

A Variable Gene in a Conserved Region of the *Helicobacter pylori* Genome: Isotopic Gene Replacement or Rapid Evolution?

Armelle MÉNARD^{1,2}, Antoine DANCHIN³, Sandrine DUPOUY^{1,2}, Francis MÉGRAUD^{1,2}, and Philippe LEHOURS^{1,2,*}

INSERM U853, Laboratoire de Bactériologie, Université Victor Segalen Bordeaux 2, 146 rue Léo Saignat, F-33076 Bordeaux cedex, France¹; Université Victor Segalen Bordeaux 2, Laboratoire de Bactériologie, Bordeaux F-33076, France² and Institut Pasteur, Génétique des Génomes Bactériens - CNRS URA2171, Paris F-750015, France³

(Received 6 November 2007; accepted on 3 April 2008; published online 27 April 2008)

Abstract

The present study concerns the identification of a novel coding sequence in a region of the *Helicobacter pylori* genome, located between JHP1069/HP1141 and JHP1071/HP1143 according to the numbering of the J99 and 26695 reference strains, respectively, and spanning three different coding DNA sequences (CDSs). The CDSs located at the centre of this locus were highly polymorphic, as determined by the analysis of 24 European isolates, 3 Asian, and 3 African isolates. Phylogenetic and molecular evolutionary analyses showed that the CDSs were not restricted to the geographical origin of the strains. Despite a very high variability observed in the deduced protein sequences, significant similarity was observed, always with the same protein families, i.e. ATPase and bacteriophage receptor/invasion proteins. Although this variability could be explained by isotopic gene replacement via horizontal transfer of a gene with the same function but coming from a variety of sources, it seems more likely that the very high sequence variation observed at this locus is the result of a strong selection pressure exerted on the corresponding gene product. The CDSs identified in the present study could be used as strain specific markers.

Key words: *Helicobacter pylori*; coding DNA sequence; genetic diversity; diversifying selection

Comparative analyses conducted on *Helicobacter pylori* genome sequences, i.e. from *H. pylori* strain J99 associated with peptic ulcer,^{1,2} strain 26695³ associated with gastritis, and strain HPAG1 associated with atrophic gastritis,⁴ revealed a significant macrodiversity (presence or absence of genes) and microdiversity (high polymorphism among orthologous genes).^{5,6,7} The plasticity zones and the *cag* pathogenicity island (*cag* PAI) are considered to be the main variable genomic areas. The remaining variable genes are distributed throughout the *H. pylori* genome and some of them have been individualized

in clusters of instability concerning blocks of 5–8 coding DNA sequences (CDSs).^{5,8,9}

Subtractive hybridization is a powerful tool for comparative prokaryotic genomics and was validated on *H. pylori* by several authors.^{10,11} In a previous study, we used subtractive hybridization to compare the genetic content of one *H. pylori* strain isolated from a gastric MALT lymphoma strain (strain B34) and one chronic gastritis only strain.¹² One original 1092 bp sequence was identified, with no significant nucleotide similarity in comparison to the *H. pylori* reference strains 26695 and J99 genomes which were available. The aim of the present study was to localize this sequence in the *H. pylori* genome, to determine its prevalence, and to analyze its genetic diversity in *H. pylori*.

Using an in-house genome walking method as previously described,¹³ the original region was localized

Edited by Masahira Hattori

* To whom correspondence should be addressed. Tel. +33 5-57-57-12-86. Fax. 33 5-56-51-41-82. E-mail: philippe.lehours@labhel.u-bordeaux2.fr

© The Author 2008. Kazusa DNA Research Institute

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

in the *H. pylori* genome and a new CDS was subsequently identified using the CDS finder website (<http://www.ncbi.nlm.nih.gov/gCDS/CDSig.cgi>). This new CDS, called CDS2, is located between two CDS homologous to JHP1069/HP1141 and JHP1071/HP1143 according to the numbering of the J99 and 26695 reference strains, respectively.^{1,3} CDS2 replaced JHP1070/HP1142, called CDS1, in *H. pylori* reference strains J99 and 26695.

The percentage of identity between the nucleotide sequences of CDS1 and CDS2 was determined using the LALIGN software,¹⁴ which identifies multiple matching subsegments in two sequences (http://www.ch.embnet.org/software/LALIGN_form.html). CDS2 showed 54.9% identity in a 2046 nucleotides overlap with JHP1070 and 55.5% identity in a 2083 nucleotides overlap with HP1142. CDS2 encodes a putative polypeptide of 820 residues (Genbank accession number EF492441, EMBL Nucleotide Sequence AM902682). Regarding the protein homology, CDS2 shared 23.6% identity with JHP1070 in a 628 amino acid overlap and 24.4% identity with HP1142 in a 630 amino acid overlap. Finally, a strong nucleotide identity was found with the HPAG1_1080 sequence⁴ with 89.3% identity in a 2469 nucleotides overlap.

The prevalence and the genetic diversity of the identified genomic locus were first determined for 24 *H. pylori* strains: 13 *H. pylori* strains isolated from gastric MALT lymphoma patients obtained from two multicentre French protocols and 11 strains isolated from French chronic gastritis only patients, as previously described^{12,15} by PCR amplification using primers hybridizing to the conserved sequence of the flanking genes (JHP1069/HP1141 and JHP1071/HP1143) according to the numbering of the J99 and 26695 strains, respectively. The primers were designed using the web Primer3 software (http://www.broad.mit.edu/cgi-bin/primer/primer3_www.cgi).¹⁶ Direct sequencing was carried out on both strands, and nucleotide and deduced protein sequences were compared with the NCBI Blast program (<http://www.ncbi.nlm.nih.gov/BLAST/>). A CDS was always present at this locus: CDS1 was found in 54% of the strains, CDS2 in 29% of the strains, and an additional CDS, called CDS3, was identified in 17% of the strains. In the chronic gastritis only *H. pylori* strain G2, CDS3 had a 53.4% identity in a 2005 nucleotide overlap with CDS1 and a 52.9% identity in a 2063 nucleotides overlap with CDS2, and it encodes a putative polypeptide of 861 residues (GenBank accession number EF492442, EMBL Nucleotide Sequence AM902683). CDS3 still has no counterpart in databases. Considering the three CDSs, no significant association with a virulence factor was found, nor with a pathology (data not

shown). The presence or absence of these CDSs was also verified by dot blot hybridization, as previously described.¹² It showed that the presence of one of these three CDSs was exclusive (no local duplication, data not shown).

We first focused on the role of the genes present around the polymorphic locus. According to the revised annotation of the *H. pylori* genome,¹⁷ JHP1069/HP1141 encodes a methionyl-tRNA formyltransferase (*fmt*) and JHP1071/HP1143, a conserved hypothetical protein. *fmt* is considered to be an essential gene which links general metabolism with the translation process (protein biosynthesis).^{18,19} As shown in Fig. 1, JHP1069/HP1141 and JHP1071/HP1143 are surrounded by genes of hypothetical function. Considering the G + C% content of the region, all of the CDSs contained a G + C% similar to the rest of the *H. pylori* genome (~39%) except for these variable regions: CDS1, CDS2, and CDS3 had 29, 30, and 31% G + C% content, respectively. The lower G + C% content suggests an external origin of these CDSs or a rapid adaptation.²⁰ Indeed, Saunders et al.²¹, using a tetranucleotide and hexanucleotide signature analysis, identified substantial differences between JHP1070 and HP1142 genes and hypothesized that they were horizontally transferred.

CDS1 has been annotated as a predicted coding region JHP1070 with no homolog in the databases. It codes for a putative polypeptide of 759 residues. Using a Blastp search, significant homologies were found with (i) Rlo proteins (R-linked ORF) from *Campylobacter* (e.g. RloG, $E = e-11$ in *Campylobacter jejuni* strain RM1167, or RloC, $E = 7e-13$ in *C. jejuni* strain RM11221),^{22,23} (ii) an ATP/GTP binding-site (GXXXXGKT), and (iii) a putative phage murein transglycosylase *YomI* (SPbeta phage protein; lytic transglycosylase, $E = 4.52e-05$ in *Bacillus subtilis*).²⁴ The same significant homologies with Rlo and *YomI* were also found in CDS2 and CDS3. Finally, another interesting point to consider is that the three CDSs shared significant homology with chromosome partition protein SMC: for example, *Treponema denticola* ATCC 35405 chromosome NC_002967 $E = 1.11e-11$ for CDS1, *Fusobacterium nucleatum* subsp. *nucleatum* ATCC 25586 $E = 8.66e-10$ for CDS2, *Fusobacterium nucleatum* subsp. *nucleatum* ATCC 25586 $E = 3.86e-08$ for CDS3.²⁵ We did not find any significant motifs indicating that the proteins could be secreted and/or present in the membrane, but this does not preclude an association with the membrane via interaction with an integral membrane protein partner.

How can one explain the apparent variability of the locus identified in the present study? One potential hypothesis is that the region is a hot spot for gene insertion/deletion, with a specific selection pressure

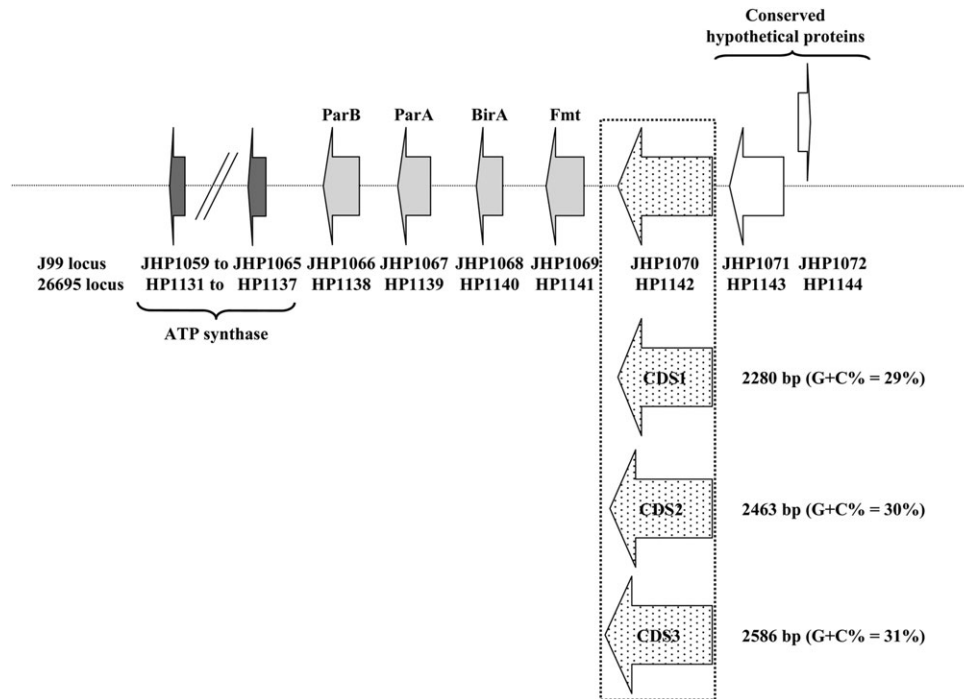


Figure 1. Representation of the genomic area of interest, according to the genome sequences of the two *Helicobacter pylori* reference strains J99 and 26695, which contain the variable CDS identified in the present study. Each CDS is represented by an arrow with the direction indicating the translational direction. The numbering under each CDS corresponds to the number of CDS in *H. pylori* J99 (top line) and 26695 (bottom line). The function of each CDS is indicated according to the revised annotation of Boneca *et al.*¹⁷ Fmt, methionyl-tRNA formyltransferase; BirA, biotin ligase bifunctional protein; ParB, replication/partition-related protein; ParA, chromosome partition protein (Soj).

maintaining a particular function at that precise location in the genome. Suerbaum and Josenhans²⁶ recently reviewed the current data on the genetic diversity of *H. pylori* and argued that this bacterium uses mutation and recombination processes to adapt to its individual host by modifying molecules that interact with the host.²⁶ Because the three CDSs retain the same similarities, it is likely that (i) these proteins share the same function or (ii) the gene is submitted to specific selection pressure making it evolve at a very rapid rate. We proposed that such a protein could be a phage receptor/translocator or that it could allow the DNA phage to enter host cells by remodelling the cell wall.^{27–29} Indeed, as already described in *Escherichia coli*, this kind of protein is subjected to a strong positive selection.³⁰

Helicobacter pylori genotypes vary markedly with their geographical region, and this is particularly the case for genes under positive selection. Therefore, the corresponding genes were looked for in three East Asian strains and three African strains. All three CDS were found: CDS1 was found in one Asian (strain 8038) and one African strain (strain TALLAN), CDS2 in two Asian strains (strains 12001 and strain 8033), and CDS3 in one Asian (strain 19A) and one African strain (strain BAPOOI) (Fig. 2). A phylogenetic analysis was conducted on the deduced amino acid sequences

of CDS. Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 4.³¹ Phylogenetic trees were generated by the neighbour-joining method.³² Molecular distances were determined using the Kimura two-parameter model.³³ The tree showed three independent clusters which were clearly separated and corresponded to CDS1, CDS2, and CDS3, respectively (Fig. 2). However, the exact organization of these different CDS cannot be determined since this consensus tree cannot be rooted to other species. Indeed, no CDS with significant homology has ever been found in other species (in databases). Interestingly, even though the testing was performed on a limited number of non-European strains, these results indicate that the presence of one of the three CDS cannot be restricted to the geographical origin of the strains.

The type of selection operating at the amino acid level was also evaluated by comparing non-synonymous substitutions (Ka) and synonymous substitutions (Ks).³⁴ The overall mean of Ks and Ka substitutions was determined using the Nei–Gojobori method.³⁵ The codon based Z-test of selection³⁶ was used to evaluate the significance of Ka/Ks substitution values. Bootstrap confidence levels were determined by randomly resampling the sequencing data 1000 times. The results are indicated

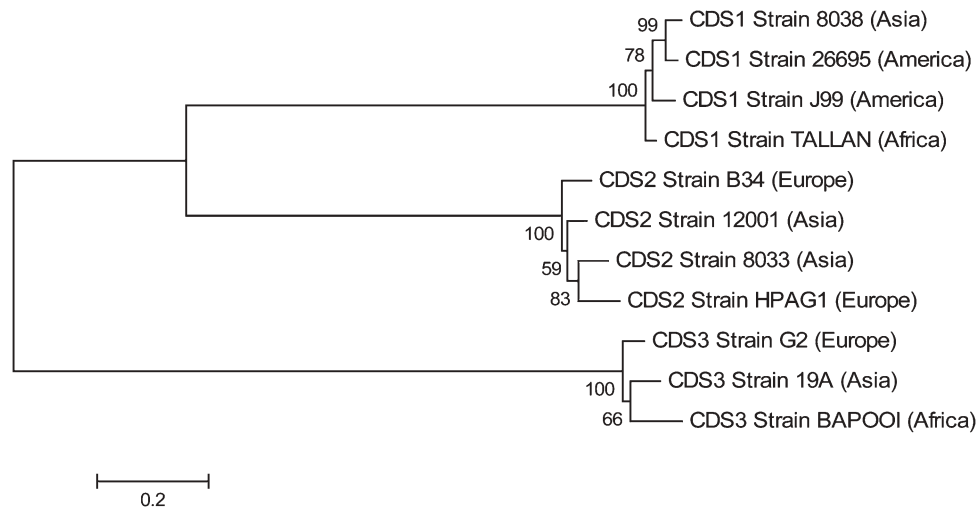


Figure 2. Phylogenetic analysis of CDS, CDS1, CDS2, and CDS3, proteins generated with the neighbour-joining method. The phylogeny presented is based on the alignment of the entire deduced protein. The bootstrap values are indicated next to each node. Nucleotide and protein sequences are available in GenBank (EMBL) for the *Helicobacter pylori* strains B34, G2, 8038, TALLAN, 12001, 8033, 19A, and BAPOOI under the accession numbers EF492441 (AM902682), EF492442 (AM902683), EU553483 (AM946633), EU553485 (AM946634), EU553505 (AM946635), EU553482 (AM946636), EU553481 (AM946637), and EU556504 (AM946638), respectively. Nucleotide and protein sequences of *H. pylori* reference strains 26695, J99, and HPAG1 are available in GenBank under the accession numbers AE000511, AE001439, and ABF85147, respectively.

for each CDS in Table 1. Since K_a/K_s was < 1 for the three CDSs analyzed, the purifying selection hypothesis was tested and the significant P -value obtained supports the hypothesis of conservation at the protein level for each CDS (Z -test $P < 0.001$).

Finally, we propose that the very high variation observed in the protein sequences reflects the permanent selection pressure exerted by phages or other elements interacting with the organism's cell envelope. If this is the case, this locus could be used as a marker for constraints operating in the environmental niches in which particular *H. pylori* strains evolve. The presence of phages in *H. pylori* has been rarely described.³⁷ For example, Marsich et al.³⁸ postulated that *H. pylori* lysozyme gene (*lys*) had a prophage origin. Numerous other explanations cannot be excluded, such as bacterial

mammalian host interaction, protozoan predation, or porin specificity. Indeed several publications have focused on cases of genes that vary markedly among *H. pylori* isolates. One example is the replacement of *babA* by *babB* as reported by Solnick et al.³⁹ *Helicobacter pylori* BabA is the ABO blood group antigen binding adhesin, which has a closely related paralogue (BabB) whose function is unknown. An extensive genotypic diversity in *babA* and *babB* across different strains, as well as within a strain colonizing an individual patient has been shown in line with the hypothesis that diverse profiles of *babA* and *babB* reflect selective pressures for adherence, which may differ across different hosts and within an individual over time.⁴⁰

In summary, a novel polymorphic locus comprised of a single gene was identified in the

Table 1. Analysis of molecular distances and synonymous and non-synonymous nucleotide substitutions within CDSs, CDS1 ($n = 4$), CDS2 ($n = 4$), and CDS3 ($n = 3$), in different *Helicobacter pylori* strains

	CDS1	CDS2	CDS3
Mol. distance (nt)	$0.045 \pm 0.003^{\&}$	0.080 ± 0.005	0.069 ± 0.004
No. differences (nt)	98.167 ± 7.246	182.167 ± 9.753	170.00 ± 9.334
Ks	0.088 ± 0.012	0.161 ± 0.015	0.123 ± 0.013
Ka	0.035 ± 0.004	0.061 ± 0.004	0.057 ± 0.004
Ka/Ks	$0.398 \pm 0.071^{\dagger}$	$0.379 \pm 0.043^{\dagger}$	$0.463 \pm 0.059^{\dagger}$

nt, nucleotides; Ks, synonymous substitutions; Ka, non-synonymous substitutions.

[†] $P_{Z\text{-Test}} < 0.001$ for purifying selection hypothesis ($K_a/K_s < 1$).

[&]Value \pm standard error.

The GenBank accession numbers of the sequences used in this study are listed in Fig. 2.

H. pylori genome. Although this variation could be explained by isotopic gene replacement via horizontal transfer of a gene with the same function but coming from a variety of sources, it seems more likely that the very high sequence variation observed at this locus is the result of a strong selection pressure exerted on the corresponding gene product. We propose that the evolution of CDS1, CDS2, and CDS3 is due to the occurrence of a specific environmental event, such as interaction with a biological structure, e.g. bacteriophage which are involved in surface cell secretion. The genes identified in the present study could be used as strain specific markers for particular niches. The predicted function of the gene products, although highly speculative, should encourage investigators to explore the presence of phages in the *H. pylori* environment and study their relationship regarding pathogenicity.

Acknowledgments: The authors want to thank Dr Monica Oleastro from the Departamento de Doenças Infecciosas of the Instituto Nacional Saúde Dr Ricardo Jorge (Lisbon, Portugal) for the comparison of non-synonymous substitutions and synonymous substitutions among the three CDSs described in this study and Dr Jorge M. B. Vítor and Dr Vale Filipa from the Faculty of Pharmacy in Lisbon for strains. The study was financially supported by the Institut de Recherche des Maladies de l'Appareil Digestif (IRMAD), the Association pour la Recherche contre le Cancer (ARC), and the Conseil Régional d'Aquitaine, France.

References

1. Alm, R. A., Ling, L. S. L., Moir, D. T., *et al.* 1999, Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*, *Nature*, **397**, 176–180.
2. Alm, R. A. and Trust, T. J. 1999, Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes, *J. Mol. Med.*, **77**, 834–846.
3. Tomb, J. F., White, O., Kerlavage, A. R., *et al.* 1997, The complete genome sequence of the gastric pathogen *Helicobacter pylori*, *Nature*, **388**, 539–547.
4. Oh, J. D., Kling-Backhed, H., Giannakis, M., *et al.* 2006, The complete genome sequence of a chronic atrophic gastritis *Helicobacter pylori* strain: evolution during disease progression, *Proc. Natl. Acad. Sci. USA*, **103**, 9999–10004.
5. Salama, N., Guillemin, K., McDaniel, T. K., Sherlock, G., Tompkins, L. and Falkow, S. 2000, A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains, *Proc. Natl. Acad. Sci. USA*, **97**, 14668–14673.
6. Falush, D., Wirth, T., Linz, B., *et al.* 2003, Traces of human migrations in *Helicobacter pylori* populations, *Science*, **299**, 1582–1585.
7. Raymond, J., Thiberge, J. M., Chevalier, C., *et al.* 2004, Genetic and transmission analysis of *Helicobacter pylori* strains within a family, *Emerg. Infect. Dis.*, **10**, 1816–1821.
8. Gressmann, H., Linz, B., Ghai, R., *et al.* 2005, Gain and loss of multiple genes during the evolution of *Helicobacter pylori*, *PLoS Genet.*, **1**, e43.
9. Chanto, G., Occhialini, A., Gras, N., Alm, R. A., Megraud, F. and Marais, A. 2002, Identification of strain-specific genes located outside the plasticity zone in nine clinical isolates of *Helicobacter pylori*, *Microbiol. Sgm.*, **148** (11), 3671–3680.
10. Akopyants, N. S., Fradkov, A., Diatchenko, L., *et al.* 1998, PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*, *Proc. Natl. Acad. Sci. USA*, **95**, 13108–13113.
11. Kersulyte, D., Mukhopadhyay, A. K., Shirai, M., Nakazawa, T. and Berg, D. E. 2000, Functional organization and insertion specificity of IS607, a chimeric element of *Helicobacter pylori*, *J. Bacteriol.*, **182**, 5300–5308.
12. Lehours, P., Dupouy, S., Bergey, B., *et al.* 2004, Identification of a genetic marker of *Helicobacter pylori* strains involved in gastric extranodal marginal zone B cell lymphoma of the MALT-type, *Gut*, **53**, 931–937.
13. Abdelbaqi, K., Menard, A., Prouzet-Mauleon, V., Bringaud, F., Lehours, P. and Megraud, F. 2007, Nucleotide sequence of the *gyrA* gene of *Arcobacter* species and characterization of human ciprofloxacin-resistant clinical isolates, *FEMS Immunol. Med. Microbiol.*, **49**, 337–345.
14. Huang, X. and Miller, M. 1991, A time-efficient, linear-space local similarity algorithm, *Adv. Appl. Math.*, **12**, 337–357.
15. Lehours, P., Menard, A., Dupouy, S., *et al.* 2004, Evaluation of the association of nine *Helicobacter pylori* virulence factors with strains involved in low-grade gastric mucosa-associated lymphoid tissue lymphoma, *Infect. Immun.*, **72**, 880–888.
16. Rozen, S. and Skaletsky, H. 2000, Primer3 on the WWW for general users and for biologist programmers, *Methods Mol. Biol.*, **132**, 365–386.
17. Boneca, I. G., de Reuse, H., Epinat, J.-C., Pupin, M., Labigne, A. and Moszer, I. 2003, A revised annotation and comparative analysis of *Helicobacter pylori* genomes, *Nucl. Acids Res.*, **31**, 1704–1714.
18. Meinnel, T., Guillon, J. M., Mechulam, Y., *et al.* 1993, The *Escherichia coli fmt* gene, encoding methionyl-tRNA(fMet) formyltransferase, escapes metabolic control. Disruption of the gene for Met-tRNA(fMet) formyltransferase severely impairs growth of *Escherichia coli*, *J. Bacteriol.*, **175**, 993–1000.
19. Guillon, J. M., Mechulam, Y., Schmitter, J. M., Blanquet, S. and Fayat, G. 1992, Disruption of the gene for Met-tRNA(fMet) formyltransferase severely impairs growth of *Escherichia coli*, *J. Bacteriol.*, **174**, 4294–4301.

20. Rocha, E. P. and Danchin, A. 2002, Base composition bias might result from competition for metabolic resources, *Trends Genet.*, **18**, 291–294.
21. Saunders, N. J., Boonmee, P., Peden, J. F. and Jarvis, S. A. 2005, Inter-species horizontal transfer resulting in core-genome and niche-adaptive variation within *Helicobacter pylori*, *BMC Genom.*, **6**, 9.
22. Miller, W. G., Pearson, B. M., Wells, J. M., Parker, C. T., Kapitonov, V. V. and Mandrell, R. E. 2005, Diversity within the *Campylobacter jejuni* type I restriction-modification loci, *Microbiology*, **151**, 337–351.
23. Fouts, D. E., Mongodin, E. F., Mandrell, R. E., *et al.* 2005, Major structural differences and novel potential virulence mechanisms from the genomes of multiple campylobacter species, *PLoS Biol.*, **3**, e15.
24. Lazarevic, V., Dusterhoft, A., Soldo, B., Hilbert, H., Muel, C. and Karamata, D. 1999, Nucleotide sequence of the *Bacillus subtilis* temperate bacteriophage SPbetac2, *Microbiology*, **145**, 1055–1067.
25. Kapatral, V., Anderson, I., Ivanova, N., *et al.* 2002, Genome sequence and analysis of the oral bacterium *Fusobacterium nucleatum* strain ATCC 25586, *J. Bacteriol.*, **184**, 2005–2018.
26. Suerbaum, S. and Josenhans, C. 2007, *Helicobacter pylori* evolution and phenotypic diversification in a changing host, *Nat. Rev. Microbiol.*, **5**, 441–452.
27. Qasba, P. K. and Kumar, S. 1997, Molecular divergence of lysozymes and alpha-lactalbumin, *Crit. Rev. Biochem. Mol. Biol.*, **32**, 255–306.
28. Rydman, P. S. and Bamford, D. H. 2000, Bacteriophage PRD1 DNA entry uses a viral membrane-associated transglycosylase activity, *Mol. Microbiol.*, **37**, 356–363.
29. Blackburn, N. T. and Clarke, A. J. 2001, Identification of four families of peptidoglycan lytic transglycosylases, *J. Mol. Evol.*, **52**, 78–84.
30. Petersen, L., Bollback, J. P., Dimmic, M., Hubisz, M. and Nielsen, R. 2007, Genes under positive selection in *Escherichia coli*, *Genome Res.*, **17**, 1336–1343.
31. Tamura, K., Dudley, J., Nei, M. and Kumar, S. 2007, MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0, *Mol. Biol. Evol.*, **24**, 1596–1599.
32. Saitou, N. and Nei, M. 1987, The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Mol. Biol. Evol.*, **4**, 406–425.
33. Kimura, M. 1980, A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences, *J. Mol. Evol.*, **16**, 111–120.
34. Vandamme, A. 2003, Basic concepts of molecular evolution. In: Salemi, M. and Vandamme, A. (eds.), *The Phylogenetic Handbook - A Practical Approach To Dna And Protein Phylogeny*, Cambridge: Cambridge University Press, pp. 1–23.
35. Nei, M. and Gojobori, T. 1986, Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions, *Molecular Biol. Evol.*, **3**, 418–426.
36. Nei, M. and Kumar, S. 2000, Synonymous substitutions and non synonymous nucleotide substitutions. In: Nei, M. (ed.), *Molecular Evolution and Phylogenetics*, Volume 1, New York: Oxford University Press, pp. 52–61.
37. Schmid, E. N., von Recklinghausen, G. and Ansorg, R. 1990, Bacteriophages in *Helicobacter (Campylobacter) pylori*, *J. Med. Microbiol.*, **32**, 101–104.
38. Marsich, E., Zuccato, P., Rizzi, S., Vetere, A., Tonin, E. and Paoletti, S. 2002, *Helicobacter pylori* expresses an autolytic enzyme: gene identification, cloning, and theoretical protein structure, *J. Bacteriol.*, **184**, 6270–6279.
39. Solnick, J. V., Hansen, L. M., Salama, N. R., Boonjakuakul, J. K. and Syvanen, M. 2004, Modification of *Helicobacter pylori* outer membrane protein expression during experimental infection of rhesus macaques, *Proc. Natl. Acad. Sci. USA*, **101**, 2106–2111.
40. Colbeck, J. C., Hansen, L. M., Fong, J. M. and Solnick, J. V. 2006, Genotypic profile of the outer membrane proteins BabA and BabB in clinical isolates of *Helicobacter pylori*, *Infect. Immun.*, **74**, 4375–4378.