


# Complete Genome Sequences of Eight *Helicobacter pylori* Strains with Different Virulence Factor Genotypes and Methylation Profiles, Isolated from Patients with Diverse Gastrointestinal Diseases on Okinawa Island, Japan, Determined Using PacBio Single-Molecule Real-Time Technology

 Kazuhito Satou,<sup>a</sup> Akino Shiroma,<sup>a</sup> Kuniko Teruya,<sup>a</sup> Makiko Shimoji,<sup>a</sup> Kazuma Nakano,<sup>a</sup> Ayaka Juan,<sup>a</sup> Hinako Tamotsu,<sup>a</sup> Yasunobu Terabayashi,<sup>a</sup> Misako Aoyama,<sup>a</sup> Morimi Teruya,<sup>b</sup> Rumiko Suzuki,<sup>c</sup> Miyuki Matsuda,<sup>c</sup> Akihiro Sekine,<sup>d</sup> Nagisa Kinjo,<sup>e</sup> Fukunori Kinjo,<sup>e</sup> Yoshio Yamaoka,<sup>c,f</sup> Takashi Hirano<sup>a</sup>

Okinawa Institute of Advanced Sciences, Uruma, Japan<sup>a</sup>; Okinawa Industrial Technology Center, Uruma, Japan<sup>b</sup>; Department of Environmental and Preventive Medicine, Oita University Faculty of Medicine, Yufu, Japan<sup>c</sup>; Pharmacogenomics Project, Kyoto University Graduate School of Medicine, Kyoto, Japan<sup>d</sup>; Department of Endoscopy, University Hospital, University of the Ryukyus, Nishihara, Japan<sup>e</sup>; Department of Medicine—Gastroenterology, Baylor College of Medicine and Michael E. DeBakey Veterans Affairs Medical Center, Houston, Texas, USA<sup>f</sup>

**We report the complete genome sequences of eight *Helicobacter pylori* strains isolated from patients with gastrointestinal diseases in Okinawa, Japan. Whole-genome sequencing and DNA methylation detection were performed using the PacBio platform. *De novo* assembly determined a single, complete contig for each strain. Furthermore, methylation analysis identified virulence factor genotype-dependent motifs.**

Received 17 March 2014 Accepted 25 March 2014 Published 17 April 2014

**Citation** Satou K, Shiroma A, Teruya K, Shimoji M, Nakano K, Juan A, Tamotsu H, Terabayashi Y, Aoyama M, Teruya M, Suzuki R, Matsuda M, Sekine A, Kinjo N, Kinjo F, Yamaoka Y, Hirano T. 2014. Complete genome sequences of eight *Helicobacter pylori* strains with different virulence factor genotypes and methylation profiles, isolated from patients with diverse gastrointestinal diseases on Okinawa Island, Japan, determined using PacBio single-molecule real-time technology. *Genome Announc.* 2(2):e00286-14. doi:10.1128/genomeA.00286-14.

**Copyright** © 2014 Satou et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Kazuhito Satou, [kazuhiro.satou@oias.or.jp](mailto:kazuhiro.satou@oias.or.jp).

*Helicobacter pylori* is a spiral-shaped, Gram-negative, microaerophilic bacterium that colonizes the stomach. Its genome consists of a circular chromosome with a size of around 1.6 Mb, an average G+C content of 39% (1), and high allelic diversity (2). Approximately half of the world's population harbors the bacterium (3). Infection with *H. pylori* occurs worldwide, but the prevalence varies greatly among populations (4). The vast majority of infected patients are asymptomatic; however, *H. pylori* infection is linked with the development of certain gastrointestinal diseases (5).

Okinawa is the southernmost prefecture of Japan, with a population of approximately 1.4 million people. Although there is no significant difference in *H. pylori* prevalence between Okinawa (42% in 2004) and other areas in Japan (6, 7), the incidence of gastric cancer in Okinawa (6.0 deaths/100,000 population in 2012) is by far the lowest in Japan (10.5 deaths/100,000 population in 2012) [Center for Cancer Control and Information Services, National Cancer Center, Japan; [http://ganjoho.jp/data/professional/statistics/odjrh3000000hwsa-att/pref\\_CancerSite\\_mortalityASR75\(1995-2012\).xls](http://ganjoho.jp/data/professional/statistics/odjrh3000000hwsa-att/pref_CancerSite_mortalityASR75(1995-2012).xls)]. In our previous study, we determined the genotypes of *cagA* and *vacA* virulence factors using PCR and Sanger-based sequencing technology and revealed an association between *H. pylori* virulence factors and gastroduodenal diseases in Okinawa (8). In the present study, we performed whole-genome sequencing and DNA methylation detection for eight *H. pylori* Okinawa strains using

PacBio single-molecule real-time (SMRT) sequencing technology (9) to gain broader insights into the virulence of *H. pylori* (10, 11).

Next-generation sequencing (NGS) technologies are widely used in genomics studies. However, due to their PCR-amplification bias and shortness of read lengths, they are inadequate to generate the finished genome assemblies of *H. pylori* strains because of the low G+C content and large numbers of repetitive regions in such strains (12). In contrast, the SMRT technology provides real-time analysis of biomolecules at single-molecule resolution. It achieves unbiased G+C coverage (13), extraordinarily long, multikilobase reads (14), and direct methylation sequencing (15).

*H. pylori* Okinawa strains examined in the present study had been previously isolated from patients with gastric atrophy, gastric ulcer, or duodenal ulcer (8). Whole-genome sequencing of eight *H. pylori* Okinawa strains was carried out using the PacBio RS (Pacific Biosciences, Menlo Park, CA) platform with a 10-kb insert library and XL/C2 chemistry. *De novo* assembly was conducted using the hierarchical genome assembly process (HGAP) workflow (16), including consensus polishing with Quiver. This workflow resulted in a single, complete contig for each genome. Annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (17). DNA methylation detection was also carried out using the kinetics data collected during the sequencing process. A total of 24 methylation motifs were found to be common to two or more strains. Interestingly, some motifs were asso-

TABLE 1 Summary of statistics for the eight *H. pylori* Okinawa strains

Strain name	Clinical presentation	Coverage depth <sup>a</sup> (fold)	Genome length (bases)	G+C content (%)	No. of genes	No. of CDS <sup>b</sup>	Accession no.
Oki102	Atrophy	489	1,633,212	38.81	1,610	1,478	CP006820
Oki112	Atrophy	328	1,637,925	38.81	1,620	1,477	CP006821
Oki128	Atrophy	261	1,553,826	38.97	1,534	1,393	CP006822
Oki154	Duodenal ulcer	822	1,599,700	38.80	1,584	1,439	CP006823
Oki422	Atrophy	432	1,634,852	38.83	1,620	1,468	CP006824
Oki673	Gastric ulcer	193	1,595,058	38.82	1,587	1,437	CP006825
Oki828	Duodenal ulcer	434	1,600,345	38.80	1,590	1,445	CP006826
Oki898	Duodenal ulcer	613	1,634,875	38.83	1,615	1,488	CP006827

<sup>a</sup> Postfilter polymerase read bases.

<sup>b</sup> CDS, coding sequences.

ciated with the genotypes of virulence factors. A summary of statistics for these eight *H. pylori* Okinawa strains is shown in Table 1. A detailed comparative analysis of these genomes will be included in a future publication.

**Nucleotide sequence accession numbers.** The complete genome sequences of all eight *H. pylori* Okinawa strains have been deposited in DDBJ/EMBL/GenBank under the accession numbers listed in Table 1.

#### ACKNOWLEDGMENT

This work was supported by the Okinawa Prefectural Government.

#### REFERENCES

- Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA, Nelson K, Quackenbush J, Zhou L, Kirkness EF, Peterson S, Loftus B, Richardson D, Dodson R, Khalak HG, Glodek A, McKenney K, Fitzgerald LM, Lee N, Adams MD, Hickey EK, Berg DE, Gocayne JD, Utterback TR, Peterson JD, Kelley JM, Cotton MD, Weidman JM, Fujii C, Bowman C, Watthey L, Wallin E, Hayes WS, Borodovsky M, Karp PD, Smith HO, Fraser CM, Venter JC. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388:539–547. <http://dx.doi.org/10.1038/41483>.
- Jolley KA, Maiden MC. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11:595. <http://dx.doi.org/10.1186/1471-2105-11-595>.
- Covacci A, Telford JL, Del Giudice G, Parsonnet J, Rappuoli R. 1999. *Helicobacter pylori* virulence and genetic geography. *Science* 284:1328–1333. <http://dx.doi.org/10.1126/science.284.5418.1328>.
- Bauer B, Meyer TF. 2011. The human gastric pathogen *Helicobacter pylori* and its association with gastric cancer and ulcer disease. *Ulcers* 2011:340157. <http://dx.doi.org/10.1155/2011/340157>.
- McCull KE. 2010. Clinical practice. *Helicobacter pylori* infection. *N. Engl. J. Med.* 362:1597–1604. <http://dx.doi.org/10.1056/NEJMc1001110>.
- Ito S, Azuma T, Murakita H, Hirai M, Miyaji H, Ito Y, Ohtaki Y, Yamazaki Y, Kuriyama M, Keida Y, Kohli Y. 1996. Profile of *Helicobacter pylori* cytotoxin derived from two areas of Japan with different prevalence of atrophic gastritis. *Gut* 39:800–806. <http://dx.doi.org/10.1136/gut.39.6.800>.
- Nobuta A, Asaka M, Sugiyama T, Kato M, Hige S, Takeda H, Kato T, Ogoshi K, Keida Y, Shinomura J. 2004. *Helicobacter pylori* infection in two areas in Japan with different risks for gastric cancer. *Aliment. Pharmacol. Ther.* 20(Suppl 1):1–6. <http://dx.doi.org/10.1111/j.1365-2036.2004.01976.x>.
- Matsunari O, Shiota S, Suzuki R, Watada M, Kinjo N, Murakami K, Fujioka T, Kinjo F, Yamaoka Y. 2012. Association between *Helicobacter pylori* virulence factors and gastroduodenal diseases in Okinawa, Japan. *J. Clin. Microbiol.* 50:876–883. <http://dx.doi.org/10.1128/JCM.05562-11>.
- Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, deWinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. 2009. Real-Time DNA sequencing from single polymerase molecules. *Science* 323:133–138. <http://dx.doi.org/10.1126/science.1162986>.
- Yamaoka Y. 2010. Mechanisms of disease: *Helicobacter pylori* virulence factors. *Nat. Rev. Gastroenterol. Hepatol.* 7:629–641. <http://dx.doi.org/10.1038/nrgastro.2010.154>.
- Yamaoka Y. 2012. Pathogenesis of *Helicobacter pylori*-related gastroduodenal diseases from molecular epidemiological studies. *Gastroenterol. Res. Pract.* 2012:371503. <http://dx.doi.org/10.1155/2012/371503>.
- Perkins TT, Tay CY, Thirriot F, Marshall B. 2013. Choosing a benchtop sequencing machine to characterise *Helicobacter pylori* genomes. *PLoS One* 8:e67539. <http://dx.doi.org/10.1371/journal.pone.0067539>.
- Shin SC, Ahn do H, Kim SJ, Lee H, Oh TJ, Lee JE, Park H. 2013. Advantages of single-molecule real-time sequencing in high-GC content genomes. *PLoS One* 8:e68824. <http://dx.doi.org/10.1371/journal.pone.0068824>.
- Koren S, Harhay GP, Smith TP, Bono JL, Harhay DM, McVey SD, Radune D, Bergman NH, Phillippy AM. 2013. Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biol.* 14:R101. <http://dx.doi.org/10.1186/gb-2013-14-9-r101>.
- Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7:461–465. <http://dx.doi.org/10.1038/nmeth.1459>.
- Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10:563–569. <http://dx.doi.org/10.1038/nmeth.2474>.
- Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, Kodira CD, Kyrpides N, Madupu R, Markowitz V, Tatusova T, Thomson N, White O. 2008. Toward an online repository of standard operating procedures (SOPs) for (meta)genomic annotation. *OMICS* 12:137–141. <http://dx.doi.org/10.1089/omi.2008.0017>.