Journal of Proteomics & Bioinformatics - Open Access

Research Article JPB/Vol.1/August 2008

Genomic Methylation Status for Discrimination Among *Helicobacter* Species: A Bioinformatics Approach

Filipa F. Vale[1]*, Jorge M. B. Vítor[2]

[1]Engineering Faculty, Catholic University of Portugal, Estrada Octávio Pato, 2635-631 Rio de Mouro, Portugal
[2] iMed.UL (CECF), Faculty of Pharmacy, University of Lisbon, Av. Forças Armadas, 1649-003 Lisboa, Portugal

*Corresponding author: Filipa F. Vale. Faculty of Engineering, Catholic University of Portugal, Estrada Octávio Pato, 2635-631 Rio de Mouro, Portugal, Tel: +351.214269824; Fax: +351.214269800. 21; E-mail: filipavale@fe.lisboa.ucp

Received July 10, 2008; Accepted 08, August 2008; Published August 14, 2008

Citation: Filipa FV, Jorge MBV (2008) Genomic Methylation Status for Discrimination Among *Helicobacter* Species: A Bioinformatics Approach. J Proteomics Bioinform 1: 258-266. doi:10.4172/jpb.1000033

Abstract

The genus *Helicobacter* comprises several species of both gastric and enterohepatic intestinal bacteria. *H. pylori*, the type species of the genus, is associated with gastritis, peptic ulcer and gastric cancer in humans. *H. pylori* genome has a high number of restriction and modification (R-M) systems and their diversity is useful for strain typing. To analyse if such a high number of expressed methyltransferases is a characteristic of the genus *Helicobacter*, the genomic methylation of five non-pylori *Helicobacter* spp. (*H. canadensis*, *H. canis*, *H. felis*, *H. mustelae* and *H. pullorum*) was determined. The results revealed that the number of R-M systems among non-pylori *Helicobacter* spp. is smaller than those observed among a group of 221 *H. pylori* strains (p<0,001), but is greater than those observed for the mean of all bacteria sequenced genomes (p=0,005). 16S ribosomal RNA analysis of *H. pylori* sequenced strains and five non-pylori *Helicobacter* spp. clearly isolate *H. pylori* species. Surprisingly, the analysis of the genomic methylation status by MCRM algorithm performs similarly. This suggests that R-M systems do not appear to be spread in a miscellaneous manner, once even that these genes may be subjected to acquisition and loss; their expression still allows discriminating among *Helicobacter* spp.

Keywords: MCRM algorithm; Genomic methylation; *Helicobacter* spp.; Bioinformatics; Restriction and modification ribosomal RNA

Abbreviations

MCRM: Minimum Common Restriction and Modification; REase: Restriction Endonuclease; MTase: Methyltransferase; RM: Restriction and Modification.

Introduction

Helicobacter pylori colonizes the stomach of about half of the human population and is associated with several disease outcomes, like gastritis, peptic ulcer and gastric cancer (Dunn et al., 1997; Kusters et al., 2006). Other similar spiral bacteria have been isolated from several animals, like cat, dog or mice, among others. *Helicobacter* species can be subdivided into gastric *Helicobacter* species and enterohepatic (nongastric) *Helicobacter* species. The two lineages demonstrate a high level of organ specificity, such that gastric *Helicobacter* spp. in general does not colonize the intestine or liver, and vice versa (Kusters et al., 2006).

The genomic DNA of *H. pylori* is characterised by the presence of a high unusual number of restriction and modification (R-M) system (Nobusato et al., 2000; Lin et al., 2001; Takata et al., 2002). The type II R-M systems are composed by least two genes: one coding for a restriction endonuclease (REase) that recognizes a specific DNA sequence and cuts both strands; and other gene coding for a cognate MTase that methylates the same DNA sequence, thus protecting the genomic DNA from being cleaved by the companion REase (Roberts et al., 2003). Type II R-M systems have been referred as selfish genetic elements, because the descendants of cells that had lost these genes appeared unable to modify a sufficient number of recogni-

tion sites in their chromosomes to protect them from lethal attack by the remaining restriction enzyme molecules (Naito et al., 1995).

Recently we have demonstrated that the diversity of R-M systems among *H. pylori* strains is high enough to be used as a typing method (Vale and Vitor, 2007). Cluster analysis by conventional methods does not consider the propensity for R-M systems conservation after acquisition, due to the "selfish behavior" (Naito et al., 1995). Considering this we have recently developed a new clustering algorithm [Minimum Common Restriction Modification (MCRM) algorithm] that takes into account the pressure of REases on MTases, and that is based on the hypothesis that each strain evolves by acquiring new RM systems without loosing acquired RM systems (Vale et al., 2008). In this algorithm it is considered that: i) the strain with less RM systems is the one that has the core set of the most abundant R-M systems expressed among the typed strains; ii) these core set of R-M systems was the first to be acquired by *H. pylori*, so that they exhibit a large dissemination (expression) among several daughter strains (Vale et al., 2008). MCRM analysis of the genomic methylation data from *H. pylori* strains isolated from different geographic revealed a clustering according to strain's continent of origin (Vale et al., 2008), which is in agreement with *H. pylori* coevolution with its human host (Covacci et al., 1999; Linz et al., 2007; Vale et al., 2008). This observation led to the suggestion that R-M systems may trace *H. pylori* geographic distribution and, by default also its human host migrations (Vale et al., 2008).

In this study it was investigated if non-pylori *Helicobacter* strains also have a high number expressed MTases and, if the genomic methylation status followed be MCRM algorithm analysis permitted to discriminate between *H. pylori* and non-pylori *Helicobacter* spp. (*H. canadensis*, *H. canis*, *H. felis*, *H. mustelae* and *H. pullorum*). Following, these results were compared with the phylogenetic analysis of 16S rRNA gene sequences for the same species. To our knowledge this is the first study that systematically analysis the diversity of expression of R-M systems in *H. canadensis*, *H. canis*, *H. felis*, *H. mustelae* and *H. pullorum*.

Material and Methods

Helicobacter spp

H. pylori strains (26695 and J99) were cultured on *H. pylori* selective agar (Wilkins-Chalgren agar supplemented with 10% horse blood, vancomycin [10 mg liter^{-1}], cefsulodin [5 mg liter^{-1}], trimethoprim [5 mg liter^{-1}], and cycloheximide [100 mg liter^{-1}] [Biogerm, Porto, Portugal]) and incubated at 37°C for 48 h in an anaerobic jar (Oxoid, UK; or BBL, USA) with a gas generator system (CampyGen; Oxoid, UK) (Megraud, 1996). Non-pylori *Helicobacter* spp. were cultured on Muller Hinton agar (Oxoid, UK) supplemented with 10% (v/v) defibrinated horse blood (Probiologica, Portugal) and incubated in similar conditions. Genomic DNA was extracted by standard methods.

Helicobacter spp. R-M systems diversity

To evaluate the expression of the cognate methyltransferase, the genomic DNA was digested with 27 REases [AciI, AseI, BseRI, BssHII, BstUI, DdeI, DpnI, DpnII, DraI, EagI, FauI, Fnu4HI, FokI, HaeIII, HhaI, Hpy188I, Hpy188III, Hpy99I, HpyCH4III, HpyCH4IV, HpyCH4V, MspI, NaeI, NlaIII, Sau96I, ScrFI, and TaqI (New England Biolabs, USA)]. The results were coded as "0" for digestion observed (DNA is unmethylated), and "1" for absence of digestion, suggesting an active methyltransferase (Vale and Vitor, 2007).

Genomic methylation status comparison

The mean number of active MTases on non-pylori *Helicobacter* spp. was compared with: i) the mean number of expressed MTases of 221 *H. pylori* tested by us; ii) the mean number of M genes predicted by REBASE for all sequenced bacteria (Roberts et al., 2007). The Kruskal-Wallis test was performed using the statistical package SPSS v.15 (SPSS Inc., Chicago, IL).

Ribosomal RNA alignment

16S rRNA sequences available on public data bases from the *Helicobacter* spp. (table 1) were aligned using ClustalW, producing a cladogram (Chenna et al., 2003).

MCRM clustering

Several dendrograms were produced using MCRM algorithm after genomic methylation analysis of 7 *Helicobacter* spp. Most of the dendrograms produced by MCRM algorithm are indeed similar, but it is possible that distinct dendrograms are generated as different choices of strain or R-M system at ties may result in different clustering. Thus, 10 different dendrograms were produced from the same data in order to increase the confidence on the clustering results.

Strain	Classification	Number of expressed type II MTases	Number of MTases predicted by REBASE	16S rRNA Locus
H. pylori J99	gastric	12	21	jhpr3
H. pylori 26695	gastric	9	20	HPrrnA16S
H. pylori HPAG1	gastric	ND	20	HPAG1_r06
H. pylori Shi 470	gastric	ND	24	HPSH_r08358
H. mustelae	gastric	6	ND	M88156
H. canadensis	enterohepatic	8	ND	DQ438122
H. canis	enterohepatic	12	ND	EU144018
H. felis	gastric	8	ND	AY366427
H. pullorum	enterohepatic	6	ND	AY631956

ND – Not determined

Table 1: *Helicobacter* species used in the present work.

Results

Helicobacter spp. genomic methylation

After genomic DNA hydrolysis with the selected REases it was observed that among non-pylori *Helicobacter* spp. the mean number of expressed MTases was 8 (SD=2.4). A similar analysis for 221 *H. pylori* strains (Vale et al. (2008) unpublished results) revealed a mean of 17 active MTases (SD=3.4). The mean number of genes coding for methyltransferases for the overall 862 sequenced genomes is 4.2 (SD=5.0) [data from REBASE (Roberts et al., 2007)]. Table 1 resumes the number of active methyltransferases in tested *Helicobacter* spp.

A significant statistical difference between mean number of active MTases from non-pylori *Helicobacter* spp. and *H. pylori* ($p<0,001$) and also between non-pylori *Helicobacter* spp. and the overall sequenced genomes ($p=0,005$) was verified. Present study results showed that the number of expressed MTases in a decreasing order by organism, or group of organisms, was: *H. pylori*, non-pylori *Helicobacter* and all sequenced bacteria available at REBASE (Roberts et al., 2007).

Helicobacter spp. 16S rRNA cladogram

Construction of a cladogram after multiple sequence

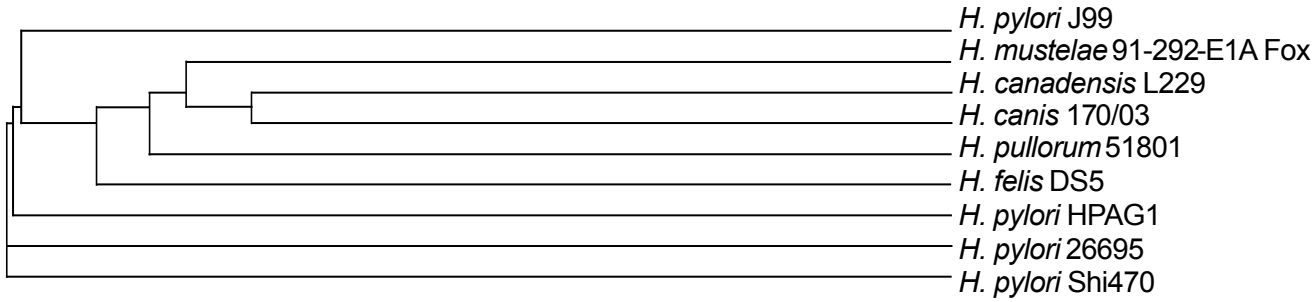

Figure 1: Cladogram based on 16S rRNA of *Helicobacter* species used

	H. pylori J99	*H. pylori* 26695	*H. pylori* HPAG1	*H. pylori* Shi470	*H. mustelae*	*H. canadensis*	*H. canis*	*H. felis*	*H. pullorum*
H. pylori J99	100								
H. pylori 26695	98	100							
H. pylori HPAG1	99	98	100						
H. pylori Shi470	99	99	99	100					
H. mustelae	93	92	93	93	100				
H. canadensis	94	93	93	93	94	100			
H. canis	76	75	76	75	80	83	100		
H. felis	93	92	92	92	90	89	78	100	
H. pullorum	95	94	95	95	94	97	79	92	100

Table 2: 16S rRNA similarity among *Helicobacter* spp. used in the present work.

alignment using ClustalW (Chenna et al., 2003) 16S rRNA sequences (from *Helicobacter* spp.) available on public databases clearly isolated *H. pylori* species (Figure 1). As expected the *H. pylori* sequenced strains (26695, J99, HPAG1 and Shi470) presented a similarity level =98%, and clustered together (>97% defines a species). Moreover, *Helicobacter* gastric and enterohepatic species appear to be in different clusters, according to previous work (Dewhirst et al., 2005). The similarity levels among *Helicobacter* spp. based on 16S rRNA is presented in Table 2.

MCRM clustering analysis

After genomic DNA hydrolysis with the selected REases the codified data were analysed using MCRM algorithm (Vale et al., 2008). The Simpson index of diversity, which reflects the capacity of the method to distinguish unrelated strains, was 100% (Hunter and Gaston, 1988). The produced dendrogram is present in Figure 2. Surprisingly, this analysis based on the genomic methylation status clearly isolated *H. pylori* species from non-pylori *Helicobacter* spp., as it was observed when the analysis focuses on the 16S rRNA. Out of 10 produced dendrograms, 60% cluster *H. pylori* together and 40% of the dendrograms also discriminate between *H. pylori* and non-pylori *Helicobacter* spp. All of these last mentioned dendrograms gathered *H. pylori* strains (data not shown). *H. pylori* was discriminated from non-pylori *Helicobacter* spp. at k/nM=0.04 (where, k=1 and nM=27,

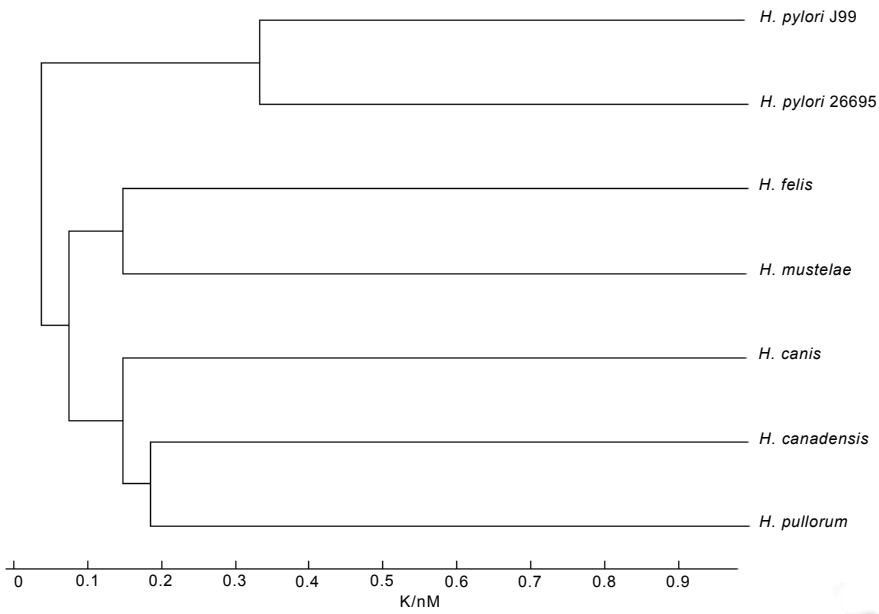

Figure 2: MCRM algorithm clustering of *Helicobacter* species used. Similarity represented by K/nM, where K is the number of R-M systems common to all tested strains and nM the total number of R-M systems screened.

i.e. one MTase common to all *Helicobacter* species used). This MTase common to all tested *Helicobacter* spp. was M.NaeI (Table 3).

Discussion

When comparing the number of MTases expressed in each *Helicobacter* spp. it was verified that *H. pylori* expresses a number of MTases higher than tested non-pylori *Helicobacter* spp. ($p<0,001$) and, that non-pylori *Helicobacter* spp. expresses a number higher than all sequenced bacteria analysed by REBASE ($p=0,005$). This analysis reveals that the increased number of MTases genes expressed is probably a characteristic of *Helicobacter* genus and not only of *H. pylori*. To our knowledge the evaluation of the expressed MTases in non-*pylori Helicobacter* spp. has only been referred for the sequenced genomes of *H. acinonychis* Sheeba (Eppinger et al., 2006) and *H. hepaticus* ATCC 51449 (Suerbaum et al., 2003) with 29 and 8 M genes, according to REBASE (Roberts et al., 2007), respectively. Besides this analysis only the sequence GATC has been screened for methylation in *H. mustelae* (Edmonds et al., 1992). Present study and Edmonds et al. (1992) study (Edmonds et al., 1992) found that the GATC methylation is absent in *H. mustelae* (Table3).

The only MTase that it was found to be expressed among all tested *Helicobacter* spp. is M.NaeI (Table 3). Previously we have reported that this MTase is probably conserved in all *H. pylori* strains (Vale and Vitor, 2007; Vale et al., 2008). In order to confirm if this MTase is common to the *Helicobacter* genus, other non-pylori *Helicobacter* spp. should be included in the study, and for each species several stains should be characterized.

Analysis of 16S rRNA gene sequences has become the primary method for determining prokaryotic phylogeny, which is the current basis for prokaryotic systematics. Although it has been described that *Helicobacter* is susceptible to horizontal transfer of 16S rRNA gene so that it can be misleading in *Helicobacter* spp. identification (Vandamme et al., 2000; Dewhirst et al., 2005), the 16S rRNA cladogram clearly discriminate the *Helicobacter* spp. used in the present study as expected. Moreover, Figure 1 presents gastric *Helicobacter* spp. and enterohepatic *Helicobacter* spp. in different clusters, as described elsewhere (Dewhirst et al., 2005). However, the cladogram obtained from 60 kDa heat-shock protein (HSP60), referred as better marker for *Helicobacter* species phylogeny (Mikkonen et al., 2004) was similar to the one obtained with 16S rRNA (data not shown).

REase	Organism						
	H. pylori J99	*H. pylori* 26695	*H. mustelae*	*H. canadensis*	*H. canis*	*H. felis*	*H. pullorum*
DpnI	1	1	0	0	0	1	0
DpnII	1	1	0	0	0	**1**	0
HhaI	1	1	0	0	0	0	0
AseI	1	1	0	0	1	0	0
FokI	0	0	0	0	0	0	0
DdeI	0	0	1	0	0	0	0
HaeIII	0	0	0	1	1	0	1
MspI	1	0	0	0	0	1	0
Hpy188I	0	0	0	1	1	0	0
Hpy99I	1	0	1	1	1	1	0
HpyCH4III	0	0	0	0	1	0	0
HpyCH4IV	1	0	0	1	1	0	1
HpyCH4V	0	0	0	0	0	0	0
DraI	0	0	0	0	0	0	0
ScrFI	0	0	0	0	1	0	0
Fnu4HI	0	0	0	0	0	0	0
EagI	0	0	1	1	1	1	1
NaeI	1	1	1	1	1	1	1
BssHII	1	1	0	1	0	0	1
BstUI	0	0	1	0	0	0	0
FauI	0	0	1	0	1	1	0
TaqI	1	1	0	0	0	0	0
NlaIII	1	1	0	0	1	1	0
AciI	0	0	0	0	0	0	0
BseRI	1	1	0	1	1	0	0
Sau96I	0	0	0	0	0	0	1
Hpy188III	0	0	0	0	0	0	0
Total	12	9	6	8	12	8	6

"0" – DNA digested (absence of an active MTase);

"1" – DNA not digested (suggestion of an active MTase).

Table 3: Type II MTases expressed in tested *Helicobacter* spp.

A surprising result was the capacity of the genomic methylation and of MCRM algorithm to cluster separately gastric *Helicobacter* spp. and enterohepatic *Helicobacter* spp. in 40% of the produced dendrograms. Also this methodology discriminate *H. pylori* from non-pylori *Helicobacter* spp. in 60% of the produced dendrograms. When genomic methylation and MCRM analysis is compared with the current phylogeny approach the results are remarkably similar. The genomic methylation status appears to be an interesting new tool to characterize genomes with a high number of MTases expressed. It has been described that R-M systems are subject to horizontal transfer (Jeltsch and Pingoud, 1996; Gressmann et al., 2005). The separation between gastric *Helicobacter* spp. and

MTases predicted *H. pylori* Shi470	*Helicobacter* sequenced strains*			
	H. pylori 26695	*H. pylori* J99	*H. pylori* HPAG1	*H. acinonychis* str. Sheeba
>M2.HpySORF245P CCTC 1227 nt	1018/1067 (95%)	185/197 (93%)	1013/1068 (94%)	967/1085 (89%)
>M.HpySORF265P CCTTC 201 nt	59/170 (93%)	120/135 (88%)	169/180 (93%)	
>M.HpySORF465P GATC 834 nt	786/825 (95%)	756/829 (91%)	745/793 (93%)	731/818 (89%)
>M.HpySORF1370P CCGG 759 nt	726/759 (95%)	712/762 (93%)	724/760 (95%)	711/762 (93%)
>M.HpySORF1620P CTNAG 1110 nt			758/782 (96%) 221/226 (97%)	724/786 (92%)
>M.HpySORF1840P GCNGC 930 nt				345/369 (93%)
>M.HpySORF2355P ATTAAT 2451 nt	2318/2456 (94%)	2302/2459 (93%)	2337/2455 (95%)	2302/2459 (93%)
>HpySORF2365P ACNGT 624 nt	608/622 (97%)	585/625 (93%)	604/626 (96%)	
>M.HpySORF2375P ACGT 1047 nt	979/1025 (95%)	967/1023 (94%)	499/526 (94%)	
>M.HpySORF2550P GGCC 984 nt				910/987 (92%)
>M.HpySDamP GATC 540 nt	513/540 (95%)	509/541 (94%)	519/541 (95%)	496/541 (91%)
>M.HpySORF3420P CCNNGG 1215 nt		439/457 (96%) 369/379 (97%) 251/259 (96%)	453/457 (99%) 382/387 (98%) 254/258 (98%)	
>HpySORF3515P 4827 nt	2429/2755 (88%) 1886/2087 (90%)	4247/4939 (85%)	2400/2760 (86%) 1854/2062 (89%)	
>M.HpySORF4790P GTNNAC 1149 nt	1093/1142 (95%)	1099/1151 (95%)	1093/1142 (95%)	1053/1146 (91%)
>M.HpySORF5570P TCNNGA 963 nt	657/673 (97%) 164/171 (95%)	924/964 (95%)	937/963 (97%)	909/963 (94%),
>M.HpySORF5770P GCGC 963 nt	904/967 (93%)	887/952 (93%)	907/964 (94%)	871/956 (91%)
>M.HpySORF6255P CATG 990 nt	926/993 (93%)	910/993 (91%)	928/991 (93%)	882/987 (89%)
>HpySORF6990BP 3510 nt	1484/1589 (93%) 864/940 (91%) 512/543 (94%) 108/113 (95%) 46/47 (97%)	1442/1594 (90%) 835/945 (88%) 740/816 (90%) 104/113 (92%)	1435/1554 (92%) 835/937 (89%) 501/547 (91%) 105/113 (92%) 45/47 (95%)	1434/1587 (90%) 831/956 (86%) 709/805 (88%)
>M.HpySORF7000P GANTC 1080 nt	1040/1081 (96%)	1035/1081 (95%)	1035/1081 (95%)	
>HpySORF7070P GAAGA 1059 nt	635/653 (97%) 349/359 (97%)	539/565 (95%) 340/361 (94%)		608/655 (92%) 333/362 (91%)
>M.HpySORF7080P GATATC 894 nt	45/46 (97%)			
>HpySORF7540P 2676 nt	158/169 (93%) 176/208 (84%)	670/692 (96%)	894/927 (96%) 776/809 (95%) 228/238 (95%) 173/183 (94%) 57/60 (95%)	1250/1348 (92%) 869/930 (93%) 212/241 (87%)
>HpySORF7805P 3840 nt	914/1006 (90%) 683/769 (88%) 216/223 (96%) 193/203 (95%) 164/191 (85%) 67/69 (97%)	689/732 (94%) 690/757 (91%) 678/761 (89%) 193/210 (91%) 69/72 (95%) 68/75 (90%)	704/766 (91%) 654/729 (89%) 682/760 (89%) 183/199 (91%) 65/69 (94%)	567/648 (87%) 561/652 (86%) 164/194 (84%) 62/66 (93%)
>M.HpySORF7990P TGCA 897 nt	142/143 (99%) 81/86 (94%)		858/897 (95%)	818/901 (90%)

"* For *H. hepaticus* it was not found any similarity."

Table 4: Blast results of DNA sequence of MTases from H. pylori strain Shi470 predicted by REBASE against all other Helicobacter spp. sequenced strains.

enterohepatic *Helicobacter* spp. suggests that probably theses species, which have a high level of organ specificity (Kusters et al., 2006) may have access to different sets of R-M systems through horizontal gene transfer. The horizontal gene transfer might occur only in ideal conditions provided by the specific tissue environment characteristic of each species. This could justify the presence of gastric *Helicobacter* spp. and enterohepatic *Helicobacter* spp. in different clusters. Similarly, *H. pylori* unique reservoir may justify the presence of the two tested strains in a different cluster. Finally, the results suggest that possibly some R-M systems are not as mobile as previously described, or are not available for horizontal transfer due to the isolation provided by the human (or animal) reservoir of each species, because the genomic methylation analysis permits to discriminate among *Helicobacter* spp. The R-M systems do not appear to be spread in a miscellaneous manner.

Indeed, a blast analysis (Zhang et al., 2000) of all methyltransferases predicted by REBASE (Roberts et al., 2007) for the recently sequenced *H. pylori* strain Shi470 clearly shows that most methyltransferases are identical to other *H. pylori* sequenced strains and to *H. acinonychis* strain Sheeba (Table 4, supplementary material). It is clear from analysis of table 4 that most of the methyltransferases of the recent sequenced *H. pylori* strain Shi470 are identical to the other *H. pylori* sequenced strains and *H. acinonychis* strain Sheeba, but none is observed in *H. hepaticus* sequenced strain. We postulate that similarly to *H. pylori*, non-pylori *Helicobacter* may also present a high diversity of MTases expressed which could be used for strain typing, but this still needs to be confirmed with further investigation. R-M systems probably play an important role in *Helicobacter* genus biology that has not been ascertained, yet.

In conclusion, it was observed a high number of MTases expressed in non-pylori *Helicobacter* spp. as it was previously determined for *H. pylori*. The discrimination of *Helicobacter* species by the dendrogram produced with MCRM algorithm and by 16S rRNA alignment performed in a similar way. The results suggest that some R-M systems do not appear to be spread in a miscellaneous manner, once genomic methylation analysis allows discrimination among *Helicobacter* spp. Future work should include increasing the number of *Helicobacter* species analysed and also the number of tested strains from each species, in order to confirm present study results.

Acknowledgments

We thank Lurdes Monteiro, Francis Mégraud, Jay Solnick and, Nuno Azevedo for the *Helicobacter* spp. strains. This work was partially supported by New England Biolabs, Inc. (USA).

References

1. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, et al. (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 31: 3497-3500. » CrossRef » Pubmed » Google Scholar
2. Covacci A, Telford JL, Del Giudice G, Parsonnet J, Rappuoli R (1999) *Helicobacter pylori* virulence and genetic geography. Science 284: 1328-1333. » CrossRef » Pubmed » Google Scholar
3. Dewhirst FE, Shen Z, Scimeca MS, Stokes LN, Boumenna T, et al. (2005) Discordant 16S and 23S rRNA gene phylogenies for the genus *Helicobacter*: implications for phylogenetic inference and systematics. J Bacteriol 187: 6106-6118. » CrossRef » Pubmed » Google Scholar
4. Dunn BE, Cohen H, Blaser MJ (1997) Helicobacter pylori. Clin Microbiol Rev 10: 720-741. » CrossRef » Pubmed » Google Scholar
5. Edmonds P, Hall BM, Edwards WR, Hartline KM (1992) Presence of methylated adenine in GATC sequences in chromosomal DNAs from Campylobacter species. J Bacteriol 174: 8156-8157. » CrossRef » Pubmed » Google Scholar
6. Eppinger M, Baar C, Linz B, Raddatz G, Lanz C, et al. (2006) Who ate whom? Adaptive *Helicobacter* genomic changes that accompanied a host jump from early humans to large felines. PLoS Genet 2: e120. » CrossRef » Pubmed » Google Scholar
7. Gressmann H, Linz B, Ghai R, Pleissner KP, Schlapbach R, et al. (2005) Gain and loss of multiple genes during the evolution of Helicobacter pylori. PLoS Genet 1: e43. » CrossRef » Pubmed » Google Scholar
8. Hunter PR, Gaston MA (1988) Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol 26: 2465-2466. » CrossRef » Pubmed » Google Scholar
9. Jeltsch A, Pingoud A (1996) Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. J Mol Evol 42: 91-96. » CrossRef » Pubmed » Google Scholar
10. Kusters JG, van Vliet AH, Kuipers EJ (2006) Pathogenesis of *Helicobacter pylori* infection. Clin Microbiol Rev 19: 449-490. » CrossRef » Pubmed » Google Scholar

11. Lin LF, Posfai J, Roberts RJ, Kong H (2001) Comparative genomics of the restriction-modification systems in *Helicobacter pylori*. Proc Natl Acad Sci USA 98: 2740-2745. » CrossRef » Pubmed » Google Scholar

12. Linz B, Balloux F, Moodley Y, Manica A, Liu H, et al. (2007) An African origin for the intimate association between humans and *Helicobacter pylori*. Nature 445: 915-918. » CrossRef » Pubmed » Google Scholar

13. Megraud F (1996) Diagnostic bactériologique standart de l'infection à *Helicobacter pylori*. In Megraud,F. and Lamouliatte,H. (eds), *Helicobacter pylori*. Elsevier Amsterdam pp249-266.

14. Mikkonen TP, Karenlampi RI, Hanninen ML (2004) Phylogenetic analysis of gastric and enterohepatic *Helicobacter* species based on partial HSP60 gene sequences. Int J Syst Evol Microbiol 54: 753-758. » CrossRef » Pubmed » Google Scholar

15. Naito T, Kusano K, Kobayashi I (1995) Selfish behavior of restriction-modification systems. Science 267: 897-899. » CrossRef » Pubmed » Google Scholar

16. Nobusato A, Uchiyama I, Kobayashi I (2000) Diversity of restriction-modification gene homologues in *Helicobacter pylori*. Gene 259: 89-98. » CrossRef » Pubmed » Google Scholar

17. Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, et al. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res 31: 1805-1812. » CrossRef » Pubmed » Google Scholar

18. Roberts RJ, Vincze T, Posfai J, Macelis D (2007) REBASE-enzymes and genes for DNA restriction and modification. Nucleic Acids Res 35: D269-D270. » CrossRef » Pubmed » Google Scholar

19. Suerbaum S, Josenhans C, Sterzenbach T, Drescher B, Brandt P, et al. (2003) The complete genome sequence of the carcinogenic bacterium *Helicobacter hepaticus*. Proc Natl Acad Sci USA 100: 7901-7906. » CrossRef » Pubmed » Google Scholar

20. Takata T, Aras R, Tavakoli D, Ando T, Olivares AZ, et al. (2002) Phenotypic and genotypic variation in methylases involved in type II restriction-modification systems in *Helicobacter pylori*. Nucleic Acids Res 30: 2444-2452. » CrossRef » Pubmed » Google Scholar

21. Vale FF, Encarnacao P, Vitor JM (2008) A new algorithm for cluster analysis of genomic methylation: the *Helicobacter pylori* case. Bioinformatics 24: 383-388. » CrossRef » Pubmed » Google Scholar

22. Vale FF, Vitor JM (2007) Genomic Methylation: a Tool for Typing Helicobacter pylori Isolates. Appl Environ Microbiol 73: 4243-4249. » CrossRef » Pubmed » Google Scholar

23. Vandamme P, Harrington CS, Jalava K, On SL (2000) Misidentifying helicobacters: the *Helicobacter cinaedi* example. J Clin Microbiol 38: 2261-2266. » CrossRef » Pubmed » Google Scholar

24. Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. J Comput Biol 7: 203-214. » CrossRef » Pubmed » Google Scholar