SHORT COMMUNICATION

Genome sequences of 65 *Helicobacter pylori* strains isolated from asymptomatic individuals and patients with gastric cancer, peptic ulcer disease, or gastritis

Thomas G. Blanchard¹, Steven J. Czinn¹, Pelayo Correa², Teruko Nakazawa³, Monika Keelan⁴, Lindsay Morningstar¹, Ivette Santana-Cruz⁵, Ankit Maroo⁵, Carri McCracken⁵, Kent Shufelt⁵, Sean Daugherty⁵, Yang Song⁵, Claire M. Fraser⁵ & W. Florian Frickie⁵

1 Department of Pediatrics, University of Maryland, Baltimore, MD, USA
2 Department of Medicine, Vanderbilt University, Nashville, TN, USA
3 Graduate School of Medicine, Yamaguchi University, Ube, Japan
4 Department of Laboratory Medicine & Pathology, University of Alberta, Edmonton, AB, Canada
5 Institute for Genomic Sciences, University of Maryland, Baltimore, MD, USA

This study provides a great resource for the *Helicobacter pylori* research community. It is the largest genome sequencing effort in this species so far and it provides an unprecedented overview of genomic diversity within the species. Both the raw data and assembled scaffolds are made available.

Keywords

stomach; pathogenesis; human.

Correspondence

Thomas G. Blanchard, University of Maryland
School of Medicine, Bresler Research Building, 13-043, 655 West Baltimore Street, Baltimore, MD 21201, USA.
Tel.: +1410-706-1772
fax: +1410-328-1072
e-mail: tblanchard@peds.umaryland.edu

Received 27 March 2013; revised 22 April 2013; accepted 28 April 2013. Final version published online 23 May 2013.
doi:10.1111/2049-632X.12045

Abstract

*Helicobacter pylori*, inhabitant of the gastric mucosa of over half of the world population, with decreasing prevalence in the U.S., has been associated with a variety of gastric pathologies. However, the majority of *H. pylori*-infected individuals remain asymptomatic, and negative correlations between *H. pylori* and allergic diseases have been reported. Comprehensive genome characterization of *H. pylori* populations from different human host backgrounds including healthy individuals provides the exciting potential to generate new insights into the open question whether human health outcome is associated with specific *H. pylori* genotypes or dependent on other environmental factors. We report the genome sequences of 65 *H. pylori* isolates from individuals with gastric cancer, preneoplastic lesions, peptic ulcer disease, gastritis, and from asymptomatic adults. Isolates were collected from multiple locations in North America (USA and Canada) as well as from Columbia and Japan. The availability of these *H. pylori* genome sequences from individuals with distinct clinical presentations provides the research community with a resource for detailed investigations into genetic elements that correlate either positively or negatively with the epidemiology, human host adaptation, and gastric pathogenesis and will aid in the characterization of strains that may favor the development of specific pathology, including gastric cancer.

The Gram-negative bacterium *Helicobacter pylori* has been colonizing the gastric mucosa of modern humans since their migration out of Africa 60 000 years ago (Linz *et al.*, 2007). Infection is typically established in childhood, maintained for life, and invariably results in histologic gastritis (Dooley *et al.*, 1989; Dehesa *et al.*, 1991; Marshall, 1995; Suerbaum & Michetti, 2002). *Helicobacter pylori* has been identified as the etiologic agent of most peptic ulcer disease (Marshall & Warren, 1984; NIH, 1994), can cause gastric mucosa-associated lymphoid tissue (MALT) lymphoma (Wotherspoon *et al.*, 1991; Genta *et al.*, 1993), and is a risk factor for the development of gastric adenocarcinoma (Parsonnet *et al.*, 1991; Talley *et al.*, 1991). The World Health Organization lists *H. pylori* as a Class I carcinogen (World Health Organization, 1994). Today, over 50% of the world population is colonized by *H. pylori*, with infection rates ranging from as high as 90% in developing nations to less than 10% in children from Western populations (Roosendaal *et al.*, 2013).
Helicobacter pylori genomes associated with gastric health and disease

T.G. Blanchard et al.

However, less than 20% of all infected individuals will develop H. pylori-associated disease. Additionally, although significantly associated with deaths due to gastric cancer, colonization is not linked to increased all-cause mortality in the U.S. (Chen et al., 2013). A negative association between H. pylori colonization and the occurrence of asthma and allergy has also been reported (Zevit et al., 2012). Whether the background of the infected host is responsible for the different H. pylori-associated health effects (human genotype, diet, influence of the surrounding stomach microbiota), or whether the outcome of H. pylori colonization is determined by the genotype of the infecting H. pylori strain remains an open question. Today, gastric cancer remains a leading cause of death in many developing and advanced nations (Yamaoka et al., 2008).

The mechanisms of H. pylori pathogenesis are complex. While some factors such as the Type IV secretion system and the CagA oncoprotein have been demonstrated to play a role in inflammation and carcinogenesis (Rieder et al., 1997; Odenbreit et al., 2000; Fischer et al., 2001; Kwok et al., 2007; Ohnishi et al., 2008; Jimenez-Soto et al., 2009), few other virulence factors have been identified that explain why certain infections cause disease, or which specific disease will develop. The H. pylori genome is diverse, typically carrying plasticity zones that vary between strains (Tomb et al., 1997; Alm et al., 1999; Occhialini et al., 2000; Kersulyte et al., 2009). In light of the unclear role of H. pylori for human health, genome sequencing of H. pylori strains isolated from individuals with different disease backgrounds, including asymptomatic carriers, can be used to identify potential H. pylori genotypes associated with specific human host disease phenotypes. We sequenced the genomes of 65 clinical H. pylori isolates recovered from subjects in geographically diverse locations with distinct clinical outcomes, including seven strains recovered from postmenopausal asymptomatic women (Table 1).

Genomic DNA was isolated from H. pylori plate cultures spread on Columbia blood agar using Genomic tip-20/G DNA isolation kits (Qiagen, Germantown, MD) and sequenced at the Institute for Genome Sciences, University of Maryland School of Medicine (http://www.igs.umaryland.edu). A hybrid sequencing approach was used combining data from 3-kb insert paired-end libraries sequenced on the 454 GS FLX Titanium platform (Roche, Branford, CT) and 300 to 400-bp insert paired-end libraries sequenced with 100-bp read length on the HiSeq 2000 platform (Illumina, San Diego, CA). On average, 27.5-fold genome coverage of 454- and 38.6-fold coverage of Illumina sequence data was used to create draft genome assemblies with the Celera assembler (http://www.ccbcb.umd.edu/research/CeleraAssembler.shtml) consisting of between 2 and 14 scaffolds per genome (average: 8.6 scaffolds) (Table 1). Draft genomes were annotated with the IGS annotation engine pipeline (http://ae.igs.umaryland.edu/cgi/ae_pipeline_outline.cgi). Between 1557 and 1831, genes were identified per genome, similar to gene numbers predicted for previously sequenced H. pylori genomes (McClain et al., 2009). Evolutionary relationships of all sequenced isolates were analyzed by whole-genome alignment with the Mugsy tool (Angiuoli & Salzberg, 2011) followed by alignment processing (concatenation, removal of gap-containing positions) with Phylomark (Sahl et al., 2012) and phylogenetic tree prediction with FastTree 2 (Price et al., 2010). This method identified a core alignment of 73kbp length consisting of multiple syntenic sequence blocks shared between all 65 compared genomes.

Phylogenetic analysis of the 65 newly sequenced H. pylori genomes revealed that the evolutionary relationships of the selected strains are not entirely determined by either the human host gastric health background (Fig. 1a) or geographic location (Fig. 1b). Only eight H. pylori isolates, which comprise strains isolated from patients with advanced stages of H. pylori-associated disease, that is, peptic ulcer disease (four isolates) or gastric cancer (four isolates) and represent the only isolates from Japan, are phylogenetically distinct from all other isolates. These isolates share a unique, relatively recent, common ancestor, that is, they are more closely related to each other than to any of the other isolates. An East Asian H. pylori lineage that diverged from European and Amerind lineage ancestors was also reported by Kawai et al. (2011). The remaining strains did not cluster by clinical outcome or by geographic location, suggesting that, based on the presented data, the evolution of sublineages within the H. pylori species is not associated with the development of more or less pathogenic potential. However, additional genome sequences from a larger number of H. pylori isolates from disparate parts of the world and from patients with distinct pathologies and including asymptomatic donors will be required to conclusively determine the role of H. pylori species evolution for human health.

Using the animal isolate H. felis ATCC 49179 as an outgroup (GenBank accession number: FQ670179), 64 of 65 isolates could be assigned to one of two sublineages within the H. pylori species. One sublineage contained a single isolate from an asymptomatic patient collected in Alberta, Canada, as well as 30 isolates collected from a diverse patient population in Cleveland, OH, USA presenting with gastritis, gastric ulcers, or duodenal ulcers. The second sublineage contained the remaining eleven isolates from the U.S. collection, the eight isolates from Japanese patients, seven isolates from Colombian patients with gastritis presenting with various stages of premalignant lesions, and six isolates from asymptomatic patients collected in Alberta, Canada. Interestingly, all H. pylori isolates are separated on the phylogenetic tree by relatively large unique branches. In most cases, any two H. pylori strains show a greater phylogenetic distance from their last common ancestor than different ancestors have from each other, suggesting that the sequenced isolates, although collected in four different countries from three continents, represent a very small fraction of the existing diversity within prevalent H. pylori populations.

In addition to allowing new insights into the genome evolution of the H. pylori species, the genome sequence data from 65 newly sequenced H. pylori strains provide a new, valuable resource for comparative genome analysis in order to elucidate the genetic markers of virulence, adaptability, antibiotic resistance, and epidemiology in this species.
Table 1 Sequenced *Helicobacter pylori* genomes.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Background*</th>
<th>Source</th>
<th>Acc.-No.</th>
<th>Scaffolds†</th>
<th>Total length</th>
<th>CDS</th>
<th>Assembler</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPY1124</td>
<td>Gastric ulcer Japan</td>
<td>AKNJ0000000000</td>
<td>13</td>
<td>1 563 197</td>
<td>1648</td>
<td>CA6.1</td>
<td>21X 454; 30X Illumina</td>
<td></td>
</tr>
<tr>
<td>CPY1313</td>
<td>Duodenal ulcer Japan</td>
<td>AKNK0000000000</td>
<td>5</td>
<td>1 581 437</td>
<td>1647</td>
<td>CA6.1</td>
<td>22X 454; 40X Illumina</td>
<td></td>
</tr>
<tr>
<td>CPY1962</td>
<td>Gastric ulcer Japan</td>
<td>AKNL0000000000</td>
<td>8</td>
<td>1 561 561</td>
<td>1607</td>
<td>CA6.1</td>
<td>20X 454; 60X Illumina</td>
<td></td>
</tr>
<tr>
<td>CPY3281</td>
<td>Duodenal ulcer Japan</td>
<td>AKNM0000000000</td>
<td>8</td>
<td>1 606 528</td>
<td>1708</td>
<td>CA6.1</td>
<td>22X 454; 40X Illumina</td>
<td></td>
</tr>
<tr>
<td>CPY6081</td>
<td>Gastric cancer Japan</td>
<td>AKNN0000000000</td>
<td>9</td>
<td>1 599 100</td>
<td>1679</td>
<td>CA6.1</td>
<td>23X 454; 20X Illumina</td>
<td></td>
</tr>
<tr>
<td>CPY6261</td>
<td>Gastric cancer Japan</td>
<td>AKNO0000000000</td>
<td>5</td>
<td>1 608 996</td>
<td>1649</td>
<td>CA6.1</td>
<td>24X 454; 50X Illumina</td>
<td></td>
</tr>
<tr>
<td>CPY6271</td>
<td>Gastric cancer Japan</td>
<td>AKNP0000000000</td>
<td>8</td>
<td>1 602 120</td>
<td>1645</td>
<td>CA6.1</td>
<td>24X 454; 60X Illumina</td>
<td></td>
</tr>
<tr>
<td>CPY6311</td>
<td>Gastric cancer Japan</td>
<td>AKNQ0000000000</td>
<td>9</td>
<td>1 596 646</td>
<td>1625</td>
<td>CA6.1</td>
<td>22X 454; 40X Illumina</td>
<td></td>
</tr>
<tr>
<td>CpY1124</td>
<td>Gastric ulcer Japan</td>
<td>AKNJ0000000000</td>
<td>13</td>
<td>1 563 197</td>
<td>1648</td>
<td>CA6.1</td>
<td>21X 454; 30X Illumina</td>
<td></td>
</tr>
<tr>
<td>CpY1313</td>
<td>Duodenal ulcer Japan</td>
<td>AKNK0000000000</td>
<td>5</td>
<td>1 581 437</td>
<td>1647</td>
<td>CA6.1</td>
<td>22X 454; 40X Illumina</td>
<td></td>
</tr>
<tr>
<td>CpY1962</td>
<td>Gastric ulcer Japan</td>
<td>AKNL0000000000</td>
<td>8</td>
<td>1 561 561</td>
<td>1607</td>
<td>CA6.1</td>
<td>20X 454; 60X Illumina</td>
<td></td>
</tr>
<tr>
<td>CpY3281</td>
<td>Duodenal ulcer Japan</td>
<td>AKNM0000000000</td>
<td>8</td>
<td>1 606 528</td>
<td>1708</td>
<td>CA6.1</td>
<td>22X 454; 40X Illumina</td>
<td></td>
</tr>
<tr>
<td>CpY6081</td>
<td>Gastric cancer Japan</td>
<td>AKNN0000000000</td>
<td>9</td>
<td>1 599 100</td>
<td>1679</td>
<td>CA6.1</td>
<td>23X 454; 20X Illumina</td>
<td></td>
</tr>
<tr>
<td>CpY6261</td>
<td>Gastric cancer Japan</td>
<td>AKNO0000000000</td>
<td>5</td>
<td>1 608 996</td>
<td>1649</td>
<td>CA6.1</td>
<td>24X 454; 50X Illumina</td>
<td></td>
</tr>
<tr>
<td>CpY6271</td>
<td>Gastric cancer Japan</td>
<td>AKNP0000000000</td>
<td>8</td>
<td>1 602 120</td>
<td>1645</td>
<td>CA6.1</td>
<td>24X 454; 60X Illumina</td>
<td></td>
</tr>
<tr>
<td>CpY6311</td>
<td>Gastric cancer Japan</td>
<td>AKNQ0000000000</td>
<td>9</td>
<td>1 596 646</td>
<td>1625</td>
<td>CA6.1</td>
<td>22X 454; 40X Illumina</td>
<td></td>
</tr>
</tbody>
</table>

*Note: *Helicobacter pylori* genomes associated with gastric health and disease.
Table 1 (continued)

<table>
<thead>
<tr>
<th>Strain</th>
<th>Background*</th>
<th>Source</th>
<th>Acc.-No.</th>
<th>Scaffolds†</th>
<th>Total length</th>
<th>CDS</th>
<th>Assembler</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>NQ4161</td>
<td>Multifocal atrophic gastritis w/o metaplasia</td>
<td>Columbia</td>
<td>AKNY00000000</td>
<td>9</td>
<td>1 643 218</td>
<td>1678</td>
<td>CA6.1</td>
<td>22X 454; 30X Illumina</td>
</tr>
<tr>
<td>NQ4200</td>
<td>Intestinal metaplasia</td>
<td>Columbia</td>
<td>AKNS00000000</td>
<td>14</td>
<td>1 646 737</td>
<td>1647</td>
<td>CA6.1</td>
<td>20X 454; 20X Illumina</td>
</tr>
<tr>
<td>NQ4216</td>
<td>Indefinite for dysplasia</td>
<td>Columbia</td>
<td>AKNT00000000</td>
<td>6</td>
<td>1 653 281</td>
<td>1627</td>
<td>CA6.1</td>
<td>42X 454; 40X Illumina</td>
</tr>
<tr>
<td>R030b</td>
<td>Asymptomatic</td>
<td>Canada</td>
<td>AMOR00000000</td>
<td>11</td>
<td>1 610 617</td>
<td>1619</td>
<td>CA7</td>
<td>46X 454; 20X Illumina</td>
</tr>
<tr>
<td>R037c</td>
<td>Asymptomatic</td>
<td>Canada</td>
<td>AMOU00000000</td>
<td>9</td>
<td>1 614 130</td>
<td>1637</td>
<td>CA7</td>
<td>47X 454; 20X Illumina</td>
</tr>
<tr>
<td>R038b</td>
<td>Asymptomatic</td>
<td>Canada</td>
<td>AMOV00000000</td>
<td>14</td>
<td>1 632 757</td>
<td>1626</td>
<td>CA7</td>
<td>46X 454; 20X Illumina</td>
</tr>
<tr>
<td>R046Wa</td>
<td>Asymptomatic</td>
<td>Canada</td>
<td>AMOW00000000</td>
<td>8</td>
<td>1 585 443</td>
<td>1610</td>
<td>CA7</td>
<td>36X 454; 10X Illumina</td>
</tr>
<tr>
<td>R056a</td>
<td>Asymptomatic</td>
<td>Canada</td>
<td>AMOS00000000</td>
<td>13</td>
<td>1 646 110</td>
<td>1652</td>
<td>CA7</td>
<td>38X 454; 20X Illumina</td>
</tr>
<tr>
<td>R32b</td>
<td>Asymptomatic</td>
<td>Canada</td>
<td>AMOS00000000</td>
<td>10</td>
<td>1 582 461</td>
<td>1607</td>
<td>CA7</td>
<td>53X 454; 10X Illumina</td>
</tr>
</tbody>
</table>

*Human host disease background.
†Scaffolds can consist of multiple contigs that are concatenated via linker sequences based on matching paired-end read information.

Fig. 1 Evolutionary relationships of sequenced Helicobacter pylori strains. (a) Comparison with human host disease background, that is, asymptomatic (green), gastritis (black), ulcer (red), or gastric cancer (blue). (b) Comparison with geographic location of human host, that is, Colombia (green), USA (black), Japan (red), or Canada (blue). Phylogenetic trees were generated with FastTree 2. The scale refers to Jukes–Cantor distances. Node labels show FastTree support values.

Acknowledgments

This project has been funded in part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under contract number HHSN2722009 00009C, and by National Institutes of Health grants AI055710, AI082655, DK46461, and CA028842.
Nucleotide sequence accession numbers

The genome sequence data have been deposited in GenBank with accession numbers listed in Table 1. This information is also available through the Institute for Genomic Sciences at the University of Maryland in Baltimore at the following site: http://gsqid.igs.umaryland.edu/wp.php?wp=comparative_sequence_analysis_of_H_pylori_isolates_from_subjects_with_distinct_gastric_pathologies

References


