Comparative Whole Genome Sequence Analysis of the Carcinogenic Bacterial Model Pathogen *Helicobacter felis*

Isabelle C. Arnold1, Zuzana Zigova1, Matthew Holden2, Trevor D. Lawley2, Roland Rad2, Gordon Dougan2, Stanley Falkow3, Stephen D. Bentley2, and Anne Müller*,1

1Institute of Molecular Cancer Research, University of Zürich, Zürich, Switzerland
2Pathogen Genomics, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom
3Department of Microbiology and Immunology, Stanford University School of Medicine

*Corresponding author: E-mail: mueller@imcr.uzh.ch.
Accepted: 8 March 2011
GenBank accession number FQ670179

Abstract

The gram-negative bacterium *Helicobacter felis* naturally colonizes the gastric mucosa of dogs and cats. Due to its ability to persistently infect laboratory mice, *H. felis* has been used extensively to experimentally model gastric disorders induced in humans by *H. pylori*. We determined the 1.67 Mb genome sequence of *H. felis* using combined Solexa and 454 pyrosequencing, annotated the genome, and compared it with multiple previously published *Helicobacter* genomes. About 1,063 (63.6%) of the 1,671 genes identified in the *H. felis* genome have orthologues in *H. pylori*, its closest relative among the fully sequenced *Helicobacter* species. Many *H. pylori* virulence factors are shared by *H. felis*: these include the gammaglutamyl transpeptidase GGT, the immunomodulator NapA, and the secreted enzymes collagenase and HtrA. *Helicobacter felis* lacks a Cag pathogenicity island and the vacuolating cytotoxin VacA but possesses a complete comB system conferring natural competence. Remarkable features of the *H. felis* genome include its paucity of transcriptional regulators and an extraordinary abundance of chemotaxis sensors and restriction/modification systems. *Helicobacter felis* possesses an episomally replicating 6.7-kb plasmid and harbors three chromosomal regions with deviating GC content. These putative horizontally acquired regions show homology and synteny with the recently isolated *H. pylori* plasmid pHPPC4 and homology to *Campylobacter* bacteriophage genes (transposases, structural, and lytic genes), respectively. In summary, the *H. felis* genome harbors a variety of putative mobile elements that are unique among *Helicobacter* species and may contribute to this pathogen's carcinogenic properties.

Key words: *Helicobacter felis*, comparative genomics, genome sequence analysis, horizontal gene transfer, mobile elements, bacteriophage.

*Helicobacter felis* is a close relative of the human gastric bacterial pathogen *H. pylori*, the causative agent of gastritis and gastric ulcers (Marshall and Warren 1984) and known risk factor for gastric adenocarcinoma and gastric lymphoma (Parsonnet et al. 1991, 1994). *Helicobacter felis* was originally isolated from cats (Lee et al. 1990). Its ability to persistently colonize laboratory mice was discovered in 1990 (Lee et al. 1990) and was subsequently exploited to generate convenient and highly reproducible mouse models of *Helicobacter*-induced chronic active gastritis (Lee et al. 1990, 1993), gastric atrophy (Fox et al. 2000), lymphoma (Enno et al. 1995, 1998), and adenocarcinoma (Fox et al. 2002). It remains the most broadly used strain for modeling gastric pathology associated with virulent *Helicobacter* infection to date (Houghton et al. 2004; Sayi et al. 2009; Craig et al. 2010a, 2010b; Toller, Altmeyer, et al. 2010; Toller, Hitzler, et al. 2010). Whereas fully annotated whole genome sequence information is now available for seven *H. pylori* isolates (Tomb et al. 1997; Alm et al. 1999; Oh et al. 2006; Giannakis et al. 2008; Baltrus et al. 2009; Farnbacher et al. 2010; Fischer et al. 2010), one strain each of *H. hepaticus* (Suerbaum et al. 2003), *H. mustelae* (O’Toole et al. 2010a, 2010b), and *H. acinonychis* (Eppinger et al. 2006) and for numerous other related species of the Campylobacterales family (Parkhill et al. 2000; Baar et al. 2003), the *H. felis* genome had not been sequenced to date. Due to its carcinogenic
properties and complement of virulence factors that are suspected to differ from those of *H. pylori*, the genome sequence information of *H. felis* is expected to generate novel insights into the molecular pathogenesis of this model organism.

The genome sequence of *H. felis* CS1 (ATCC 49179) was determined at the Wellcome Trust Sanger Institute using a combination of 454 pyrosequencing (which generated a total number of 345,320 reads, 47 contigs, 54× genome coverage) and Solexa sequencing (28.4 million reads, 900× genome coverage). The average read lengths were 388 bp (454) and 54 bp (Solexa). The combined 454 and Solexa sequencing data were joined to generate a draft sequence based on the assembly scaffold information; gaps between contigs were closed by a single round of polymerase chain reaction-based finishing/gap closure. The resulting final *H. felis* genome assembly revealed a 1,672,681 bp genome that includes a 6,700 bp epismally replicating plasmid. The genome of *H. felis* is comparable in size to the sequenced *H. pylori* genomes (1.59–1.67 Mb) is slightly larger than the *H. mustelae* and *H. acinonychis* genomes (1.58 and 1.55 Mb) but smaller than the *H. hepaticus* genome (1.8 Mb). Its GC content of 44.51% is higher than that determined for *H. pylori* (39%) but well within the range of Campylobacter genera (30.6–48.5%). Coding sequence predictions made by Orpheus, Glimmer2, and EasyGene software identified 1,671 genes and 1 pseudogene in a coding area of 92%, placing *H. felis* squarely in the range determined for other *Helicobacter* species (1,403–1,875 genes) and Campylobacter (1,403–2,046 genes). The average length of an *H. felis* gene is 921 bp (0.998 genes per kb). We detected 35 tRNA genes using tRNAscan-SE (Lowe and Eddy 1997). The existence of a previously identified *H. felis* plasmid could be confirmed (see below) (De Ungria et al. 1998). The *H. felis* genome sequence was annotated using Artemis software (Rutherford et al. 2000) version 12 and visualized with DNA Plotter; protein domains and patterns were marked up using Pfam and Prosite (Bateman et al. 2004). Annotation was transferred from the previously annotated *H. pylori* P12 and *H. mustelae* genomes (Fischer et al. 2010; O’Toole et al. 2010) to orthologous genes and then manually curated using FASTA and Blast results. Orthologous proteins were identified as reciprocal best matches using FASTA; all-against-all FASTA searches were performed, and reciprocal best matches were defined if the top hit covered at least 80% of the length of both sequences with at least 60% identity for both proteins. Of the 1,671 genes, 83% could be annotated with high confidence using these criteria. The fully annotated genome is publicly available at GenBank (accession number FQ670179). All genes were assigned to COG categories using the integrated microbial (IM)B genomes tool (Markowitz et al. 2009); annotated genes are color coded according to COG category (fig. 1).

**Helicobacter felis** Factors Involved in Colonization, Motility, Chemotaxis, Virulence, and Natural Competence

*Helicobacter felis* shares numerous features with other *Helicobacter* species that facilitate colonization of the gastric acidic environment and are required for motility and chemotaxis (fig. 1 and table 1). The *H. felis* genome harbors a complete urease gene cluster (ureABIEFGH), which in *H. pylori* is essential for gastric colonization (Karita et al. 1995) (fig. 1). In line with the requirement for nickel as a cofactor of urease, the orthologue of the *H. pylori* nickel transporter NixA and another predicted high-affinity nickel transport protein are encoded directly downstream of the urease gene cluster. An additional ureA,B2 operon is present in *H. felis*, as was reported for *H. mustelae* (O’Toole et al. 2010), where UreA2 and UreB2 are known to be expressed under conditions of nickel limitation (Stoof et al. 2008). The *H. felis* genome harbors at least 40 motility/chemotaxis-related genes encoded by the fla, flg, fil, and fli gene families (table 1), which in *H. pylori* and other flagellated bacteria are involved in the regulation, secretion, and assembly of the flagellum (O’Toole et al. 2000). *Helicobacter felis* further possesses an extraordinary number of chemotaxis genes. At least 20 predicted methyl-accepting inner membrane chemotaxis proteins (MCPs) are present in the *H. felis* genome, many of which share homology with the *Bacillus subtilis* MCPs TlpA (present in 5 copies in the *H. felis* genome), Tlp8 (6 copies), and TlpC (3 copies) and tend to be clustered together on the chromosome (fig. 1). In other organisms, the binding of chemotactants such as urea, bicarbonate, or amino acids to MCPs is transduced to the autophosphorylating kinase CheA via CheW. CheA donates a phosphoryl group from a histidine residue to an aspartate of CheY, which then interacts with switching proteins to change the direction of flagellar motor rotation. cheA, W, and Y orthologues are each present in one copy in the *H. felis* genome. Three additional chemotaxis genes share homology with the *B. subtilis* CheV protein (termed CheV, V1, and V2 in *H. felis*). Like *H. pylori*, *H. felis* lacks cheB and cheR, which in other bacteria are responsible for modulating the chemotactic response by addition and removal of methyl groups to/from MCPs. The abundance of predicted MCP-like chemotaxis sensors in *H. felis*, especially in comparison to other Campylobacterales (*H. pylori*: 4, *H. hepaticus*: 9, *C. jejuni*: 10), is striking and suggests an elaborate spatial orientation in a diverse habitat.

Apart from a type III secretion system that exports the flagellar subunit components across both membranes (encoded in *H. felis* by predicted orthologues of the *H. pylori* genes flaA,B and flaH,I,P,Q,R), the *H. felis* genome harbors only one additional secretion system. The comB regulon encoding the type IV secretion components comB2,3,4,6,8,9,10 is required for natural competence of *H. pylori* (Hofreuter et al. 1998, 2003; Karnholz et al.
2006), and its orthologues are expected to encode the identical function in H. felis. Whereas the comB components of H. pylori are organized in only two operons encoding comB2-4 and comB6-10, respectively, the H. felis orthologues are dispersed across the genome in three operons (fig. 1). A second H. pylori-specific type IV secretion system, encoded by the Cag pathogenicity island, is clearly absent in the H. felis genome, as is the vacuolating cytotoxin VacA (Montecucco et al. 2001). Other virulence-associated genes of H. pylori, in contrast, are present in H. felis and typically share a high degree of similarity. An H. pylori virulence factor involved in immunomodulation, NapA (Satin et al. 2000), is present in H. felis, as are three enzymes recently implicated in H. pylori virulence—a collagenase, the secreted serine protease HtrA and the gamma-glutamyl transpeptidase GGT (Gong et al. 2010; Hoy et al. 2010). The cytolethal distending toxin shared by H. hepaticus and C. jejuni lacks an orthologue in H. felis. The H. felis genome encodes a total of 52 outer membrane proteins belonging to the Hor, Hop, Hof, and Hom gene families. Orthologues of the Sab and Bab adhesins could not be identified in the H. felis genome.

All Helicobacter genomes sequenced so far share several features that set the Helicobacter genus apart from other enteropathogenic bacteria; one striking characteristic is the scarcity of transcriptional regulators. We could identify only three sigma factors, \( \sigma^{54}/rpoN \), \( \sigma^{70}/rpoD \), and \( \sigma^{28}/FliA \). FliA, as well as the anti-sigma factor FlgM of H. felis contribute to the regulation of the fla/fli/flg/flh motility regulon. The CstA regulator of the “stringent response” to carbon starvation is present in the H. felis genome. Two-component systems consisting of a membrane histidine kinase sensor protein and a cytoplasmic DNA-binding response regulator are represented in the H. felis genome by only two sensors and two response regulators (in addition to the aforementioned CheA/Y system regulating chemotaxis), which all share high homology with their orthologues in other Helicobacter genomes. Additional transcriptional regulators identified in the H. felis genome include the ferric uptake regulator (Fur), a main regulator of iron acquisition, the nickel-responsive repressor NikR and the carbon storage regulator CsrA. Like other bacterial pathogens colonizing environments limited in ferric and ferrous iron, H. felis possesses a number of iron uptake and storage systems. In addition to Fur, the H. felis genome encodes orthologues of the Escherichia coli Fec and Feo siderophore-mediated iron uptake systems; multiple copies of frp genes encoding

**FIG. 1.**—Circular genome atlas of Helicobacter felis CS1. Rings from outside to inside: 1, Selected H. felis orthologues of H. pylori factors associated with virulence, colonization, natural competence, and chemotaxis. 2, Nucleotide coordinates in bp. 3, open reading frame (ORF) distribution, plus strand. 4, ORF distribution, negative strand. ORFs are color coded based on COG classifications. Abbreviations: B2-10, comB2-10; coll., collagenase.
### Table 1

<table>
<thead>
<tr>
<th>Trait</th>
<th><em>Helicobacter pylori</em> Locus</th>
<th>Orthologous <em>H. felis</em> System</th>
<th>Role in Colonization/Virulence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type IV secretion</td>
<td>Cag PAI</td>
<td>Absent</td>
<td>Severe inflammation and secretion elevated gastric cancer risk</td>
</tr>
<tr>
<td>Urease production</td>
<td>ComB</td>
<td>ComB2,3,4,6,8,9,10 present</td>
<td>Required for natural competence</td>
</tr>
<tr>
<td></td>
<td>Urease gene cluster</td>
<td>UreABEFGH present</td>
<td>Required for acid resistance and gastric colonization</td>
</tr>
<tr>
<td>Vacuolating cytotoxin</td>
<td>VacA</td>
<td>Absent</td>
<td>Vacuolization, cytotoxicity, T-cell inhibition</td>
</tr>
<tr>
<td>Neutrophil activation</td>
<td>NapA</td>
<td>NapA present</td>
<td>Neutrophil activation, Th1 polarization</td>
</tr>
<tr>
<td>Gamma-glutamy transpeptidase</td>
<td>GGT</td>
<td>GGT present</td>
<td>Oxidative DNA damage, colonization</td>
</tr>
<tr>
<td>Outer membrane proteins</td>
<td>Hop, Hor, Hof, Horn</td>
<td>Hop, Hor, Hof, Hom</td>
<td>Binding to various glycosylated host cell surface proteins</td>
</tr>
<tr>
<td>Motility</td>
<td>Fla, Flg, Flh, Fli</td>
<td>FlaABG; FlgBCEE2GG2H2IKLM; FlADEFGHILMPQRSTWW2Y; FlhABF</td>
<td>Regulation, assembly, and function of flagella</td>
</tr>
<tr>
<td>Secreted serine protease</td>
<td>HtrA</td>
<td>Present</td>
<td>E-cadherin cleavage; access to intercellular space</td>
</tr>
<tr>
<td>Cytolethal distending toxin</td>
<td>Absent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sigma factors</td>
<td>σ^54, σ^70, σ^28</td>
<td>σ^54, σ^70, σ^28fliA</td>
<td>Very few σ factors present</td>
</tr>
<tr>
<td>DNA repair recombination</td>
<td>RecA, AddA,B</td>
<td>RecA,N present; AddB present RecB,C,D,G absent; mutS present</td>
<td>Required for gastric colonization</td>
</tr>
<tr>
<td>Collagenase</td>
<td>Present</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iron uptake</td>
<td>Feo, FeC, Frn, Fur, TonB, ExbB, ExbD</td>
<td>FeoA,B, FeC,A, 2xFrpB, 2xTonB, ferritin, SodB, Fur, ExbB2,D,D2</td>
<td>Required for gastric colonization</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Iron uptake likely critical for early colonization and persistence</td>
</tr>
</tbody>
</table>

heme- or lactoferrin-binding proteins are present. The non-heme iron storage protein ferritin is also encoded in the *H. felis* genome. Overall, the iron uptake systems of *H. felis* are highly conserved within the *Helicobacter* genus (Tomb et al. 1997; Suerbaum et al. 2003), highlighting the evolutionary restrictions created by iron limitation in mammalian hosts.

### Comparative Analysis of Sequenced and Annotated Helicobacter Genomes

Of the 1,671 genes identified in the *H. felis* genome, 1,033 are predicted to have orthologues in *H. pylori* P12 (61.8%), 1,048 have orthologues in *H. pylori* B8 (62.7%), 1,029 have orthologues in *H. acinonychis* (61.6%), 946 genes have orthologues in *H. mustelae* (56.6%), and 940 have orthologues in *H. hepaticus* (56.2%). A circular plot comparing the *H. felis* genome with these five *Helicobacter* genomes is shown in figure 2 along with GC content and GC skew maps. The genes that are *H. felis*-specific, that is, not shared with these representatives of the other sequenced *Helicobacter* species (purple ring, fig. 2A), are enriched for very few gene classes. Genes encoding restriction/modification systems are particularly overrepresented among *H. felis*-specific genes. The *H. felis* genome encodes eight complete restriction/modification systems as well as methyltransferases for which no matching restriction endonuclease could be identified; several of the complete systems and most of the adenine methyltransferases lack orthologues in the other *Helicobacter* genomes (annotated in red, fig. 2A). The number of restriction/modification systems is thus quite variable across the *Helicobacter* genus, with *H. pylori* encoding at least 11 and *H. hepaticus* encoding only 2 complete systems (Tomb et al. 1997; Suerbaum et al. 2003).

Other classes of *H. felis*-specific genes include additional copies of the chemotaxis sensors *tipA, B*, and *C* mentioned earlier and transposases encoded by putative insertion elements (annotated in fig. 2A). As indicated above, *H. felis* possesses a plasmid of 6,712 bp that shares no homology with the plasmids characterized in *H. pylori* (fig. 2B). The plasmid encodes five predicted proteins, two of which—a putative murein transglycosylase and an N-acetylmuramoyl L-alanine-amiadase—are peptidoglycan-modifying enzymes with predicted autolytic activity. A replication initiation protein A (RepA) is not encoded by the plasmid; seven chromosome-encoded full-length copies of repA (*Hfelis200, 16140, 16300, and 16400*) may compensate for the plasmid’s repA deficiency and may regulate its replication in trans. Of the seven repA gene copies, two are encoded on two of three putative mobile elements (fig. 2C and D), which differ in GC content from the rest of the genome (fig. 2A). Region I (~8 kb in length, positions 1596300–1604640) shows homology and synteny with an *H. pylori* plasmid isolated from a gastric cancer patient in Peru, pHPPC4. Two genes of
region I encode an ABC-transporter-like multidrug resistance protein and a mersacidin-modifying enzyme (fig. 2C), which introduces lanthionin rings into lantibiotics such as mersacidin. Both genes show >70% similarity with their *H. pylori* homologues encoded on pHPPC4 (accession number CP002075). pHPPC4 in turn shares sequence homology with many other *H. pylori* plasmids (pHPe4, pHP69, pHPAG1, pHPG27, pHP12), which in contrast to pHPPC4 lack both the multidrug resistance and mersacidin-modifying enzyme genes. Despite the fact that region I harbors a repA gene, we have no evidence that the region replicates episomally. A second region, region II (29.8 kb in length, positions 4000–33757, fig. 2C), harbors elements with extensive homology to *Campylobacter* bacteriophage genes. At least eight region II-encoded phagic structural and morphogenesis proteins and resolvases share a high degree of homology (but no synteny) with *Campylobacter jejuni* and *hominis* phage sequences. Several of the bacteriophage genes (*Hf*elis0290, 0350, 0450, 0460, 0610) have homologues in either *H. hepaticus* and/or *H. bilis*, indicating that the sequenced strains of these two murine pathogens may also harbor prophages. Interestingly, region II is flanked by two IS605 transposases (*Hf*elis0370 and *Hf*elis0630), one of which is >60% identical to chromosomally encoded transposases of *H. pylori*. Region III (9.5 kb in length, positions 1624000–1633500, fig. 2D), in contrast, neither bears resemblance to prophages of *Campylobacter* (despite sharing homology with nonphagic, chromosomally encoded *Campylobacter* genes) nor do the mobilization and replication initiation proteins encoded by region III resemble their functional counterparts on *H. pylori* plasmids. Again, we could not find evidence for episomal replication of region III, indicating that this region may represent an integrated, no longer autonomously replicating plasmid derived from *Campylobacter*.

In summary, the sequence of *H. felis* reaffirmed many known characteristics of the *Helicobacter* genus such as the general paucity of transcriptional regulators and the abundance of restriction/modification systems and chemotaxis sensors. Our evidence for the existence of at least one *H. felis* prophage makes it unique among *Helicobacter* species, which in contrast to *Campylobacter* and with the exception of *H. acinonychis*, are not known to harbor phages. Efforts are currently under way to establish the genetic tools to manipulate *H. felis* and will hopefully allow us to dissect experimentally which virulence factors are required for gastric carcinogenesis and lymphomagenesis induced in animal models by this interesting bacterial pathogen.

**Literature Cited**


**Fig. 2.** —Helicobacter felis-specific genes, plasmid and genomic regions. (A) Circular plot showing the genomic differences and similarities of six *Helicobacter* species as indicated in the legend. Rings from outside to inside: 1, Positions of three putative horizontally acquired mobile elements (regions I–III) and the *H. felis* plasmid pHPS51 in the *H. felis* genome. 2, Positions of *H. felis*-specific genes; functional categories are color coded as indicated in the legend. 3, Nucleotide coordinates in bp. 4, *H. felis*-specific ORFs not present in any of the other five Helicobacter genomes (purple). 5, *H. felis* ORFs on the plus and minus strands (dark green). 6–10, ORFs of the indicated Helicobacter species with orthologues in *H. felis*. 11, GC content. 12, GC skew. The GC content and GC skew were calculated in Artemis with a window size of 2,000 and 5,000 bp, respectively, and an overlap of 200 bp between windows. (B) Schematic showing the positions of the five predicted ORFs of the *H. felis* plasmid pHPS51. (C) and (D) Schematics of the putative horizontally acquired regions I, II, and III with nucleotide coordinates; annotated genes are indicated using the following color code: dark green, phage structural, and morphogenesis genes; orange, phage lytic genes; red, (phage) transposases and resolvases; yellow, replication initiation proteins; blue, antibiotic resistance and modification genes; olive, mobilization proteins; gray, uncharacterized or putative genes. A schematic of the *H. pylori* plasmid pHPPC4 is shown in C for comparison.


Associate editor: Bill Martin