Journal of Bacteriology

Draft Genome Sequences of Helicobacter pylori Isolates from Malaysia, Cultured from Patients with Functional Dyspepsia and Gastric Cancer

Selva Perumal Gunaletchumy, Xinsheng Teh, Yalda Khosravi, Nur Siti Khadijah Ramli, Eng Guan Chua, Thevakumar Kavitha, Joanne N. Mason, Huey Tyng Lee, Halimah Alias, Nur Zafirah Zaidan, Norzawani Buang M. Yassin, Liang Chung Tay, Stephen Rudd, Hazel M. Mitchell, Nadeem O. Kaakoush, Mun Fai Loke, Khean Lee Goh and Jamuna Vadivelu

J. Bacteriol. **2012, 194(20):5695. DOI: 10.1128/JB.01278-12.**

JB Journals.ASM.org

GENOME ANNOUNCEMENT

Draft Genome Sequences of *Helicobacter pylori* Isolates from Malaysia, Cultured from Patients with Functional Dyspepsia and Gastric Cancer

Selva Perumal Gunaletchumy,[a] Xinsheng Teh,[a] Yalda Khosravi,[a] Nur Siti Khadijah Ramli,[a] Eng Guan Chua,[a] Thevakumar Kavitha,[a] Joanne N. Mason,[b] Huey Tyng Lee,[b] Halimah Alias,[b] Nur Zafirah Zaidan,[b] Norzawani Buang M. Yassin,[b] Liang Chung Tay,[b] Stephen Rudd,[b] Hazel M. Mitchell,[c] Nadeem O. Kaakoush,[c] Mun Fai Loke,[a] Khean Lee Goh,[a] and Jamuna Vadivelu[a]

University of Malaya, Kuala Lumpur, Malaysia[a]; Malaysian Genomics Resource Centre Berhad, Kuala Lumpur, Malaysia[b]; and University of New South Wales, Sydney, Australia[c]

Helicobacter pylori **is the main bacterial causative agent of gastroduodenal disorders and a risk factor for gastric adenocarcinoma and mucosa-associated lymphoid tissue (MALT) lymphoma. The draft genomes of 10 closely related *H. pylori* isolates from the multiracial Malaysian population will provide an insight into the genetic diversity of isolates in Southeast Asia. These isolates were cultured from gastric biopsy samples from patients with functional dyspepsia and gastric cancer. The availability of this genomic information will provide an opportunity for examining the evolution and population structure of *H. pylori* isolates from Southeast Asia, where the East meets the West.**

Malaysia is among countries with an intermediate gastric cancer incidence, demonstrating significant differences in the three major ethnic groups (Malay, Chinese, and Indian) in *Helicobacter pylori* prevalence and gastric cancer incidence (2). *H. pylori* hspIndia (colonizing mainly Indian and Malay subjects) and hspEAsia (found mainly in Chinese subjects) are the major subpopulations isolated in this region, accounting for 41.5% and 39.0% of all isolates, respectively (8). Given the limited information on genomes of *H. pylori* isolated from Southeast Asia, located at the crossroads between East and West, the current study focused on the investigation of similarities and differences in genomes of *H. pylori* isolated from subjects of different ethnic backgrounds residing in Malaysia.

Whole-genome sequencing was performed using 100-base, paired-end reads on the Illumina HiSeq2000 instrument (Illumina, Inc., San Diego, CA) at the Malaysian Genomics Resource Centre Berhad (MGRC), Malaysia). *De novo* assembly was performed using the ABySS software program with a *k*-mer of 55 (7). Contigs produced were then grouped and reassembled using the software program Phrap. Paired-end information on reads was used to scaffold contigs together using the program MIP Scaffolder 0.5 (6). Sequencing statistics and genome information for each genome are summarized in Table 1.

All isolates were positive for the well-described housekeeping genes, which include *atpA* (a gene encoding the ATP synthase subunit A chain), *glr* (a glutamate racemase gene), *ppa* (an inorganic pyrophosphatase gene), *efp* (an elongation factor p gene), *trpC* (a bifunctional indole-3-glycerol phosphate synthase gene), *fur* (a ferric uptake regulation protein gene), and *cysS* (a cysteinyl-tRNA synthetase gene). In addition, all isolates were also positive for virulence genes: the *cag* pathogenicity island (PAI), *vacA*, and *homAB*.

It was predicted that the assembled genomes in this study contain approximately 1,620 genes (average), which is consistent with the *H. pylori* 26695 and J99 genomes, which contain 1,590 and 1,495 genes, respectively (1, 9). Based on the genomes of 26695 and J99, Salama et al. (5) and Gressmann et al. (3) attempted to provide an estimate of the number of genes belonging to the core

Received 17 July 2012 **Accepted** 3 August 2012
Address correspondence to Jamuna Vadivelu, jamuna@ummc.edu.my.

doi:10.1128/JB.01278-12

TABLE 1 Sequencing statistics and genome information

Sample ID	No. of contigs (≥500 bp)	No. of bases (≥500 bp)	No. of scaffolds	No. of bases	Maximum scaffold size	Mean scaffold size	N_{50}	Avg sequencing coverage (times)	Genome size (bp)	GC content (%)	Predicted no. of protein-coding sequences	Remarks[a]
FD568	155	1,592,946	72	1,604,711	437,216	17,497	78,927	172	1,494,164	38.54	1,587	hspEAsia; FD
GC26	135	1,582,963	77	1,597,232	203,946	18,173	54,475	156	1,534,861	38.58	1,613	hspEAsia; GC
FD506	245	1,573,056	58	1,608,477	505,763	22,085	94,183	82	1,535,289	38.18	1,672	hspEAsia; FD
FD577	115	1,592,290	59	1,609,138	314,417	24,775	59,288	197	1,580,091	38.40	1,601	hspEAsia; FD
FD662	134	1,632,154	58	1,659,218	174,469	25,557	62,179	164	1,481,802	38.52	1,640	hspIndia; FD
FD719	140	1,617,260	72	1,628,999	220,324	19,662	92,906	193	1,471,471	38.90	1,612	hspIndia; FD
FD703	126	1,622,324	70	1,637,264	204,996	21,842	53,160	173	1,540,544	38.85	1,611	hspIndia; FD
FD430	161	1,620,357	70	1,638,660	214,779	20,009	75,391	149	1,459,901	38.89	1,644	hspIndia; FD
FD535	117	1,625,362	62	1,633,165	319,872	23,029	86,768	211	1,556,024	38.91	1,588	hspIndia; FD
FD423	190	1,599,675	80	1,620,118	147,294	17,089	74,214	146	1,547,587	38.89	1,627	hspIndia; FD

[a] FD, functional dyspepsia; GC, gastric cancer.

genome of *H. pylori*, their estimates being 1,281 and 1,111 genes, respectively. In comparison, using the predicted genes from this study, which spans two subpopulations (hpAsia2/hspIndia and hpEastAsia/hspEAsia) and two disease groups, the core genome of *H. pylori* was extrapolated to contain no more than 760 genes. With less than 50 percent of its gene pool being well conserved across the entire *H. pylori* species, this study suggests that *H. pylori* may be genetically even more diverse that previously thought.

In conclusion, the availability of sequences of these closely related isolates will provide a platform for further analysis of genomic variability and plasticity, as well as bacterial evolution. Most importantly, data presented in this study have highlighted a need to take into consideration geographical and population variations in future genomic studies.

Nucleotide sequence accession numbers. The *H. pylori* draft genomes in this study have been deposited as a whole-genome shotgun project (BioProject ID no. PRJNA165757) at DDBJ/EMBL/GenBank under the accession numbers AKHM00000000 (*H. pylori* FD423), AKHN00000000 (FD430), AKHO00000000 (FD506), AKHP00000000 (FD535), AKHQ00000000 (FD568), AKHR00000000 (FD577), AKHS00000000 (FD703), AKHT00000000 (FD662), AKHU00000000 (FD719), and AKHV00000000 (GC26). The version described in this article is the first version, accession numbers AKHM01000000 to AKHV01000000.

ACKNOWLEDGMENT

We thankfully acknowledge support received from the University of Malaya-Ministry of Higher Education (UM-MOHE) High Impact Research (HIR) grant (reference UM.C/625/1/HIR/MOHE/CHAN-02; account no. A000002-50001, "Molecular Genetics").

REFERENCES

1. **Alm RA, et al.** 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. Nature **397**:176–180.
2. **Fock KM, Ang TL.** 2010. Epidemiology of *Helicobacter pylori* infection and gastric cancer in Asia. J. Gastroenterol. Hepatol. **25**:479–486.
3. **Gressmann H, et al.** 2005. Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. PLoS Genet. **1**:e43. doi:10.1371/journal.pgen.0010043.
4. Reference deleted.
5. **Salama N, et al.** 2000. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. Proc. Natl. Acad. Sci. U. S. A. **97**:14668–14673.
6. **Salmela L, Mäkinen V, Välimäki N, Ylinen J, Ukkonen E.** 2011. Fast scaffolding with small independent mixed integer programs. Bioinformatics **27**:3259–3265.
7. **Simpson JT, et al.** 2009. ABySS: a parallel assembler for short read sequence data. Genome Res. **19**:1117–1123.
8. **Tay CY, et al.** 2009. Population structure of *Helicobacter pylori* among ethnic groups in Malaysia: recent acquisition of the bacterium by the Malay population. BMC Microbiol. **9**:126. doi:10.1186/1471-2180-9-126.
9. **Tomb JF, et al.** 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature **388**:539–547.