Genome Sequencing Reveals a Phage in Helicobacter pylori

Philippe Lehours, Filipa F. Vale, Magnus K. Bjursell, Ojar Melefors, Reza Advani, Steve Glavas, Julia Guegueniat, Étienne Gontier, Sabrina Lacomme, António Alves Matos, Armelle Menard, Francis Mégraud, Lars Engstrand, and Anders F. Andersson

INSERM U853, Bordeaux, France; Université de Bordeaux, Centre National de Référence des Campylobacters et des Hélicobacters, Bordeaux, France; Faculty of Engineering, Catholic University of Portugal, Lisbon, Portugal; Science for Life Laboratory and Center for Molecular Medicine, Karolinska Institute, Stockholm, Sweden; Swedish Institute for Infectious Disease Control, Solna, Sweden; Université de Bordeaux, Bordeaux Imaging Center, Bordeaux, France; Pathological Anatomy Department, Curry Cabral Hospital, Lisbon, Portugal; University of Aveiro, Center for Environmental and Marine Studies, Aveiro, Portugal; and Karolinska Institute, Department of Microbiology, Tumor and Cell Biology, Stockholm, Sweden

*Present address: Science for Life Laboratory, School of Biotechnology, KTH Royal Institute of Technology, Stockholm, Sweden

ABSTRACT Helicobacter pylori chronically infects the gastric mucosa in more than half of the human population; in a subset of this population, its presence is associated with development of severe disease, such as gastric cancer. Genomic analysis of several strains has revealed an extensive H. pylori pan-genome, likely to grow as more genomes are sampled. Here we describe the draft genome sequence (63 contigs; 26× mean coverage) of H. pylori strain B45, isolated from a patient with gastric mucosa-associated lymphoid tissue (MALT) lymphoma. The major finding was a 24.6-kb prophage integrated in the bacterial genome. The prophage shares most of its genes (22/27) with prophage region II of Helicobacter acinonychis strain Sheeba. After UV treatment of liquid cultures, circular DNA carrying the prophage integrase gene could be detected, and intracellular tailed phage-like particles were observed in H. pylori cells by transmission electron microscopy, indicating that phage production can be induced from the prophage. PCR amplification and sequencing of the integrase gene from 341 H. pylori strains from different geographic regions revealed a high prevalence of the prophage (21.4%). Phylogenetic reconstruction showed four distinct clusters in the integrase gene, three of which tended to be specific for geographic regions. Our study implies that phages may play important roles in the ecology and evolution of H. pylori.

IMPORTANCE Helicobacter pylori chronically infects the gastric mucosa in more than half of the human population, and while most of the infected individuals do not develop disease, H. pylori infection doubles the risk of developing gastric cancer. An abundance and diversity of viruses (phages) infect microbial populations in most environments and are important mediators of microbial diversity. Our finding of a 24.6-kb prophage integrated inside an H. pylori genome and the observation of circular integrase gene-containing DNA and phage-like particles inside cells upon UV treatment demonstrate that we have discovered a viable H. pylori phage. The additional finding of integrase genes in a large proportion of screened isolates of diverse geographic origins indicates that the prevalence of prophages may have been underestimated in H. pylori. Since phages are important drivers of microbial evolution, the discovery should be important for understanding and predicting genetic diversity in H. pylori.

Received 3 October 2011 Accepted 12 October 2011 Published 15 November 2011

Citation Lehours P, et al. 2011. Genome sequencing reveals a phage in Helicobacter pylori. mBio 2(6):e00239-11. doi:10.1128/mBio.00239-11.

Editor Rino Rappuoli, Novartis Vaccines and Diagnostics

Copyright © 2011 Lehours et al. This is an open-access article distributed under the terms of the Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License, which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Philippe Lehours, philippe.lehours@u-bordeaux2.fr.
strong urease to resist the acidic gastric pH and flagella to move in the thick mucus and evade host immunity. *H. pylori* pathogenicity has been associated with the production of numerous virulence factors, in particular those encoded by the *cag* pathogenicity island (*cag* PAI), which have been associated with peptic ulceration, atrophic gastritis, and gastric adenocarcinoma (14, 15). These factors can target tumor suppression functions (reduction of p53 on delivery of CagA), in a mode similar to that of DNA tumor viruses (16). Currently, eradication of *H. pylori* is performed with antibiotics that have severe effects on the intestinal microbiota, including persistence of elevated levels of antibiotic-resistant strains years after treatment (17); thus, novel therapies targeting *H. pylori* specifically are needed.

Very little is known about *H. pylori* phages; however, shortly after the discovery of *H. pylori*, Marshall et al. (18) and Goodwin et al. (19) described intracellular phage-like particles observed in human gastric mucosa. Two other studies from the early 1990s showed spontaneous production of small amounts of phage particles by the *H. pylori* strain Schreck (20). Furthermore, Vale et al. (21) described *H. pylori* temperate phage induction of phage-like particles using UV. Reports of prophages in other *Helicobacter* species are also rare. Until now, there have been only two reports of a prophage: one in the genome of *Helicobacter acinonychis* strain Sheeba (22) and one in *Helicobacter felis* CS1 (ATCC 49179) (23). Although the reports concerning *H. pylori* phages are rare, genetic characteristics of *H. pylori* isolates suggest that the organism is challenged by viruses; a multitude of restriction-modification (R-M) system genes are found that display high evolutionary rates (24–26). In addition, R-M system genes are highly diverse among strains, representing more than half of the strain-specific genes present in *H. pylori* sequenced genomes (27). The repertoire of R-M system genes in a bacterial cell influences the pattern of DNA methylation, which in *H. pylori* has been shown to influence gene expression (28); however, R-M system genes are generally believed to function as a protection against invading DNA (2). *H. pylori* cell surface proteins are rapidly evolving (29–31). This has been inferred as an adaptation to differences in mucosal surface structures among human hosts or as a selection to counteract the immune response, but it could also reflect escape mechanisms in response to bacteriophage infection.

In this study, we sequenced the genome of an isolate from a gastric mucosa-associated lymphoid tissue (MALT) lymphoma patient with the aim to identify genes that could be associated with disease development. An unexpected finding was the presence of a prophage integrated in the bacterial genome, from which we were able to induce production of phage particles.

**RESULTS**

The *H. pylori* B45 genome. Whole-genome shotgun sequencing carried out using the 454 Titanium platform (454 Life Sciences, Branford, CT) generated 42 Mb of data that assembled into 63 contigs with a total size of 1,602,587 bp and 26× average coverage (GenBank accession numbers AFAO01000001 to AFAO01000063). Mean contig length was 25,437 bp (minimum, 581 bp; maximum, 214,447 bp), and average G+C content was 39%. Some contigs had G+C contents that strongly deviated from the genome average (minimum, 32.99%; maximum, 48.19%), which suggests integration of external DNA in the bacterial chromosome (32).

The AMIGene software (33) predicted 1,602 coding sequences (CDs) on the B45 contigs. We used BLASTp and the Markov clustering (MCL) algorithm (34, 35) to define orthologous genes within *H. pylori* genomes, including that of B45 and seven reference genomes (26695, J99, HPAG1, P12, G27, Shi 470, and B38) (13, 32, 36–39). The MCL processes built 2,293 ortholog clusters. The 1,602 CDs identified on B45 contigs were distributed over 1,182 clusters ubiquitous in all reference genomes (corresponding to 1,195 B45 CDs), 223 clusters were absent in at least one reference genome (nonubiquitous; corresponding to 232 B45 CDs), and, finally, 174 clusters (corresponding to 175 B45 CDs) had no counterpart in any of the 7 reference genomes.

The 175 CDs found in clusters unique to B45 were analyzed further by BLASTp searches against the seven *H. pylori* reference genomes, two *H. pylori* draft genomes available at that time (*H. pylori* strains 98-10 and B128 [40]), and *H. acinonychis* strain Sheeba (22) and against all other complete bacterial genomes in GenBank. Thirty-three of the CDs lacked significant BLASTp hits to any of the above genomes. Since these CDs were unusually short (mean length, 33 bp), we considered them artifacts. Fifteen CDs had significant BLASTp hits with Hp_B128 CDs (40), six had hits with Hp_98-10 (40), 27 had hits to *H. acinonychis* strain Sheeba (22), two had hits to other bacteria, and 92 had hits to one or more of the seven *H. pylori* reference genomes initially considered.

The B45 genome contains a complete *cag* pathogenicity island (*cag* PAI) comprised of 27 CDs divided over three contigs (C26, C50, and C53), as was expected considering its ability to promote interleukin-8 (IL-8) production and hummingbird phenotype on AGS cells (41) (see Table S1 in the supplemental material) (GenBank accession numbers AFAO01000023, AFAO01000045, and AFAO01000048, respectively).

The B45 genome has a prophage. Interestingly, among the 27 CDs in the orthologous clusters unique to B45 with significant BLASTp hits to *H. acinonychis* strain Sheeba, 22 matched genes in the *H. acinonychis* strain Sheeba prophage II (22). These CDs were originally found on two contigs: contig C47 (GenBank accession number AFAO01000043) and contig C21 (GenBank accession number AFAO01000020). The two contigs could be joined by PCR, using primers targeting the ends of C21 and C47 and Sanger sequencing of the resulting sequence fragment; hence, the B45 prophage constituted a continuous genomic region. A 1,665-bp region separated C21 from C47 (which corresponded exactly to a small contig C73 [AFAO01000062]). The reconstructed area (called C21B450028-C73B450001-C47B450001) contained a large CDS (5,958 bp), homologous to Hac_1615. The final prophage sequence constituted 24,645 bp with a G+C content of 37% and contained 27 CDs encoded on the same strand (GenBank accession number JF734911).

Nine of the 32 genes carried by the 28.4-kb *H. acinonychis* strain Sheeba prophage II lack orthologs in the B45 prophage, while five B45 genes lack orthologs in the Ha_Sheeba prophage (Fig. 1) (see also Table S2 in the supplemental material). Six of the most upstream B45 prophage genes were also found in a 5.5-kb region of the B38 genome, which probably corresponds to a remnant prophage (39). The average amino acid identity between B45 and Ha_Sheeba prophage genes was 61.0%, and that between B45 and B38 genes was 70.5%. The reconstructed B45 prophage was flanked on each side by partial gene sequences corresponding to jhp0786, which is a type I R-M enzyme subunit. The B45 phage has therefore probably been inserted into an R-M system.
The B45 prophage can be induced by UV irradiation. To assess whether the B45 prophage is functional, several assays were used to test for phage particle production. No significant lysis plaques of B45 were observed on agar plates after mitomycin treatment (0.1 to 1 \( \mu \text{g/ml} \); see Materials and Methods) or UV inductions (1 to 10 times the \( D \) value of UV irradiation; see Materials and Methods). Lysis plaques were observed using the JP1 strain, which served as a positive control (21). Liquid medium cultures did not reveal a significant decrease in optical density (OD) values (600 nm) at low concentrations (0.1 \( \mu \text{g/ml} \)) of mitomycin. High concentrations (1 \( \mu \text{g/ml} \)) could not be used since they were highly toxic to the cells (the negative-control strain 26695 displayed markedly lower OD values at this concentration). Finally, we tried to induce the B45 prophage in liquid medium by UV irradiation (1 \( D \) value after 24 h of incubation, followed by 24 additional hours of incubation; see Materials and Methods), which was followed by a precipitation step, as described in Materials and Methods. Transmission electron microscopy (TEM) using negative staining revealed numerous phage-like particles with an eggshell structure (size, 100 \( \pm \) 30 nm) (Fig. 2) apparently lacking a tail. Negative staining does not usually reveal phage tails because of the limited density of the tail (42). Fixation embedding and ultrathin sectioning of the bacteria allowed us to have access to the interior of the bacterial cells. A few cells contained 1 or 2 phage particles in their cytoplasm. These were recognizable by a high-density mature head, with rounded appearance, on which a tail was attached. The phage head was \( \approx 62.5 \text{ nm} \) \( \pm 7.3 \text{ nm} \) in diameter, and the tail was \( \approx 92.4 \text{ nm} \) \( \pm 2.97 \text{ nm} \) long and 5 to 6 nm in diameter. The total length of the phage was \( \approx 150 \text{ nm} \). Overall, these structures were compatible with a \textit{Siphoviridae} phage (Fig. 2).

Most phages replicate by first circularizing their linear DNA (43). We were unable to purify enough circular DNA to visualize it directly on an agarose gel, which may indicate that our induction conditions were suboptimal. However, observation of supernatant from UV-treated cultures allowed us to detect phage DNA by PCR amplification of the integrase gene after two rounds of exonuclease treatment. This indicates that circular phage DNA was present, protected from exonuclease cleavage. In contrast, the \( cagA \) gene could not be amplified after exonuclease treatment, indicating that all residual genomic DNA had been removed (Fig. 3).

Genome organization compatible with the \textit{Siphoviridae} phage family. Because the phage morphology observed using TEM was compatible with the \textit{Siphoviridae} family of phages, we compared the B45 prophage genome organization with this family of viruses. \textit{Siphoviridae} genomes fall into two size classes: 121 to 134 kb and 22 to 56 kb (44, 45). Hence, the B45 prophage (as well as \textit{H. acinonychis} strain Sheeba prophage II) had a size compatible with the second class. \textit{Siphoviridae} family phages usually harbor several typical modules such as head proteins, head-tail joining, DNA packaging, tail and tail fiber proteins, host lysis, lysogeny module, replication module, and transcriptional regulators (44). Most of the genes (23 of 27) did not generate significant BLAST matches to proteins other than those of the \textit{H. acinonychis} strain Sheeba and B38 prophage proteins, indicating that no closely related virus had been sequenced before (Table 1; see also Table S2 in the sup-
However, searches based on PSI-BLAST gave functional clues about a subset of the genes. Although sometimes highly putative, the results indicated that functionally related genes could be organized in several modules. These included a putative lysogeny module, harbouring at least a phage integrase/recombinase gene (C21B450018); 2 putative transcriptional regulators (C21B450020 and C47B450016); a replication module including one putative phage replication protein (C21B450021), a DNA repair protein (C21B450024 and C47B450012), a DNA primase (C21B450025), and a putative structural maintenance of chromosomes protein (SMC) (C21B450026); a putative tail fiber protein (C47B450002); and a putative crystalline beta/gamma motif-containing protein (putative lysin) (the 5,958-bp CDS which corresponds to the fusion of C21B450028-C73B450001-C47B450001) (Table 1).

Surprisingly, C47B450009 and C47B450010 had significant homologies with the transposable element called ISHp608 of H. pylori (46). ISHp608 is a member of the IS605 transposable element family. It contains two open reading frames (orfA and orfB), each related to putative transposase genes. Interestingly, orfB is also related to the Salmonella virulence gene gipA described in the lysogenic phage Gifsy-1 that affects Salmonella enterica Typhimurium survival in Peyer’s patches (47). C47B450009 had significant homology with orfA found in H. pylori Lith5 (AAL06574.1) (expect value/H11005 4e/H110087, identities/H11005 155/155 [100%]), and C47B450010 had significant homology with orfB found in H. pylori strain

---

**FIG 2** Transmission electron microscopy (TEM) images obtained after induction of the B45 prophage. (A) Examples of TEM images obtained using negative staining: numerous phage-like particles with an eggshell structure (size, 100 ± 30 nm). (B) (1) TEM images obtained after fixation embedding and ultrathin sectioning of the bacteria. Two phage particles on the top of the bacterial cell are visualized in the cytoplasm. The dark zone in the lower part of the cell corresponds to the condensed genomic DNA. On each side of the cell, two other bacterial cells are partially visible (one with genomic DNA). Some flagellar remnants can also be seen in between the cells. (2 and 3) Higher magnification focused on these 2 particles showed that they were recognizable by the high-density mature head, with a rounded appearance, to which a tail was attached. The phage head was ~62.5 nm (±7.3 nm) in diameter, and the tail was ~92.4 nm (±2.97 nm) long and 5 to 6 nm in diameter. The total length of the phage was ~150 nm.

**FIG 3** PCR detection of free circular phage DNA. To ensure complete elimination of bacterial genomic DNA, the DNA extracted from concentrated phage particles was treated twice with exonucleases (for 4 or 24 h) in order to digest any linear bacterial genomic DNA, leaving the circular DNA (i.e., phage DNA), which cannot be degraded by these enzymes. The extracted DNA was then tested for the presence of phage DNA and bacterial genomic DNA by PCR amplification of the phage integrase gene (using the primers F1, AAGYTTTTTAGMGTTTTGYG, and R1, CGCCCTGGCTTAGCATC, generating a 529-bp amplicon) and the cagA gene (750-bp amplicon) as already described (67, 70). Lane M corresponds to the 1-kb DNA ladder (Promega). Lanes 1 and 7, B45 extracted phage DNA plus exonucleases, 24 hours; lanes 2 and 8, B45 extracted phage DNA plus exonuclease buffer only, 24 hours; lanes 3 and 9, B45 extracted phage DNA plus exonuclease, 4 hours; lanes 4 and 10, B45 extracted phage DNA plus exonuclease buffer only, 4 hours; lanes 5 and 11, B45 DNA (positive control); lanes 6 and 12, H2O (negative control).
TABLE 1 Annotation of B45 prophage coding sequences

| B45 prophage CDS     | Hac     | HELPY | Putative annotation                     |
|----------------------|---------|-------|----------------------------------------|
| C21B450017           | Hac_1604| 1520  | Unknown function                        |
| C21B450018           | Hac_1606| 1521  | Bacteriophage-related integrase         |
| C21B450019           | Hac_1607| 1522  | Unknown function                        |
| C21B450020           | Absent  | 1523  | Putative transcriptional regulator     |
| C21B450021           | Hac_1609Hac_1608| 1525 | Putative phage replication protein     |
| C21B450022           | Absent  |       | Putative ABC superfamily ATP binding   |
| C21B450023           | Hac_1610| 1527  | Unknown function                        |
| C21B450024           | Hac_1611| Absent| Putative DNA repair protein (putative helicase) |
| C21B450025           | Hac_1612| Absent| Putative DNA primase                    |
| C21B450026           | Hac_1614| Absent| Putative SMC (structural maintenance of chromosomes) proteins |
| C21B450027           | Hac_1614| Absent| Unknown function                        |
| C21B450028-C73B450001-C47B450001 | Hac_1615| Absent| Putative crystalline beta/gamma motif-containing protein (putative lysis) |
| C47B450002           | Hac_1617| Absent| Putative tail fiber assembly protein    |
| C47B450003           | Hac_1618| Absent| Unknown function                        |
| C47B450004           | Hac_1619| Absent| Putative histidine kinase               |
| C47B450005           | Hac_1620| Absent| Putative histidine kinase               |
| C47B450006           | Hac_1621| Absent| Putative sensor protein                 |
| C47B450007           | Hac_1622| Absent| Unknown function                        |
| C47B450008           | Hac_1623| Absent| Unknown function                        |
| C47B450009           | Absent  |       | Transposase                            |
| C47B450010           | Absent  |       | Transposase                            |
| C47B450011           | Hac_1631| Absent| Unknown function                        |
| C47B450012           | Hac_1632| Absent| Putative DNA repair protein             |
| C47B450013           | Hac_1633| Absent| Unknown function                        |
| C47B450014           | Hac_1634| Absent| Phage tail tape measure protein         |
| C47B450015           | Hac_1635| Absent| Unknown function                        |
| C47B450016           | Absent  |       | Putative transcriptional regulator     |

*For each predicted protein, the putative annotations are provided for Helicobacter acinonychis strain Sheeba prophage II (Hac) and Helicobacter pylori B38 (HELPY) prophage homologs. C21B450028-C73B450001-C47B450001 corresponds to the fused CDSs obtained after PCR and sequencing (see Results).*

PeruCan2A (AF357223_2) (expect value = 0.0, identities = 379/382 [99%).

**Prevalence of the H. pylori prophage.** In order to estimate the prevalence of this prophage type in *H. pylori* genomes, we screened 341 strains, isolated in different geographic regions and from patients presenting different pathologies, using degenerate PCR primers targeting the B38, B45, and *H. acinonychis* strain Sheeba prophage II integrase genes. Surprisingly, 21.4% of the isolates (73/341) had an integrase gene (GenBank accession numbers JF734912 to JF734984), suggesting that this is a much more frequent phenomenon than initially estimated. The prevalences of the gene were similar in different pathologies: 25.6% (30/117) in duodenal ulcer strains, 20.9% (23/110) in gastritis strains, 15.9% (16/63) in gastric MALT lymphoma strains, and 19.6% (10/51) in atrophic gastritis/adenocarcinoma strains. Significant association was not found with cag PAI status. The average prevalence among European strains was 21.4% (59/276). Prevalence was higher in Sweden (18/56, 32.1%) than in France (28/125, 22%), Germany (6/53, 11.3%), and Portugal (3/29, 10.3%). No significant conclusion can be made concerning the United Kingdom and Norway, for which only 7 and 6 strains were tested, respectively. Notably, the prevalence among African isolates was low, with only one strain out of 18 (5.6%) being positive by PCR screening. This could be explained either by a lower prevalence of the prophage in African isolates or by significant integrase sequence divergence leading to mismatching of the PCR primers. The only South American strain included in the present study was negative. Thirteen of the 46 Asian strains (28.3%) included in the study were also positive.

**Phylogeography of the prophage integrase gene.** For each positive strain, amplicons were sequenced on both strands. Phylogenetic reconstruction of the phage integrase gene gave two main branches (Fig. 4). One of these branches included two well-supported clusters. Interestingly, the first one (cluster A), which included the B45 integrase, was comprised of 81.0% (17/21) French isolates and the second (cluster B) was composed of 70.6% (12/17) Swedish isolates. The *H. acinonychis* strain Sheeba prophage II integrase formed a sister group with these two clusters. The other main branch of the tree had two well-supported clusters. The first one was comprised exclusively of Asian strains (cluster C) and included 10 of the 13 positive Asian strains included in the study. The last cluster (cluster D), found at the extremity of the second branch, was more heterogeneous in geographic composition and was comprised of 6 French, 4 German, 3 Portuguese, and 3 Swedish strains. Overall, the phylogenetic reconstruction indicated a strong biogeographic signal within the phage integrase genetic diversity.

**DISCUSSION**

We have identified an integrated prophage in the genome of *H. pylori* strain B45 which is inducible by UV irradiation. This prophage is morphologically similar to phages of the *Siphoviridae* family. No prophage sequence has been reported for *H. pylori*,...
except for a remnant prophage sequence in the B38 strain; however, our study confirms two previously published articles reporting spontaneous production of small amounts of phage particles by an H. pylori strain (20, 48). Interestingly, all three reports describe the morphology as compatible with that of Siphoviridae; as with the phage described in the present paper, these previous articles reported infection, propagation, and electron microscopy images of a Helicobacter phage with a morphology compatible with that of Siphoviridae. The phage characteristics described by Heintschel von Heinegg et al. (i.e., phage heads of around 50 to 60 nm and DNA estimated to be 22,000 bp in length) are also in line with the B45 prophage (20).

It remains to be determined whether the capacity to produce phage is rare among H. pylori strains or whether a substantial proportion (approximately 21%) may be capable of producing phage. The methods used for phage induction in the present study are most likely suboptimal, since relatively few phage particles were observed and there were no visible plaques (Fig. 2). Other strategies may be more effective, such as exposing the bacteria to physical stress, leading to an induction of coccoidal forms, or by using a subinhibitory concentration of an antibiotic that induces DNA damage instead of chemical stress (49, 50). However, the optimal conditions for phage induction are difficult to predict. Nevertheless, it would be of interest to know whether randomly selected H. pylori strains are susceptible to infection by phage or whether most are resistant.

**FIG 4** Phylogenetic analysis of phage integrase sequences. Phylogenetic analysis was performed on the partial (442 bp) integrase DNA sequences obtained from the 73 integrase positive strains, and from the Hac_1606 genome. The evolutionary history of phage integrase genes was inferred using the neighbor-joining method using the Kimura two-parameter model (73). The bootstrapped consensus tree, inferred from 1,000 replicates, is presented as a radial tree. Bootstrap values (percentages of replicate trees in which the associated taxa clustered together) are shown for a selected nodes in the tree. The tree is drawn to scale, with branch lengths corresponding to the evolutionary distances used to infer the phylogenetic tree. Two main branches were identified. One of these branches included two well-supported clusters: cluster A, comprised of 81.0% French isolates, and cluster B, comprised of 70.6% Swedish isolates. The *H. acinonychis* strain Sheeba prophage II integrase (Hac) formed a sister group with these two clusters. The other main branch of the tree had two well-supported clusters: cluster C, comprised exclusively of Asian strains, and cluster D, comprised of 6 French, 4 German, 3 Portuguese, and 3 Swedish strains. Country of origin is indicated at the beginning of each strain designation: Fr, France; Sw, Sweden; De, Germany; Pt, Portugal; Uk, United Kingdom; JP, Japan; Ko, South Korea; T, Taiwan; Vn, Vietnam; Eg, Egypt. Disease status is indicated in the end: G, chronic gastritis; U, ulcer; M, MALT lymphoma; C, cancer (atrophic gastritis or gastric adenocarcinoma).
A transposable element homolog to ISHp608 (46) was found in the B45 prophage. Interestingly, one copy of ISHp609 has been described in the *H. pylori* B38 prophage. The presence of such insertion sequences (IS) in B38 has been viewed as a sign of a degenerative process (39). In prophages from Gram-negative bacteria, “extra genes” called “morons” can be found near the prophage DNA ends, at strategic positions, and these can interfere with the functions required during lysogeny and for lytic infection (45). However, in prophages some morons can be flanked by such transposase genes in their vicinity, pointing to alternative methods of mobility. Therefore, the presence of ISHp608 could be also viewed as a “poison” inside the phage DNA interfering with the inducibility of the phage, which may provide one putative explanation of the small amount of phage particles observed in the present study.

Only 23 of the 27 prophage genes displayed detectable protein sequence similarity with annotated genes in databases. This is most likely due to the high evolutionary rate of bacteriophages; identification of unknown sequences is expected when analyzing a novel phage genome (45). This could explain why some of the genes typically existing in a phage genome could not be identified, including a head-like and a tail fiber protein. The origin of this *Helicobacter* prophage could not be determined based on the data obtained in the present study. *Siphoviridae* is the best-documented phage family, with more than 60 complete phage genomes sequenced. They were described among a large panel of bacteria including Gram-positive bacteria (*Lactococcus, Lactobacillus, Streptococcus*, and *Staphylococcus*) and Gram-negative bacteria (in particular, *Enterobacteriaceae* members such as *Escherichia coli*) as well as *Mycobacterium* (44). Due to this large diversity, limited conclusions could be drawn regarding the *H. pylori* phage biology based on the observation that it resembles a *Siphoviridae* phage.

In order to assess the prevalence of prophages in *H. pylori*, we performed a PCR screen for the integrase gene and found this gene in a surprisingly high proportion of screened isolates (21.4%, 73/341). Although some *H. pylori* positive strains may have been missed due to primer mispairing, the prevalence was higher than that previously observed by Thibierge et al. (39), where DNA hybridization based on three prophage CDSs was carried out. In fact, due to the genetic variability of the prophage CDSs, including the integrase gene, hybridization results may be skewed. Altogether, this does suggest that prophages are relatively widespread among *H. pylori* strains; however, integrase-positive isolates may not carry complete prophages. To address this point, an extensive genome sequencing strategy based on *H. pylori* prophage-positive strains is under way.

Considering reported associations between prophages and virulence in other bacteria (e.g., *Streptococcus agalactiae* prophages in neonatal meningitis isolates [51]), we were expecting to find implications for *H. pylori* virulence. However, no significant association was found between the presence of the integrase gene and specific *H. pylori* gastroduodenal disease. This does not exclude the possible association of complete prophage sequences, or of specific prophage genes, with virulence. However, phylogenetic analysis of the detected phage integrase gene showed a pattern of biogeographic separation. This is in agreement with a model of coevolution between the virus and its bacterial host, since biogeographic separation is also observed within *H. pylori*. A model of geographically constrained viral dispersal also fits, i.e., *H. pylori* strains from different geographical regions may have been infected by distinct phage lineages after the geographic separation of the bacterial host. Another hypothesis could be that after bacterial infection with the virus, the divergence of the bacterium is also accompanied by divergence of the integrated virus. The fact that the genetic content and organization of the B45 prophage are similar to those of *H. acinonychis* strain Sheeba prophage II is of particular interest. In fact, it has been argued that *H. acinonychis* strain Sheeba is derived from *H. pylori* and that the prophage was acquired after the host jump from human to feline (22). Our study indicates that an equally likely scenario is that the prophage was present in the bacterial genome before the host jump.

Usually, phage DNA is integrated in the vicinity of a tRNA sequence. For *Helicobacter* species, prophages are integrated into protein-encoding genes. The B45 prophage is flanked on both sides by partial gene sequences corresponding to a type I R-M enzyme subunit. Indeed, R-M systems are often linked with mobile genetic elements, such as viruses. Examples of linkage between phages and restriction and R-M systems have already been described, i.e., HindIII on Phi-flue, Sau421 on phi-42, EcoO109I on P4-like prophage, and type I RM on a prophage annotated as a genomic island 5, among others (52, 53). This integration site is not conserved among *Helicobacter* prophages; the *H. acinonychis* strain Sheeba prophage II is integrated into a camphor resistance gene and carbamoylphosphate synthase small chain, while the B38 prophage is located between a hypothetical protein and cysteinrich protein G. The integration of the prophage in the B45 genome does not seem to create any significant genome rearrangement, because the numbering of the flanking gene on each site is conserved when taking into account all reference genomes. There is ample evidence for continued exchange of genetic material between phages, bacterial genomes, and various other genetic elements (54). This explains the sometimes fuzzy distinction between phages, plasmids, and pathogenicity islands (PAIs) and the chimeric nature of some phage genomes (55, 56). Nonetheless, several factors are typical of a PAI, like the cag PAI, which are not shared by the *H. pylori* prophage: (i) GC content clearly different from that of the host, (ii) hot spot for IS, (iii) integration within an essential gene (glutamate racemase), (iv) encoding of a secretion machinery specific for effectors (type IV secretion system for cag), and (v) effector molecules (CagA and peptidoglycan) translocated into the host cell by a contact-dependent secretion system (57). The large CDSs in the middle of the B45 prophage have significant similarities to Hac_1615 (Hac_1615) [Hac prophage II orf11 mosaic CUP0956/HP1116/jhp1044-like protein] and a JHP1044-like CDS in another part of the genome. Interestingly, JHP1044-like proteins have been described in an article aimed at identifying strain-specific genes located outside the plasticity zones of *H. pylori* genomes (58). An exciting hypothesis would be that the plasticity zones, which are increasingly being considered true PAIs (13, 59), have a phage origin. This kind of association between phage and PAI has indeed already been described in other bacteria such as *Staphylococcus aureus* (60, 61).

Interestingly, the Hac1615-like protein also shows similarity to a crystalline beta/gamma motif-containing protein, in particular with a protein corresponding to a putative lysin found in *Bordetella bronchiseptica* phage BPP-1 (62). The role of phage lysin is to make the bacterial membrane more permeable, thus allowing the entry of the phage DNA. Lysin targets the integrity of the cell wall and is designed to attack one of the major bonds in the peptidogly-
can. During the lytic cycle of phages, lysin also participates in bacterial lysis. With few exceptions, lysins do not have signal sequences, so they are not translocated across the cytoplasmic membrane. When lysins are used as antimicrobial agents, they are believed to work only with Gram-positive bacteria, where they have access to the cell wall carbohydrates and can attack the peptidoglycan, whereas the outer membrane of Gram-negative bacteria prevents this interaction. However, the presence of the B45 phage carrying a putative lysin suggests that it may be possible to identify novel lytic enzymes for targeting Gram-negative bacteria (63, 64). These could potentially be engineered to lyse cells and be used as antimicrobial agents against H. pylori, for example, and possibly delivered through liposomes (65).

The phage acquisition can play an important role in short-term adaptation processes for H. pylori. The high percentage of isolates carrying an integrase gene highlights the idea that prophages are underestimated in H. pylori. It is possible that they shape the genome at least in terms of the diversity of strains found worldwide as well as contributing to virulence evolution. The high percentage of isolates carrying the integrase gene of a prophage strongly suggests its importance in the diversification and fitness of the genome architecture in H. pylori. We previously published a paper in which we described the discovery of new sequences in the H. pylori genome (66). One of our conclusions was that these new genes could correspond to bacteriophage receptor/invasion proteins. This highly speculative conclusion is further supported by the present study. We believe that the description of the B45 prophage is an important contribution that will aid future genome analysis in H. pylori.

MATERIALS AND METHODS

H. pylori strain B45. The B45 strain was isolated from a 41-year-old male patient enrolled in a French multicenter study for low-grade gastric MALT lymphoma (67). B45 harbors a functional cag PAI (41), has a vacA s1m2 genotype (67), and belongs to the European multilocus sequence typing (MLST) cluster (based on data kindly provided by Sebastian Suerbaum, Hannover, Germany). This strain is susceptible to amoxicillin and clarithromycin, the two antibiotics given as the first line in eradication treatment.

Other H. pylori strains included in the present study. The other 340 H. pylori strains included in the present study came from several origins: 134 came from the collection of the French National Reference Center for Campylobacters and Helicobacters (F. Métraud and P. Lehours, Bordeaux, France); 78 from the Faculty of Engineering, Catholic University of Portugal (F. F. Vale); 54 from the Department of Microbiology, Tumor and Cell Biology, Karolinska Institute (Lars Engstrand); and 28 from the Klinikum Rechts Der Isar II, Medical Department, Technische Universität (M. Gerhard, Munich, Germany), and 46 Asian strains were provided by Y. Yamaoka (Department of Medicine-Gastroenterology, Michael E. Debakey Veterans Affairs Medical Center and Baylor College of Medicine, Houston, TX) and M. Oleastro (Department of Infectious Diseases, National Institute of Health, Lisbon, Portugal).

Genome sequencing. Genomic DNA from the B45 strain was extracted from cultures originating from a single colony, by using a commercial kit (Qiagen SA, Courtaboeuf, France) including a RNase step. The quality and concentration of the extracted DNA were verified using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Wilmington, DE) and also by running the extracted DNA on an agarose gel. Five micrograms of DNA was used for 454 library preparations according to the manufacturer’s protocol (April 2009) (454 Life Sciences, Branford, CT) and sequenced in one lane of a 4-region FLX Titanium plate on a Roche 454 FLX instrument. Data were processed with the accompanying 2.3 software package.

Genome assembly. The sequence generated included at least 26.4-fold coverage of Roche 454 Life Sciences FLX fragment data. The Roche 454 Life Sciences sequence was assembled using the Roche 454 Life Sciences Newbler Metrics assembler (software release 2.0.0.20). Only contigs longer than 500 bp were considered for data analysis.

Gene predictions and orthology analysis. We used the free access AMIGene (Annotation of Microbial Genes) pipeline available on the Genoscope website (http://www.genoscope.cns.fr/agc/tools/amiga/Form/form.php) (33), which is an application designed for automatically identifying the most likely coding sequences (CDSs) in a large contig or complete bacterial genome sequence. The contigs were numbered according to the Newbler Metrics assembler, and the CDSs identified among each contig were numbered according to AMIGene prediction: for example, C1B450001 corresponded to the first CDS identified in contig 1.

For defining H. pylori orthologous genes, the proteins of seven reference H. pylori genomes and of Hp_B45 were cross-compared using BLASTp (68) and clustered into orthologous groups using the Markov clustering (MCL) algorithm (34, 35). The complete reference H. pylori genomes available at the time of the analysis (Hp_26695, Hp_AG1, Hp_B38, Hp_G27, Hp_J99, Hp_P12, and Hp_Shi470) were downloaded from NCBI.

Before MCL, the BLAST output file was filtered such that (i) alignments with lower than min_identity sequence identity were removed and (ii) alignments with lower than min_length_ratio (alignment length)/(protein length of the longest of the two proteins) were removed. To find optimal values for min_identity and min_length_ratio, as well as for the granularity parameter I of the MCL algorithm, simulations were run using only the reference H. pylori proteomes. All combinations of different values for the three parameters were run (min_identity: 50, 60, . . . 90; min_length_ratio: 0.3, 0.4, . . . 0.9; I: 1, 2, . . . 5), and the combination rendering the largest number of clusters with exactly one protein per genome was selected. The rationale for this was that too-stringent criteria would likely split orthologous groups and produce groups lacking representatives in some genomes, while too-sloppy criteria would likely expand groups to also include paralogs and produce groups with multiple representatives in some genomes. The optimal conditions were found to be min_identity = 70, min_length_ratio = 0.5, and I = 1 (I had only minor effect compared to the other parameters). This parameter setting was then used for running the clustering procedure on all H. pylori proteomes, including Hp_B45.

H. pylori strains and growth conditions for phage induction. The strain 26695, whose genome does not have any integrated prophages (36), was selected as the negative control. As positive control, the selected strain was JP1, which produces lysis plaques after UV induction (21). H. pylori strains were grown on conventional in-house selective agar medium (Wilkins-Chalgren agar supplemented with 10% human blood and a mixture of antibiotics: vancomycin, trimethoprim, amphotericin B [Fungi zone], and cefsulodin) or liquid medium (brucella broth supplemented with 10% Polyvitex [bioMérieux, Marcy l’Etoile, France] and 10% filtered fetal bovine serum [Gibco Invitrogen, Cergy Pontoise, France]) under microaerophilic conditions at 37°C.

For phage induction, both chemical (mitomycin C from Streptomyces caespi tus; Sigma-Aldrich, St. Louis, MO) and physical (UV irradiation at 254 nm) agents were used. For mitomycin C induction, strains were resuspended in brucella broth and mitomycin C was added to final concentrations of 1 µg/mL, 0.5 µg/mL, and 0.1 µg/mL. Twenty µL of this bacterial suspension were applied on H. pylori selective medium, dried, incubated under standard conditions, and observed for lytic plaque formation. Mitomycin C induction was also performed in liquid medium at the same concentrations, and the absorbance at 600 nm was checked at regular intervals. Controls without mitomycin C were used.

The induction with 254-nm UV irradiation (Bioblock Scientific VL 6C/6W UV lamp, 254 nm; power, 12 W; or Bulbworks BW.G8T5 UV lamp, 8 W, 253.7-nm UVC) was performed by exposing inoculated (20-liter spots) H. pylori selective medium plates to different periods of UV
irradiation at a distance of 13 cm from the UV source. For each UV lamp used, the D value (decimal reducing time, which is the time necessary to reduce 90% of the bacterial population) was determined for *Escherichia coli* K-12. The D value was used for exposure of the *H. pylori* cultures to 0, 1, 5, and 10 times the D value of UV irradiation. The plates were incubated and checked for lytic plaques. The induction of *H. pylori* in liquid medium was performed by exposing 500 ml of brucella broth culture to 1 D value after 24 h of incubation, followed by 24 additional hours of incubation. Then, after phage precipitation, the broth was used for transmission electron microscopy (TEM) analysis after phage precipitation and phage DNA extraction (see below).

**Phage particle concentration.** Phage particles from *H. pylori* strain B45 were concentrated using a protocol adapted from that of Henn et al. (69). Briefly, a 48-h culture of *H. pylori* grown in liquid medium was centrifuged to pellet cells (4,000 rpm, 10 minutes), which were discarded. The supernatant was gently decanted to a new tube. Four milliliters of phage precipitant (33% polyethylene glycol [PEG], 3 M NaCl) was added to the supernatant, followed by incubation overnight at 4°C. To pellet phage particles, a centrifugation step (10,000 rpm, 10 min at 4°C) was applied. The supernatant was gently discarded, and residual liquid was drained using a paper towel. The phage pellet was resuspended in 500 ml of distilled water. To ensure complete elimination of bacterial genomic DNA, the minipreparation was treated sequentially with two exonucleases and resuspended in the same volume of water to eliminate the enzymes (after heat inactivation and DNA precipitation). The enzymes used were (i) exonuclease I (*E. coli*) (New England Biolabs, Ipswich, MA) and (ii) lambda exonuclease (New England Biolabs) sequentially. The goal was to cut any genomic remnant DNA trace, whereas putative circular closed DNA (i.e., phage DNA) could not be degraded by any of these enzymes. The extracted DNA was then tested for the presence of phage DNA and bacterial genomic DNA by PCR amplification of the phage integrase gene (see below) and the cagA gene as already described (67, 70).

**Transmission electron microscopy.** Two kinds of electron microscopy analyses were performed: negative staining or fixation embedding and sectioning. For negative staining, Formvar membrane-covered 300-mesh copper grids were floated for 1 to 2 minutes on a drop of the virus suspension for particle adsorption to the membrane. The grids were then air dried. The resulting samples were observed using a JEOL 100SX electron microscope. For fixation embedding and sectioning, bacteria in suspension were prefixed in 0.2 mol/liter cacodylate buffer (pH 6.8) containing 5% glutaraldehyde (50/50 mixture with the culture medium) for 2 h at 4°C. After centrifugation at 8,000 rpm for 3 minutes, bacteria were postfixed in 1% osmium tetroxide buffered in 0.1 mol/liter of cacodylate (pH 6.8) for 1 h at room temperature in the dark. Pellets were included in 1% agar and cut in small pieces (2 mm³), and specimens were dehydrated in an ethanol series (50%, 70%, 95%, 100%). Dehydration was completed in propylene oxide, and the specimens were embedded in epoxy resin (Epon 812; EMS, Hatfield, PA). The resin was polymerized at 60°C for 48 h. The samples were sectioned using a diamond knife on an ultramicrotome (Ultracut-E; Leica Microsystems, Nanterre, France). Thin sections (70 nm) were picked up on copper grids and then stained with uranyl acetate and lead citrate. The grids were examined with a transmission electron microscope at 120 kV (H7650; Hitachi, Tokyo, Japan).

**Determination of the prevalence of the prophage integrase gene in a collection of *H. pylori* isolates.** To identify the prevalence of prophage sequences in *H. pylori* strains, a PCR screening strategy was used with degenerated primers F1, AAGYTTTTAGMGTTTTGYG, and R1, CGC CCGGCTTACGATC, designed using Primer 3 (http://frodo.wi.mit.edu/primer3/) on the aligned integrase sequences of B38 (HELGY_1521), B45, and *H. acnonychis* strain Sheba (Hac_1606) obtained using multiple sequence alignment with hierarchical clustering (71). These primers generated a 529-bp PCR product. A total of 341 *H. pylori* strains were screened (117 from duodenal ulcer, 110 from gastritis, 63 from gastric MALT lymphoma patients [including strain B45], and 51 from atrophic gastritis or gastric adenocarcinoma patients). Most of these strains came from Europe (France, 125; Sweden, 56; Germany, 53; Portugal, 29; United Kingdom, 7; Norway, 6), 46 strains came from Asia (Japan, 26; Thailand, 3; Vietnam, 7; Taiwan, 4; South Korea, 6), 18 strains came from Africa (Egypt, 7; Burkina Faso, 11), and 1 strain came from South America (Costa Rica). For each positive strain, amplicons were purified using MicroSpin S-400 HR columns (GE Healthcare, Saclay, France) and directly sequenced on both strands as previously described (67).

**Phylogeny analysis of phage integrase sequences.** Phylogeny analysis was performed on the integrase sequences obtained on the 73 positive strains screened in the present study: a 442-bp sequence was available for each sequenced PCR product. We also included the internal fragment of the Hac_1606 integrase sequence (YP_665314.1) corresponding to the amplified product.

The evolutionary history of phage integrase genes was inferred using the neighbor-joining method. Neighbor-joining phylogenetic tree topologies of nucleotide alignments were constructed using the MEGA (Molecular Evolutionary Genetics Analysis) 3.1 software (72), on the basis of distances estimated using the Kimura two-parameter model (73). This model corrects for multiple hits, taking into account transitional and transversal substitution rates. Branching significance was estimated using bootstrap confidence levels by randomly resampling the data 1,000 times with the referred evolutionary distance model.

**Nucleotide sequence accession numbers.** GenBank accession numbers for the 63 contigs obtained after genome assembly are AFAO01000001 to AFAO01000063. The GenBank accession number for the final prophage sequence is JF734911. The corresponding GenBank accession numbers for prophage integrase gene sequences are JF734911 to JF734984.

**ACKNOWLEDGMENTS**

We thank Alain Dublanchet, Sylvain Moineau, and Sven Löffhal for fruitful discussions.

**SUPPLEMENTAL MATERIAL**

Supplemental material for this article may be found at http://mbio.asm.orglookup/suppl/doi:10.1128/mBio.00239-11/-/DCSupplemental.

Table S1, XLS file, 0.1 MB.
Table S2, DOC file, 0.1 MB.

**REFERENCES**

1. Suttle CA. 2005. Viruses in the sea. Nature 437:356–361.
2. Labrie SJ, Samson JE, Moineau S. 2010. Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. 8:317–327.
3. Tyson GW, Banfield JF. 2008. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ. Microbiol. 10:200–207.
4. Andersson AF, Banfield JF. 2008. Virus population dynamics and acquired virus resistance in natural microbial communities. Science 320:1047–1050.
5. Rodriguez-Valera F, et al. 2009. Explaining microbial population genomics through phage predation. Nat. Rev. Microbiol. 7:828–836.
6. Daubin V, Ochman H. 2004. Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. Genome Res. 14:1036–1042.
7. Falush D, et al. 2003. Traces of human migrations in *Helicobacter pylori* populations. Science 299:1582–1585.
isolates of the human gastric pathogen _Helicobacter pylori_. Nature 397: 176–180.
33. Bocs S, Cruveiller S, Vallenet D, Nuel G, Médićuge C. 2003. AMIGene: Annotation of Microbial Genes. Nucleic Acids Res. 31:3723–3726.
34. Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30: 1575–1584.
35. van Dongen S. 2000. Graph clustering by flow simulation. Ph.D. Thesis. University of Utrecht, Utrecht, Netherlands.
36. Tomb JF, et al. 1997. The complete genome sequence of the gastric pathogen _Helicobacter pylori_. Nature 388:539–547.
37. Oh JD, et al. 2006. The complete genome sequence of a chronic atrophic gastritis _Helicobacter pylori_ colonization strain: evolution during disease progression. Proc. Natl. Acad. Sci. U. S. A. 103:9999–10004.
38. Baltrus DA, et al. 2009. The complete genome sequence of _Helicobacter pylori_ strain G27. J. Bacteriol. 191:447–448.
39. Thibeorge JM, et al. 2010. From array-based hybridization of _Helicobacter pylori_ isolates to the complete genome sequence of an isolate associated with MALT lymphoma. BMC Genomics 11:368.
40. McClain MS, Shaffer CL, Israel DA, Peek RM, Jr, Cover TL. 2009. Genome sequence analysis of _Helicobacter pylori_ strains associated with gastric ulceration and gastric cancer. BMC Genomics 10:3.
41. Ferreira-Chagas B, et al. 2007. _In vitro_ proinflammatory properties of _Helicobacter pylori_ strains causing low-grade gastric MALT lymphoma. _Helicobacter_ 12:616–617.
42. Ackermann HW. 1999. Basic phage electron microscopy. Methods Mol. Biol. (Clifton, NJ) 50:113–126.
43. Weigel C, Seitz H. 2006. Bacteriophage replication modules. FEMS Microbiol. Rev. 30:321–381.
44. Brüssow H, Desiere F. 2001. Comparative phage genomics and the evolution of _Siphoviridae_: insights from dairy phages. Mol. Microbiol. 39: 213–222.
45. Brüssow H, Canchaya C, Hardt WD. 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol. Mol. Biol. Rev. 68:560–602.
46. Kersulyte D, et al. 2002. Transposable element ISHP608 of _Helicobacter pylori:_ nonrandom geographic distribution, functional organization, and insertion specificity. J. Bacteriol. 184:992–1002.
47. Stanley TL, Ellermeier CD, Slauh CE. 2000. Tissue-specific gene expression identifies a gene in the _lysogetic_ phage _Gifsy-1_ that affects _Salmonella enterica_ serovar Typhimurium survival in _Peyer’s_ patches. J. Bacteriol. 182:4406–4413.
48. Schmid EN, von Recklinghausen G, Ansorg R. 1990. Bacteriophages of _Helicobacter pylori_ (Campylobacter) pylori. J. Med. Microbiol. 32:101–104.
49. Azvedo NF, et al. 2007. Cocccoid form of _Helicobacter pylori_ as a morphological manifestation of cell adaptation to the environment. Appl. Environ. Microbiol. 73:3423–3427.
50. McDonald JE, Smith DL, Fogg PC, McCarthy AI, Allison HE. 2010. High-throughput method for rapid induction of prophages from _bacteria_ and its application in the study of _Shiga_ toxin-encoding _Esherichia coli_ strains. Appl. Environ. Microbiol. 76:2360–2365.
51. van der Mee-Marquet N, et al. 2006. Prophagic DNA fragments in _Streptococcus agalactiae_ strains and association with neonatal meningitis. J. Clin. Microbiol. 44:1049–1058.
52. Holden MT, et al. 2004. Genomic plasticity of the causative agent of _meliodiosis_, _Burkholderia pseudomallei_. Proc. Natl. Acad. Sci. U. S. A. 101:14240–14245.
53. Kobayashi I. 2001. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. Nucleic Acids Res. 29:3742–3756.
54. Hatfull GF. 2008. Bacteriophage genomics. Curr. Opin. Microbiol. 11: 447–453.
55. Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world’s a phage. Proc. Natl. Acad. Sci. U. S. A. 96:2192–2197.
56. Novick RP, Christie GE, Penades JR. 2010. The phage-related chromosomal islands of Gram-positive bacteria. Nat. Rev. Microbiol. 8:541–551.
57. Censini S, et al. 1996. cag, a pathogenicity island of _Helicobacter pylori_, encodes type I-specific and disease-associated virulence factors. Proc. Natl. Acad. Sci. U. S. A. 93:14648–14653.
58. Chanto G, et al. 2002. Identification of strain-specific genes located outside the plasticity zone in nine clinical isolates of _Helicobacter pylori_. Microbiology (Reading, England) 148:3671–3680.
59. Kersulyte D, et al. 2009. Helicobacter pylori’s plasticity zones are novel transposable elements. PLoS One 4:e6859.

60. Tormo MA, et al. 2008. Staphylococcus aureus pathogenicity island DNA is packaged in particles composed of phage proteins. J. Bacteriol. 190: 2434–2440.

61. Tao L, Wu X, Sun B. 2010. Alternative sigma factor sigmaH modulates prophage integration and excision in Staphylococcus aureus. PLoS Pathog. 6:e1000888.

62. Liu M, et al. 2004. Genomic and genetic analysis of Bordetella bacteriophages encoding reverse transcriptase-mediated tropism-switching cas-settes. J. Bacteriol. 186:1503–1517.

63. O’Flaherty S, Rosi RP, Coffey A. 2009. Bacteriophage and their lysins for elimination of infectious bacteria. FEMS Microbiol. Rev. 33:801–819.

64. Fischetti VA. 2010. Bacteriophage endolysins: a novel anti-infective to control Gram-positive pathogens. Int. J. Med. Microbiol. 300:357–362.

65. Torchilin VP. 2006. Recent approaches to intracellular delivery of drugs and DNA and organelle targeting. Annu. Rev. Biomed. Eng. 8:343–375.

66. Menard A, Danchin A, Dupouy S, Megraud F, Lehours P. 2008. A variable gene in a conserved region of the Helicobacter pylori genome: isotopic gene replacement or rapid evolution? DNA Res. 15:163–168.

67. Lehours P, 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.

68. Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389–3402.

69. Henn MR, et al. 2010. Analysis of high-throughput sequencing and annotation strategies for phage genomes. PLoS One 5:e9083.

70. Lehours P, Zheng Z, Skoglund A, Megraud F, Engstrand L. 2009. Is there a link between the lipopolysaccharide of Helicobacter pylori gastric MALT lymphoma associated strains and lymphoma pathogenesis? PLoS One 4:e7297.

71. Corpet F. 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. 16:10881–10890.

72. Kumar S, Tamura K, Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. 5:150–163.

73. 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.