



## Phylogenetic analysis, based on EPIYA repeats in the *cagA* gene of Indian *Helicobacter pylori*, and the implications of sequence variation in tyrosine phosphorylation motifs on determining the clinical outcome

Santosh K. Tiwari<sup>1\*</sup>, Vishwas Sharma<sup>1\*</sup>, Varun Kumar Sharma<sup>2\*</sup>, Manoj Gopi<sup>1</sup>, Saikant R<sup>1</sup>, Amrita Nandan<sup>1</sup>, Avinash Bardia<sup>1</sup>, Sivaram Gunisetty<sup>1</sup>, Prasanth Katikala<sup>3</sup>, Md. Aejaz Habeeb<sup>1</sup>, Aleem A. Khan<sup>1</sup> and C.M. Habibullah<sup>1</sup>

<sup>1</sup>Centre for Liver Research and Diagnostics, Deccan College of Medical Sciences, Kanchanbagh, Hyderabad, Andhra Pradesh, India.

<sup>2</sup>Department of Biotechnology, C.S.J.M. University, Kanpur, Uttar Pradesh, India.

<sup>3</sup>Pharmaceutical Biotechnology Division, A.U. College of Pharmaceutical Sciences, Andhra University, Visakhapatnam, Andhra Pradesh, India.

### Abstract

The population of India harbors one of the world's most highly diverse gene pools, owing to the influx of successive waves of immigrants over regular periods in time. Several phylogenetic studies involving mitochondrial DNA and Y chromosomal variation have demonstrated Europeans to have been the first settlers in India. Nevertheless, certain controversy exists, due to the support given to the thesis that colonization was by the Austro-Asiatic group, prior to the Europeans. Thus, the aim was to investigate pre-historic colonization of India by anatomically modern humans, using conserved stretches of five amino acid (EPIYA) sequences in the *cagA* gene of *Helicobacter pylori*. Simultaneously, the existence of a pathogenic relationship of tyrosine phosphorylation motifs (TPMs), in 32 *H. pylori* strains isolated from subjects with several forms of gastric diseases, was also explored. High resolution sequence analysis of the above described genes was performed. The nucleotide sequences obtained were translated into amino acids using MEGA (version 4.0) software for EPIYA. An MJ-Network was constructed for obtaining TPM haplotypes by using NETWORK (version 4.5) software. The findings of the study suggest that Indian *H. pylori* strains share a common ancestry with Europeans. No specific association of haplotypes with the outcome of disease was revealed through additional network analysis of TPMs.

**Key words:** *Helicobacter pylori*, EPIYA motifs, tyrosine phosphorylation motifs, haplotypes, anatomically modern humans.

Received: May 10, 2010; Accepted: October 14, 2010.

### Introduction

The dawn of the 20<sup>th</sup> century witnessed the discovery of one of the most controversial microorganisms, *Helicobacter pylori* (*H. pylori*), responsible for provoking a rupture in contemporary popular medical doctrines, thereby enforcing changes in global conceptions of gastro-duodenal disorders (Marshall, 2001). In spite of a general consensus regarding the existence of a causal relationship, there is still disagreement as to how a single bacterium could possibly cause such a variety of disease conditions (Montecucco and Rappuoli, 2001). The apparent paradox suggests that the mere presence of *H. pylori* in the stomach

is insufficient to cause gastric disease, this requiring one or more additional conditions (Crowe, 2005). Apart from several others, *H. pylori* must bear an arsenal of specific virulence-genes such as the *cag*-pathogenicity island (*cag*-PAI), the vacuolating associated cytotoxin gene A (*vacA*), the outer membrane protein A (*oipA*), and blood group antigen binding adhesin (*babA*), to be potentially toxigenic (Censini *et al.*, 1996, Cover *et al.*, 1994, Zamboni *et al.*, 2003).

Infection by *cag*-PAI bearing *H. pylori* is recognized as increasing the risk of overt gastric disorders, such as peptic ulceration, gastric cancer and mucosa associated lymphoid tissue (MALT)-lymphoma. Seven genes of this pathogenicity island (*hp0524*, *hp0525*, *hp0527*, *hp0528*, *hp0530*, *hp0532* and *hp0544*) form a typical needle-syringe assembly called the type IV secretion apparatus (T4SS) which translocates 120-145 kDa CagA proteins, either directly into host cells or into the bacterial environment

Send correspondence to Aleem A. Khan. Centre for Liver Research and Diagnostics, Deccan College of Medical Sciences, Kanchanbagh, Hyderabad, 500058 Andhra Pradesh, India. E-mail: aleem\_a\_khan@rediffmail.com.

\*These authors contributed equally to this work.

(Bourzac and Guillemin, 2005; Christie and Cascales, 2005). During translocation, these undergo tyrosine phosphorylation by several members of the Src family kinases (SFK) such as c-Src, Fyn, Lyn and Yes (Stein *et al.*, 2002). This phosphorylation is reported to occur at specific sites in the CagA protein known as tyrosine phosphorylation motifs (TPMs), characterized by the presence-of a stretch of conserved nucleotide sequences (CNS). Three predicted motifs (TPM-A, TPM-B and TPM-C) have already been reported, based on these CNSs (Odenbreit *et al.*, 2000; Owen *et al.*, 2003). In addition, phosphorylation is also known to occur at the Glu-Pro-Ile-Tyr-Ala (EPIYA) amino acid sequence in TPMs.

The Indian population is widely known for its unique genetic and cultural diversity. Numerous phylogenetic studies based on mitochondrial DNA (mt DNA) and Y-chromosomal variation have proved that India played crucial role in the first major colonization by anatomically modern humans (AMH), at least half-a-million years ago (Chaubey *et al.*, 2008). Although, to a certain extent these studies have brought enlightenment to the evolutionary history of AMHs, the number of waves and the periods of migration still remain uncertain and subject to dogmatic views.

Several investigators have used *H. pylori* as a biological model when studying waves of immigration, owing to co-evolution with its human host (Devi *et al.*, 2007). Coincidentally, the phylogenetic analysis of *H. pylori* housekeeping gene sequences mirrors the migratory path of AMHs. The number and pattern of the five conserved amino acid sequences (EPIYA) present in the repeat-region of *cagA* have been well-explored in several phylogenetic studies, with the aim of dissecting the genetic origin of *H. pylori* strains. These have been broadly divided into Western CagA (WSS) and East-Asian CagA (ESS) specific sequences. WSS usually possess an A-B-C pattern characterized by the presence of flanking conserved amino acids, although several other subtypes of this pattern (ABC, ABCC, & ABCCC etc) are already known (Higashi *et al.*, 2002). On the contrary, ESS contains a JSR region, previously defined by Yamaoka *et al.*, (1999), besides possessing an EPIYA motif, designated “EPIYA-D, thereby justifying their classification as A-B-D (Azuma *et al.*, 2004). Furthermore, only few studies (Owen *et al.*, 2003) have focused on the status of CagA phosphorylation motifs (TPMs), with no data available from the Indian peninsula. Therefore, the present study attempted to address the prehistoric colonization of humans in India by using the *H. pylori* genome and, additionally, assess the association of haplotypes of tyrosine phosphorylation motifs with gastroduodenal diseases.

## Materials and Methods

A total of 32 indigenous *H. pylori* strains isolated from individuals with various gastrointestinal disorders, and undergoing treatment at the Department of Gastro-

enterology, Deccan College of Medical Sciences, were included for detailed analysis. Information regarding the clinical status and ethnic origin of the study-subjects are given in Table 1. Due, relevant ethical approval for undertaking the study, as well as written informed-consent from the participants prior to inclusion was obtained.

PCR based analysis was applied to the target genes, namely EPIYA motifs and tyrosine phosphorylation motifs-A, B, C, as reported previously, using designated oligonucleotide primers (Karita *et al.*, 2003, Owen *et al.*, 2003). Amplification was performed twice and the sequencing of the amplified products was done with both forward and re-

**Table 1** - Details of the *H. pylori* strains used in the study.

Strains ID	Disease type	Ethnic origin of patients	Type of EPIYA	TPM A	TPM B	TPM C
MS 2	DU	Asian	Western	Present	Present	Absent
MS 5	DU	Asian	Western	Present	Present	Absent
MS 8	DU	Asian	Western	Present	Present	Absent
MS 10	PUD	Asian	Western	Present	Present	Absent
MS 11	DU	Asian	Western	Present	Present	Absent
MS 13	GC	Asian	Western	Present	Present	Absent
MS 14	DU	Asian	Western	Present	Present	Absent
MS 15	GU	Asian	Western	Present	Present	Absent
MS 16	DU	Asian	Western	Present	Present	Absent
MS 17	DU	Asian	Western	Present	Present	Absent
MS 18	DU	Asian	Western	Present	Present	Absent
MS 20	DU	Asian	Western	Present	Present	Absent
MS 23	DU	Asian	Western	Present	Present	Absent
MS 26	GC	Asian	Western	Present	Present	Absent
MS 28	DU	Asian	Western	Present	Present	Absent
MS 30	NUD	Asian	Western	Present	Present	Absent
MS 40	GU	Asian	Western	Present	Present	Absent
MS 56	DU	Asian	Western	Present	Present	Absent
MS 233	NUD	Asian	Western	Present	Present	Present
MS 401	NUD	Asian	Western	Present	Present	Absent
GC 1	GC	Asian	Western	Present	Present	Absent
GC 2	GC	Asian	Western	Present	Present	Absent
GC 3	GC	Asian	Western	Present	Present	Absent
GC 6	GC	Asian	Western	Present	Present	Absent
GC 8	GC	Asian	Western	Present	Present	Absent
GC 11	GC	Asian	Western	Present	Present	Absent
GC 12	GC	Asian	Western	Present	Present	Absent
GC 16	GC	Asian	Western	Present	Present	Absent
GC 33	GC	Asian	Western	Present	Present	Absent
GC 83	GC	Asian	Western	Present	Present	Absent
GC 123	GC	Asian	Western	Present	Present	Absent
KL 11	GC	Asian	Western	Present	Present	Absent

DU- duodenal ulcer, PUD- peptic ulcer disease, GU- gastric ulcer, NUD- non-ulcer dyspepsia, GC-gastric cancer.

verse primers using 5.2.0 Version software from Applied Biosystems (Applied Biosystems, Foster City, USA). For analysis, Quality Values (QVs) were sought from the software itself. The sequences were edited and assembled using AutoAssembler (version 1.4) software (Applied Biosystems) to obtain consensus sequences. Furthermore, in order to minimize ambiguities in the sequences, the set-up was assembled with a minimum overlap of 20 bases and 20% percent error. The sequences were then translated into amino acid sequences by using MEGA (version 4.0) software (Tamura *et al*, 2007). Bootstrap phylogeny tests with 500 replications and 1234 seeds were used for this purpose. Finally, they were assigned to the Western or the East Asian specific groups based on the presence of C- or D- repeats in the EPIYA motifs, respectively (Figure 1 and Figure S1). Translated amino acid sequences of individual strains were comparatively analyzed using neutral sequences of the J99 *H. pylori* reference strain. The selection of model parameters was done using Median joining network, based on the Maximum Parsimony method from a set of splits, optimal realizations and reticulograms from a distance matrix. NETWORK (version 4.5) software was used for median-joining network construction (see Figures S2 a-c) (Saitou and Nei, 1987).

The sequences pertaining to three tyrosine phosphorylation motifs were aligned using AE000511 for TPM A and TPM B, as well as an AF202973 reference sequence for TPM C, the differences being noted and displayed in phylogenetic networks (Figures 2, 3 and 4). The *cagA* nucleotide sequences containing the tyrosine phosphorylation motifs of the 32 Indian isolates were deposited in GenBank (accession numbers FJ599712-FJ599743). A codon selection test was performed using an online non-synonymous to synonymous substitution ratio calculator (HIV Databases).

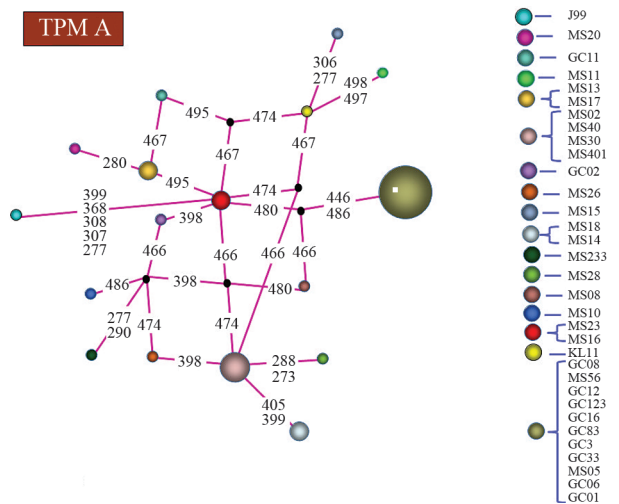
**Results**

We have analyzed a total 32 strains of *H. pylori* isolated from various clinical backgrounds (Table 1). Se-

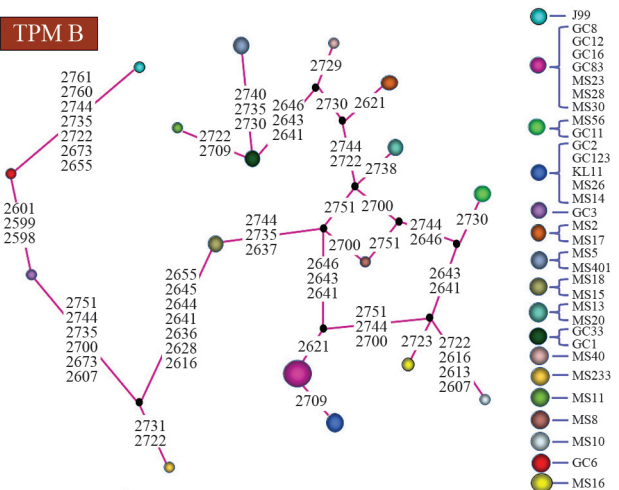
	A	B	C
GC-8	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
MS-56	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
GC-2	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
GC-12	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
GC-123	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
GC-16	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
GC-3	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
MS-2	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
MS-5	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
MS-23	EPVYA	EPVYA	HDKVDLISKVGR-EPVYA
MS-18	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
MS-28	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
MS-13	EPVYA	EPVYA	HDKVDLISKVGR-EPVYA
GC-33	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
MS-40	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
MS-233	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
MS-15	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
MS-17	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
GC-11	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
KL-11	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
MS-10	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
GC-6	EPVYA	EPVYA	HDKVDLISKVGR-EPVYA
MS-16	EPVYA	EPVYA	HDQVDLISKVGL-EPVYA
MS-30	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
GC-1	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
MS-401	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
MS-26	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
MS-14	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
GC-83	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
MS-11	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
MS-8	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA
MS-20	EPVYA	EPVYA	HDKVDLISKVGL-EPVYA

**Figure 1** - Deduced amino acid sequences of the 3repeat region of the *cagA* gene of 32 Indian strains of *H. pylori*, showing the A-B-C pattern of EPIYA repeats.

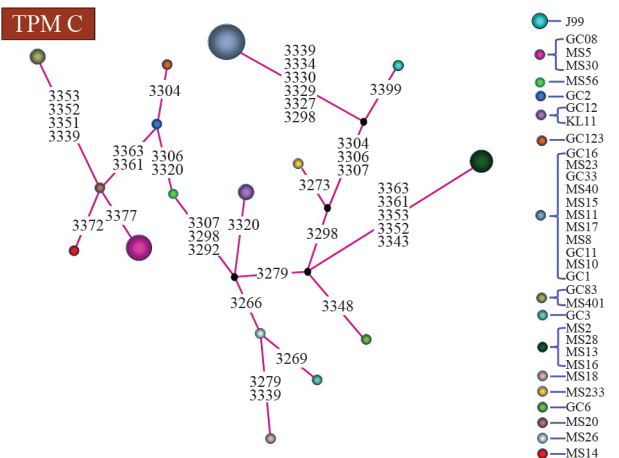
quence analysis of the repeat region in *cagA* for assessing EPIYA sequences, revealed the A-B-C pattern to be common in Europeans. Neither the A-B-D pattern (East Asian),



**Figure 2** - MJ-network analysis of TPM A showing 16 haplotypes.



**Figure 3** - MJ-network analysis of TPM B showing 16 haplotypes.



**Figure 4** - MJ-network analysis of TPM C showing 15 haplotypes.



nor any other A-B-C subtypes were present in any of the 32 isolates (Figure 1). The selection test gave a value of 1.4345, thereby signifying positive selection and co-evolution.

Amplification followed by deduced amino acid sequence analysis at each of the three phosphorylation sites, TPM-A motif, characterized as KFGDQRY, at site 122 in all the strains analyzed (100%) (Figure S2a). Similarly, the TPM-B motif, originally defined by the amino acid sequence KNS(T/g)EPIY, was found as KNEPIY at site 899 in all the sequences (Figure S2b). Nevertheless, TPM-C, as characterized by KDKDSTKY, was found at site 1029 in only 3.1% of all the Indian strains screened (Figure S2c).

High resolution analysis of the TPM-A motif revealed 16 distinct haplotypes (Figure 2), those with mutations at site 446-486 (in the strains GC-8, MS-56, GC-12, GC-123, GC-16, GC-83, GC-3, MS-5, GC-33, GC-6, GC-1) being predominant. Analysis of the tyrosine phosphorylation motif-B (TPM-B) indicated 16 different haplotypes (Figure 3). Similar high resolution analysis of the tyrosine phosphorylation motif C (TPM-C) showed 15 haplotypes (Figure 4).

Distribution of all the *cagA* tyrosine phosphorylation motifs was clinically irrelevant, as TPM A and TPM B were found to be present in 100% of the strains (Figures 2 and 3), whereas TPM C was observed in only 1 (3.1%) (Figure 4). Similarly different haplotypes of TPM A, TPM B & TPM C also showed no disease specific association.

## Discussion

The Indian microbial genome is a melting-pot for both evolutionary and pathogenic studies, seeing that it accounts for one of largest gene pools, with more than 1 billion denizens. The A-B-C pattern of EPIYA sequences in Indian strains of *H. pylori* represents a common ancestral root of origin with Europeans, as reported previously (Devi *et al.*, 2007). Although both the time and number of migrations cannot be estimated with software, exploration through various approaches, such as anthropological (Mishra, 2001), historical (Kennedy *et al.*, 1987), mitochondrial DNA and Y-chromosomal studies, have demonstrated that the Austro-Asiatic (AA) tribal groups were the first settlers of India (Majumder, 2001; Metspalu *et al.*, 2004; Thangaraj *et al.*, 2005; Kumar *et al.*, 2007; Chaubey *et al.*, 2008). Furthermore, this microbe is also known to have co-evolved with anatomically modern humans (AMH) in Africa (~50-70 kYa) (Linz *et al.*, 2007). On considering the above hypothesis, one can say that AA language speakers would have been the first to introduce *H. pylori* into India, and not Indo-European, as previously reported (Devi *et al.*, 2007). In fact Indo-Europeans played a major role in the later-day colonization and expansion of this bacterium during the Neolithic era of the Stone age (Mishra, 2001).

The AA family is broadly divided into two subfamilies, *i.e.* the Mundari and the Mon-Khmer (Ethnologue

web-site). The former, found exclusively on the Indian sub-continent, are considered to have been traditional hunters, their feeding on uncooked food having been the most probable acquisition-route of *H. pylori* in humans (~50-70 kYa) (Mishra, 2001).

Although tyrosine phosphorylation reportedly occurs at any of the three motifs (TPM-A, TPM-B and TPM-C), detailed sequence-analysis of individual distribution proved to be of no prognostic value, as no site-specific mutation in any of the three tyrosine phosphorylation motifs was observed to be directly associated with disease status (Figure S2 a, b, c), thereby implying the outcome to be TPM-independent. The presence of KNEPIY in the place of KNS(T/g)EPIY at site 899 in the strains studied, requires further investigation on large number of samples from different geographical locations of the Indian subcontinent (Figure 3). These observations are in absolute conformity with those reported by Owen *et al.*, (2003), who demonstrated there to be no association of the number and type of TPMs present, with the severity of the disease. Nonetheless, they reported relatively lower frequencies in all the three TPMs (Owen *et al.*, 2003) than was the case in the present study. Another study from Costa Rica (Occhialini *et al.*, 2001), reported relative frequencies of 100% and 58% for TPM A and B, respectively, but were unable to detect the TPM C motif in any of the strains studied. The reason for such discordance between studies is unclear thereby warranting detailed investigation of large clinical isolates from several geographical areas. Although, according to MJ-networks, TPM genes may not be pathogenetically relevant, an attempt was made to understand whether haplotypes play any specific, associated role in altering the outcome of a disease. Our high-resolution study based on MJ networks in 32 *H. pylori* strains showed this was not so.

In conclusion, the results of the present investigation support the dogma of European roots for Indian *H. pylori*. Nevertheless, their first introduction to the Indian subcontinent by Indo-Europeans remains a highly contentious issue, with sufficient reports favoring Austro-Asiatic speakers as having been the first settlers. Finally, sequence-analysis of the *cagA* tyrosine phosphorylation motifs revealed no association with their clinical presentation, as evident from frequency distribution and MJ-network analysis, thereby implying that the nature and severity of gastroduodenal diseases are independent of tyrosine phosphorylation motifs.

## Acknowledgments

The authors would like to acknowledge the kind help of Dr. Mahesh Dharne, National Centre for Cell Sciences, Pune, INDIA in sequencing clinical samples.

## References

- Azuma T, Yamazaki S, Yamakawa A, Ohtani M, Muramatsu A, Suto H, Ito Y, Dojo M, Yamazaki Y, Kuriyama M, *et al.*

- (2004) Association between diversity in the Src Homology 2 domain-containing tyrosine phosphatase binding site of *Helicobacter pylori* CagA protein and gastric atrophy and cancer. *J Infect Dis* 189:820-827.
- Bourzac KM and Guillemin K (2005) *Helicobacter pylori*-host cell interactions mediated by type IV secretion. *Cell Microbiol* 7:911-919.
- Censini S, Lange C, Xiang C, Crabtree P, Giara P, Borodovsky M, Rappuoli R and Covacci A (1996) *cag*, a pathogenicity island of *Helicobacter pylori* encodes type I-specific and disease-associated virulence factors. *Proc Natl Acad Sci USA* 93:14648-14653.
- Chaubey G, Karmin M, Metspalu E, Metspalu M, Selvi-Rani D, Singh VK, Parik J, Solnik A, Naidu BP, Kumar A, *et al.* (2008) Phylogeography of mtDNA haplogroup R7 in the Indian peninsula. *BMC Evol Biol* 8:e227.
- Christie PJ and Cascales E (2005) Structural and dynamic properties of bacterial type IV secretion systems. *Mol Membr Biol* 22:51-61.
- Cover TL, Tummuru MK, Cao P, Thompson A and Blaser MJ (1994) Divergence of genetic sequences for the vacuolating cytotoxin among *Helicobacter pylori* strains. *J Biol Chem* 269:10566-10573.
- Crowe SE (2005) *Helicobacter infection*, chronic inflammation, and the development of malignancy. *Curr Opin Gastroenterol* 21:32-38.
- Devi SM, Ahmed I, Francalacci P, Hussain MA, Akhter Y, Alvi A, Sechi LA, Mégraud F and Ahmed N (2007) Ancestral European roots of *Helicobacter pylori* in India. *BMC Genomics* 8:e184.
- Higashi H, Tsutsumi R, Fujita A, Yamazaki S, Asaka M, Azuma T and Hatakeyama M (2002) Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. *Proc Natl Acad Sci USA* 99:14428-14433.
- Karita M, Matsumoto S and Kamei T (2003) The size of *cagA* based on repeat sequence has the responsibility of the location of *Helicobacter pylori* in the gastric mucus and the degree of gastric mucosal inflammation. *Microbiol Immunol* 47:619-630.
- Kennedy KA, Deraniyagala SU, Roertgen WJ, Chiment J and Sisetell T (1987) Upper Pleistocene fossil hominids from Sri Lanka. *Am J Phys Anthropol* 72:441-461.
- Kumar V, Reddy AN, Babu JP, Rao TN, Langstieh BT, Thangaraj K, Reddy AG, Singh L and Reddy BM (2007) Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. *BMC Evol Biol* 7:e47.
- Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, Falush D, Stamer C, Prugnolle F, van der Merwe SW, *et al.* (2007) An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* 445:915-918.
- Majumder PP (2001) Ethnic population of India as seen from an evolutionary perspective. *J Biosci* 26:533-545.
- Marshall BJ (2001) One hundred years of discovery and rediscovery of *Helicobacter pylori* and its association with peptic ulcer disease. In: Mobley HLT, Mendz GL and Hazell SL (eds) *Helicobacter pylori*: Physiology and Genetics. American Society of Microbiology, Washington DC, pp 178-186.
- Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MTP, *et al.* (2004) Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genetics* 5:e26.
- Mishra VN (2001) Prehistoric human colonization of India. *J Biosci* 26:491-531.
- Montecucco C and Rappuoli R (2001) Living dangerously: How *Helicobacter pylori* survives in the human stomach. *Nat Rev Mol Cell Biol* 2:457-466.
- Occhialini A, Marais A, Urdaci M, Sierra R, Munoz N, Covacci A and Megraud F (2001) Composition and gene expression of the *cag* pathogenicity island in *Helicobacter pylori* strains from gastric carcinoma and gastritis patients in Costa Rica. *Infect Immun* 69:1902-1908.
- Odenbreit S, Püls J, Sedlmaier B, Gerland E, Fischer W and Hass R (2000) Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. *Science* 287:1497-1500.
- Owen RJ, Sharp SI, Chisholm SA and Rijpkema S (2003) Identification of *cagA* tyrosine phosphorylation DNA motifs in *Helicobacter pylori* isolates from peptic ulcer patients by novel PCR-restriction fragment length polymorphism and real-time fluorescence PCR assays. *J Clin Microbiol* 41:3112-3118.
- Saitou N and Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425.
- Stein M, Bagnoli F, Halenbeck R, Rappuoli R, Fantl WJ and Covacci A (2002) c-Src/Lyn kinases activate *Helicobacter pylori* CagA through tyrosine phosphorylation of the EPIYA motifs. *Mol Microbiol* 43:971-980.
- Tamura K, Dudley J, Nei M and Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software v. 4.0. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 24:1596-1599.
- Thangaraj K, Sridhar V, Kivisild T, Reddy AG, Chaubey G, Singh VK, Kaur S, Agarawal P, Rai A, Gupta J, *et al.* (2005) Different population histories of the Mundari-and Mon-Khmer-speaking Austro-Asiatic tribes inferred from the mtDNA 9-bp deletion/insertion polymorphism in Indian populations. *Hum Genet* 116:507-517.
- Yamaoka Y, El-Zimaity HM, Gutierrez O, Figura N, Kim JK, Kodama T, Kashima K and Graham DY (1999) Relationship between the *cagA* 3' repeat region of *Helicobacter pylori*, gastric histology, and susceptibility to low pH. *Gastroenterology* 117:342-349.
- Zambon CF, Navaglia F, Basso D, Rugge M and Plebani M (2003) *Helicobacter pylori* *babA2*, *cagA*, and *s1 vacA* genes work synergistically in causing intestinal metaplasia. *J Clin Pathol* 56:287-291.

## Internet Resources

- Ethnologue web-site, [http://www.ethnologue.com/show\\_family.asp?subid=271-16](http://www.ethnologue.com/show_family.asp?subid=271-16).
- NETWORK v. 4.5 software, <http://www.fluxusengineering.com> (February 7, 2010).

HIV Databases, non-synonymous to synonymous substitution ratio calculator, <http://www.hiv.lanl.gov>  
[www.fluxusengineering.com](http://www.fluxusengineering.com) (March 30, 2010).

### Supplementary Material

The following online material is available for this article

Figure S1 - Deduced amino acid sequences of 3' repeat region of *cagA* gene from 32 Indian strains of *H. pylori* showing EPIYA sequences.

Figure S2 - Deduced amino acid sequence of partial *cagA* gene of *H. pylori* showing (a) tyrosine phosphorylation motif A (TPM A), (b) tyrosine phosphorylation motif B (TPM B), and (c) tyrosine phosphorylation motif C (TPM C).

This material is available as part of the online article from <http://www.scielo.br/gmb>.

*Associate Editor: Luís Carlos de Souza Ferreira*

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Figure S1:** Deduced amino acid sequences of 3' repeat region of *cagA* gene from 32 Indian strains of *H. pylori* showing EPIYA sequences.

**GC-8**

\*IF\*\*IHGKNKDFSKAEETLKALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQKI  
TDKVDNLNQAVSVAKAMGDFSRVEQALADLKNFSKDQLAQQAQKNEDFNTGKKSEIYQSVKNGVNGTLVGNGLSQ  
AEATTLNFKNFSDIRKELNAKLGNFNNNNNNGLKN**EPIYA**KVNKKKAGQAASPE**EPIYA**QVAKKVNAKIDRLNQIASGL  
GGVGQAAGFPLKKHDKVDDLKSVGLSASP**EPIYA**TIDDLGGPFPLKRHDKVDDLKSVGL-

**MS-56**

\*IF\*\*IHGKNKDFSKAEETLKTGKSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQKIT  
DKVDNLSSAVSVAKATGDFSRVEQALADLKNFSKDQLAQQAQKNEDFNTGKNSELYQSVKNGVNGTLVGNGLSQAE  
ATTLNFKNFSDIKKELNEKFNFNNNNNNNGLKN**EPIYA**KVNKKKTGEVASLE**EPIYA**QVAKKVNAKIDRLNQIASGLGG  
VGQAAGFPLKRHDKVDDLKSVGLSASP**EPIYA**TIDDLGGPFPLKRHDKVDDLKSVGL-

**GC-2**

\*IF\*\*IHGKNKDFSKAEETLKALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQKI  
TDKVDNLNQAVSVAKAMGDFSRVEQALADLKNFSKDQLAQQAQKNEDFNTGKKSEIYQSVKNGVNGTLVGNGLSQ  
AEATTLNFKNFSDIRKELNAKLGNFNNNNNNGLKN**EPIYA**KVNKKKAGQTASPE**EPIYA**QVAKKVNAKIDRLNQIASGL  
GGVGQAAGFPLKKHDKVDDLKSVGLSASP**EPIYA**TIDDLGGPFPLKRHDKVDDLKSVGL-

**GC-12**

\*IF\*\*IHGKNKDFSKAEETLKALKGSVKDLGISPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQKIT  
DKVDNLNQAVSVAKATGDFSRVEQALADLKNFSKEQLAQQAQKNEDFNTGKNSALYQSVKNGVNGTLVGNGLSKA  
EATTLNFKNFSDIKKELNAKLGNFNNNNNNGLKN**EPIYA**KVNKKKTQGQAASLE**EPIYA**QVAKKVNAKIDRLNQIASGL  
GGVGQAAGFPLKKHDKVDDLKSVGLSASP**EPIYA**TIDDLGGPFPLKRHDKVDDLKSVGL

**GC-123**

\*IF\*\*IHGKNKDFSKAEETLKALKGSVEDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQKI  
TDKVDNLNQAVSVAKATGDFSRVEQALADLKNFSKDQLAQQAQKNESFNAGKKSEIYQSVKNGVNGTLVGNGLSKA  
EATTLKSNFSDIKKELNAKLGNFNNNNNNGLKN[EPIYA]KVNKKKAGQVASLE[EPIYA]QVAKKVNAKIDQLNQAASGL  
GGVGQAGFPLKRHDKVDDL SKVGLSANP[EPIYA]TIDDLGGPFPLKRHDKVDDL SKVGL--

**GC-16**

\*IF\*\*IHGKNKDFSKAEETLKALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQKI  
TDKVDNLNQAVSVAKAMGDFSRVEQALADLKNFSKDQLAQQAQKNEDFNTGKKSEIYQSVKNGVNGTLVGNGLSQ  
AEATTLKSNFSDIRKELNAKLGNFNNNNNNGLKN[EPIYA]KVNKKKAGQAASPE[EPIYA]QVAKKVNAKIDRLNQIASGL  
GGVGQVAGFPLKKHDKVDDL SKVGLSASP[EPIYA]TIDDLGGPFPLKRHDKVDDL SKVGL-

**GC-3**

\*IF\*\*IHGKNKDFSKAEETLKALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQKI  
TDKVDNLNQAVSVAKAMGDFSRVEQALADLKNFSKDQLAQQAQKNEDFNTGKKSEIYQSVKNGVNGTLVGNGLSQ  
AEATTLKSNFSDIRKELNAKLGNFNNNNNNGLKN[EPIYA]KVNKKKAGQTASPE[EPIYA]QVAKKVNAKIDRLNQIASGL  
GGVGQVAGFPLKKHDKVDDL SKVGLSASP[EPIYA]TIDDLGGPFPLKRHDKVDDL SKVGL-

**MS-2**

\*IF\*\*IHDKNKDFSKAEETLKALKGSVKDLGISPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQKIT  
DKVDNLNQAVSVAKATGDFSRVEQALADLKNFSKDQLAQQAQKNESFNTGKNSALYQSVKNGVNGTLVGNGLSKAE  
ATTLKSNFSDIKKELNAKLGNFNNNNNNGLKN[EPIYA]KVNKKKAGQADSPE[EPIYA]QVAKKVNAKIDRLNQAASGFGG  
VNPAGFPLKRHDKVDDL SKVGLSASP[EPIYA]TIDDLGGPFPLKRHDKVDDL SKVGL---

**MS-5**

\*IF\*FKNGKNKDFSKAEETLKALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQK  
ITDKVDNLNQAVSVAKAMGDFSRVEQALADLKNFSKDQLAQQAQKNEDFNTGKKSEIYQSVKNGVNGTLVGNGLSQ  
AEATTLKSNFSDIRKELNAKLGNFNNNNNNGLKN[EPIYA]KVNKKKAGQTASPE[EPIYA]QVAKKVNAKIDRLNQIASGL  
GGVGQVAGFPLKKHDKVDDL SKVGLSASP[EPIYA]TIDDLGGPFPLKRHDKVDDL SKVGL-



**MS-23**

SFDEFKNGKNKDFSKAEETLKALKGSVKDLGISPEWISKVENLNATLNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQ  
KITDKVDNLNQAVSVAKTTGDFSRVEQALADLNNFSKGQLAQQAQKNESFNAGKKSEIYQSVKNGVNGTLVGNLSQ  
AEATTLSKNFSDIKKELNAKLGNFNNNNNGLKNEPIYAKVNKKKAGQAASHEPIYAQVAKKVNAKIDRLNQIARGL  
GGVGQAAGFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDKVDDL?-----

**MS-18**

SFDEFKNGKNKDFSKAEETLKTLKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQ  
KITDKVDNLSSAVSVAKATGDFSRVEQALADLKNFSKEQLAQQAQKNEDFNTGKNSALYQSVKNGVNGTLVGNLSK  
AEATTLSKNFSDIKKELNAKLGNFNNNNNGLKNEPIYAKVNKKKTGQIAGLEPIYAQVAKKVNAKIDRLNQIASGL  
GGLGQAAGFPLKKHDKVDDLSKVGLSASPEPIYSTMDDLGGPFPLKRHDKVDDLSKVGL-

**MS-28**

SFDEFKNGKNKDFSKAEETLKALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQ  
KITDKVDNLNQAVSVAKAMGDFSRVEQALADLKNFSKDQLAQQAQKNEDFNTGKKSEIYQSVKNGVNGTLVGNLS  
QAEATTLSKNFSDIRKELNAKLGNFNNNNNGLKNEPIYAKVNKKKAGQAASPEPIYAQVAKKVNAKIDRLNQIAS  
GLGGVQVAGFPLKKHDKVDDLSKVGLSASPEPIYATIDDLGGPFPLKRHDN\*\*SQ\*GRA?-

**MS-13**

SFDEFKNGKNKDFSKAEETLKALKGSVKDLGISPEWISKVENLNATLNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQ  
KITDKVDNLNRAVSVAKATGDFSRVEQALADLNNFSKGQLAQQAQKNESFNAGKKSEIYQSVKNGVNGTLVGNLSQ  
AEATTLSKNFSDIKKELNAKLGNFNNNNNGLKNEPIYAKVNKKKAGQAASHEPIYAQVAKKVNAKIDRLNQIARGL  
GCVGQAAGFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDKVDDLSKVG?--

**GC-33**

SFDEFKNGKNKDFSKAEETLKALKGSMKDLGISPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQ  
KITDKVDNLNQAVSVAKATGDFSRVEQALADLKNFSKEQLAQQAQKNESFNAGKKSEIYQSVKNGVNGTLVGNLSG  
IEATALAKNFSDIKKELNEKFNFNNNNNNGLKNEPIYAKVNKKKAGQAASHEPIYAQVAKKVNAKIDQLNQAASG  
FGGVNQAGFPLKRHDKVDDLSKVGLSASPEPIYATIDDLGGPFPLKRHDKVDDLSKVGL--

**MS-40**

SFDEFKNGKNKDFSKAEETLKTGKSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQ  
KVTDKVDNLSSAVSVAKATGDFSRVEQALADLKNFSKEQLAQQAQKNEDFNTGKNSALYQSVKNGVNGTLVGNGLS  
KAEATTLSKNFSDIKKELNAKLGNFNNNNNNGLKNEPIYAKVNKKKVGQAASHEPIYAQVAKKVNKIDRLNQIAS  
GLGGVGQAAGYPLKRHDKVDDLSKVGLSASPEPIYATIDDLGGPFPLKRHDKVDDLSKVGL-

**MS-233**

SFDEFKNGKNKDFSKAEETLKALKGKSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQ  
KITDKVDNLNQAVSVAKAMGDFSRVEQALADLKNFSKDQLAQQAQKNEDFNTGKKSEIYQSVKNGVNGTLVGNGLS  
QAEATTLSKNFSDIRKELNAKLGNFNNNNNNGLKNEPIYAKVNKKKAGQAASPEPIYAQVAKKVNKIDRLNQIAS  
GLGGVGQAAGFPLKKHDKVDDLSKVGLSASPEPIYATIDDLGGPFPLKRHDKVDDLSKVGL-

**MS-15**

SFDEFKNDKNKDFSKAEETLKALKGKSVKDLGISPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQ  
KITDKVDNLNQAVSVAKATGDFSRVEQALADLKNFSKDQLAQQAQKNESFNTGKNSALYQSVKNGVNGTLVGNGLS  
KAEATTLSKNFSDIKKELNAKLGNFNNNNNNGLKNEPIYAKVNKKKAGQADSPEPIYAQVAKKVNKIDRLNQAASG  
FGGVNPAGFPLKRHDKVDDLSKVGLSASPEPIYATIDDLGGPFPLKRHDKVDDLSKVGL---

**MS-17**

SFDEFKNGKNKDFSKAEETLKALKGKSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQ  
KITDKVDNLNQAVSVAKAMGDFSRVEQALADLKNFSKDQLAQQAQKNEDFNTGKKSEIYQSVKNGVNGTLVGNGLS  
QAEATTLSKNFSDIRKELNAKLGNFNNNNNNGLKNEPIYAKVNKKKAGQTASPEPIYAQVAKKVNKIDRLNQIASG  
LGGVGQAAGFPLKKHDKVDDLSKVGLSASPEPIYATIDDLGGPFPLKRHDKVDDLSKVGL-

**GC-11**

GKSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQKITDKVDNLNQAVSVAKAMGDFS  
MVEQALADLKNFSKDQLAQQAQKNEDFNTGKKSEIYQSVKNGVNGTLVGNGLSQAEATTLSKNFSDIRKELNAKLG  
NFNNNNNNGLKNEPIYAKVNKKKAGQAASPEPIYAQVAKKVNKIDQLNQIASGLGGVGQVAGFPLKKHDKVDDLS  
KVGLSASPEPIYATIDDLGGPFPLKRHDKVDDLSKVGL-----

**KL-11**

SFDEFKNGKNKDFSKAEETLKTLKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQ  
KITDKVDNLSSAVSVAKATGDFSRVEQALADLKNFSKEQLAQQAQKNEDFNTGKNSALYQSVKNGVNGTLVGNGLSK  
AEATTLKSNFSDIKKELNAKLGNFNNNNNNGLKN[EPIYA]KVNKKKTGOIAGLE[EPIYA]QVAKKVNAKIDRLNQIASGL  
GGVGQAAGFPLKKHDKVDDL SKVGLSASP[EPIYA]TIDDLGGPFPLKRHDKVDDL SKVGL-

**MS-10**

SFDEFKNGKNKDFSKAEETLKTLKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQ  
KITDKVDNLSSAVSVAKATGDFSRVEQALADLKNFSKEQLAQQAQKNEDFNTGKNSALYQSVKNGVNGTLVGNGLSK  
AEATTLKSNFSDIKKELNAKLGNFNNNNNNGLKN[EPIYA]KVNKKKTGOVAGLE[EPIYA]QVAKKVNAKIDRLNQIASGL  
GGVGQAAGFPLKKHDKVDDL SKVGLSASP[EPIYA]TIDDLGGPFPLKRHDKVDDL SKVGL-

**GC-6**

SFDEFKNGKNKDFSKAEETLKALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQ  
KVTDKVDNLSSAVSMAKATGDFSRVNQALADLNNFSKEQLAQQAQKNEDFNTGKNSALYQSVKNGVNETLVGNGLS  
KAEATTLKSNFSDIKKELNAKLGNFNNNNNNGLKN[EPIYA]KVNKKKAGQAASLE[EPIYA]QVAKKVNAKIDQLNQAAS  
GLGGVGQAGFPLKRHDKVDDL SKVGRSVSP[EPIYA]TIDDLGGPFPLKRHDKVDDL SKVGF--

**MS-16**

SFDEFKNGKNKDFSKTEETLKTLKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQ  
KITDKVDNLNQAVSVAKATGDFSRVEQALVDLKNFSKEQLAQQAQKNEDFNTGKNSALYQSVKNGMNGTLVGNGLS  
KAEATTLKSNFSDIKKELNAKLGNFNNNNNNGLKN[EPIYA]KVNKKKTQGQTASLE[EPIYA]QVAKKVNAKIDRLNQIAS  
GLGGVGQAAGFPLKKHDQVDDL SKVGLSASP[EPIYA]TIDDLGGPFPLKRHDQVDDL SKVGL

**MS-30**

SFDEFKNGKNKDFSKAEETLKALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKNDLENSVKDVIIN  
QKITDKVDNLNQAVSVAKATGDFSGVEQALADLKNFSKEQLAQQAQKNEDFNAGKNSALYQSVKNGVNGTLVGNGLS  
LSKAEATTLSKNFSDIRKELNAKLGNFNNNNNNGLKN[EPIYA]KVNKKKAGQAASPE[EPIYA]QVAKKVNAKIDRLNQIA  
SGLGGVGKAAGFPLKRHDKVDDLSKVGLSASP[EPIYA]TIDDLGGPFPLKRHDKVDDLSKVGL-

**GC-1**

SFDEFKNGKNKDFSKAEETLKALKGSVEDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQ  
KITDKVDNLNQAVSVAKATGDFSRVEQALADLKNFSKDQLAQQAQKNESFNAGKKSEIYQSVKNGVNGTLVGNGLSK  
AEATTLSKNFSDIKKELNAKLGNFNNNNNNGLKN[EPIYA]KVNKKKAGQVASLE[EPIYA]QVAKKVNAKIDQLNQAASG  
LGGVGPAGFPLKRHDKVDDLSKVGLSANP[EPIYA]TIDDLGGPFPLKRHDKVDDLSKVGL--

**MS-401**

SFDEFKNGKNKDFSKAEETLKALKGSVKDLGINPEWISKVENLNTALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQ  
KITDKVDNLNQAVSVAKATDDFSMVEQALADLKNFSKEQLAQQAQKNESFNAGKKSEIYQSVKNGVNGTLVGNGLS  
QAEATTLSKNFSDIKKELNAKLGNFNNNNNNGLKN[EPIYA]KVNKKKTEQAASLE[EPIYA]QVAKKVNAKIDQLNQAAS  
GLGGVNQAGFPLKRHDKVDDLSKVGLSASP[EPIYA]TIDDLGGPFPLKKHDKVDDLSKVGL--

**MS-26**

SFDEFKNGKNKDFSKAEETLKALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQ  
KITDKVDNLNQAVSVAKAMGDFSRVEQALADLKNFSKDQLAQQAQKNEDFNTGKKSEIYQSVKNGVNGTLVGNGLS  
QAEATTLSKNFSDIRKELNAKLGNFNNNNNNGLKN[EPIYA]KVNKKKAGQAASPE[EPIYA]QVAKKVNAKIDRLNQIAS  
GLGGVGQAAGFLLKKHDKVDDLSKVGLSASP[EPIYA]KIYDLGGPFPLKRHDKVDDLSKVGL-

**MS-14**

SFDEFKNGKNKDFSKAEETLKALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQ  
KITDKVDNLNQAVSVAKAMGDFSRVEQALADLKNFSKDQLAQQAQKNENFNAGKKSEIYQSVKNGVNGTLVGNGLS  
QAEATTLSKNFSDIRKELNAKLGNFNNNNNNGLKN[EPIYA]KVNKKKAGQAASPE[EPIYA]QVAKKVNAKIDQLNQAAS  
GLGGVGQAAGFPLKKHDKVDDLSKVGLSASP[EPIYA]TIDDLGGPFPLKRHDKVDDL?-----

**GC-83**

\*IF\*\*IHGKNKDFSKAEETLKALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQKI  
TDKVDNLNQAVSVAKAMGDFSRVEQALADLKNFSKDQLAQQAQKNEDFNTGKKSEIYQSVKNGVNGTLVGNGLSQ  
AEATTLSKNFSDIRKELNAKLGNFNNNNNNGLKN**EPIYA**KVNKKKAGQTASPE**EPIYA**QVAKKVNAKIDRLNQIASGL  
GGVGQAAGFPLKKHDKVDDL SKVGLSASP**EPIYA**TIDDLGGPFPLKRHDKVDDL SKVG

**MS-11**

\*IF\*\*IHGKNKDFSKAEETLKALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQKI  
TDKVDNLNQAVSVAKAMGDFSRVEQALADLKNFSKDQLAQQAQKNEDFNTGKKSEIYQSVKNGVNGTLVGNGLSQ  
AEATTLSKNFSDIRKELNAKLGNFNNNNNNGLKN**EPIYA**KVNKKKAGQTASPE**EPIYA**QVAKKVNAKIDRLNQIASGL  
GGVGQAAGFPLKKHDKVDDL SKVGLSASP**EPIYA**TIDDLGGPFPLKRHDKVDDL SKVG--

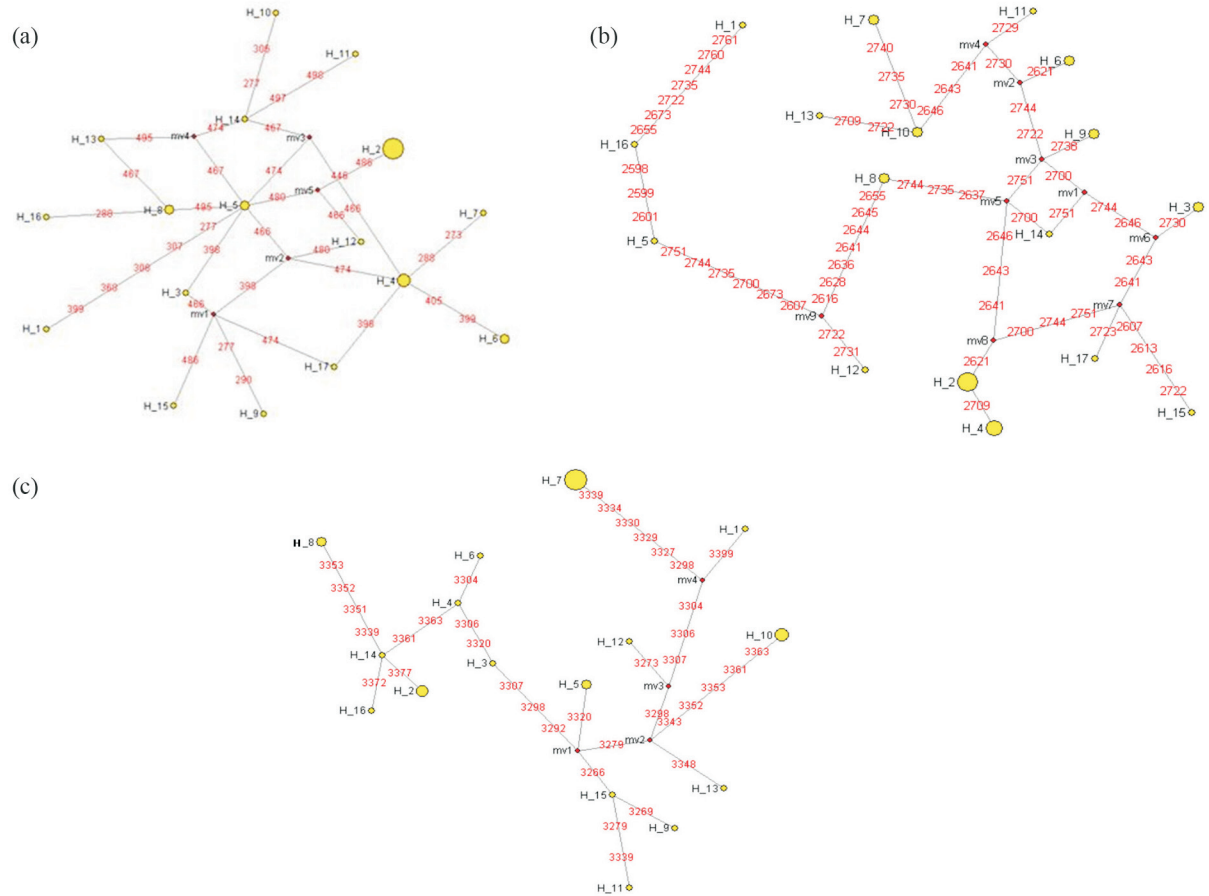
**MS-8**

\*IHGKNKDFSKAEETLKALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQKITDK  
VDNLNQAVSVAKAMGDFSRVEQALADLKNFSKDQLAQQAQKNEDFNTGKKSEIYQSVKNGVNGTLVGNGLSQAEAT  
TLSKNFSDIRKELNAKLGNFNNNNNNGLKN**EPIYA**KVNKKKAGQTASPE**EPIYA**QVAKKVNAKIDRLNQIASGLGGVG  
QAAGFPLKKHDKVDDL SKVGLSASP**EPIYA**TIDDLGGPFPLKRHDKVDDL SKVG

**MS-20**

\*IF\*\*IHGKNKDFSKAEETLKALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINQKI  
TDKVDNLNQAVSVAKAMGDFSRVEQALADLKNFSKDQLAQQAQKNEDFNTGKKSEIYQSVKNGVNGTLVGNGLSQ  
AEATTLSKNFSDIRKELNAKLGNFNNNNNNGLKN**EPIYA**KVNKKKAGQTASPE**EPIYA**QVAKKVNAKIDRLNQIASGL  
GGVGQAAGFPLKKHDKVDDL SKVGLSASP**EPIYA**TIDDLGGPFPLKRHDKVDDL SKVG





**Figure S2** - Deduced amino acid sequence of partial *cagA* gene of *H. pylori* showing (a) tyrosine phosphorylation motif A (TPM A), (b) tyrosine phosphorylation motif B (TPM B), and (c) tyrosine phosphorylation motif C (TPM C).