
58. Olsen GJ and Woese CR. Ribosomal RNA: a key to phylogeny. FASEB J. 1993;7(1):113–123.doi: 10.1096/fasebj.7.1.8422957.
59. Clarridge JE. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev. 2004;17(4):840–862.doi: 10.1128/CMR.17.4.840-862.2004.
60. Rajendhran J and Gunasekaran P. Microbial phylogeny and diversity: Small subunit ribosomal RNA sequence analysis and beyond. Microbiol Res. 2011;166(2):99–110.doi: 10.1016/j.micres.2010.02.003.
61. Billington SJ, Huggins AS, Johanesen PA, Crellin PK, Cheung JK, Katz ME, et al. Complete nucleotide sequence of the 27-kilobase virulence related locus (vrl) of *Dichelobacter nodosus*: Evidence for extrachromosomal origin. Infect Immun. 1999;67(3):1277–1286.doi: 10.1128 / IAI.67.3.1277-1286.1999.
62. Koide T, Vêncio RZN, Gomes SL. Global gene expression analysis of the heat shock response in the phytopathogen *Xylella fastidiosa*. J Bacteriol. 2006;188(16):5821–5830.doi:10.1128/JB.00182-06.
63. Loy A, Pfann C, Steinberger M, Hanson B, Herp S, Brugiroux S, et al. Lifestyle and horizontal gene transfer-mediated evolution of *Mucispirillum schaedleri*, a core member of the murine gut microbiota. mSystems. 2017;2(1):e00171-e001716.doi: 10.1128/msystems.00171-16.
64. Billington SJ, Johnston JL, Rood JI. Virulence regions and virulence factors of the ovine footrot pathogen, *Dichelobacter nodosus*. FEMS Microbiol Lett. 1996;145(2):147–156.doi:10.1016/S0378-1097(96)00379-5.
65. Marques MV. Da Silva AM, Gomes SL. Genetic organization of plasmid pXF51 from the plant pathogen *Xylella fastidiosa*. Plasmid. 2001;45(3):184–199.doi:10.1006/plas.2000.1514.
66. Nunes LR, Rosato YB, Muto NH, Yanai GM, Silva VS, Leite DB, et al. Microarray Analyses of *Xylella fastidiosa* Provide Evidence of Coordinated Transcription Control of Laterally Transferred Elements. Genome Res. 2003;13(4):570–578.doi:10.1101/gr.930 803.
67. Koonin EV, Makarova KS. Mobile genetic elements and evolution of CRISPR-Cas systems: All the way there and back. Genome Biol Evol. 2017;9(10):2812–2825.doi:10.1093/gbe/evx192.
68. Jordan IK, Rogozin IB, Wolf YI, Koonin E V. Essential Genes Are More Evolutionarily Conserved Than Are Nonessential Genes in Bacteria. Genome Res. 2002;12(6):962–968.doi:10.1101/gr.87702.
69. Oleastro M, Rocha R, Vale FF. Population genetic structure of *Helicobacter pylori* strains from Portuguese-speaking countries. *Helicobacter*. 2017;22(4): e12382. doi: 10.1111/hel.12382.
70. Muñoz-Ramírez ZY, Mendez-Tenorio A, Kato I, Bravo MM, Rizzato C, Kaisa T, et al. Whole Genome Sequence and Phylogenetic Analysis Show *Helicobacter pylori* Strains from Latin America Have Followed a Unique Evolution Pathway. Front Cell Infect Microbiol. 2017;7:50. doi:10.3389/fcimb.2017.00050.
71. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A. 1977;74(11):5088–5090.doi:10.1073/pnas.87.12.4576.
72. Woese CR. Bacterial Evolution. Microbiol Rev. 1987;51(2):221–271. doi:10.1139/m88-093.
73. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990;87(12):4576–4579.doi:10.1073/pnas.87.12.4576.
74. Garcia-Vallvé S, Romeu A, Palau J. Horizontal gene transfer in bacterial and archaeal complete genomes. Genome Res. 2000;10(11):1719–1725.doi:10.1101/gr.130000.
75. Galtier N and Daubin V. Dealing with incongruence in phylogenomic analyses. Philos Trans R Soc Lond B Biol Sci. 2008;363(1512):4023–4029.doi:10.1098/rstb.2008.0144.

76. Brown JR. Ancient horizontal gene transfer. Nat Rev Genet. 2003;4(2):121–132.doi:10.1038/nrg1000.
77. Marri PR and Golding GB. Gene amelioration demonstrated: the journey of nascent genes in bacteria. Genome. 2008;51(2):164–168. doi: 10.1139/G07-105.