JB
Journals.ASM.org

Next-Generation Sequencing and *De Novo* Assembly, Genome Organization, and Comparative Genomic Analyses of the Genomes of Two *Helicobacter pylori* Isolates from Duodenal Ulcer Patients in India

Narender Kumar,[a] Asish K. Mukhopadhyay,[b] Rajashree Patra,[b] Ronita De,[b] Ramani Baddam,[a] Sabiha Shaik,[a] Jawed Alam,[b] Suma Tiruvayipati,[a,c] and Niyaz Ahmed[a,c,d]

Pathogen Biology Laboratory, Department of Biotechnology, School of Life Sciences, University of Hyderabad, Gachibowli, Hyderabad, India[a]; Division of Bacteriology, National Institute of Cholera and Enteric Diseases (Indian Council of Medical Research), Kolkata, India[b]; Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia[c]; and Institute of Life Sciences, University of Hyderabad Campus, Gachibowli, Hyderabad, India[d]

The prevalence of different *H. pylori* genotypes in various geographical regions indicates region-specific adaptations during the course of evolution. Complete genomes of *H. pylori* from countries with high infection burdens, such as India, have not yet been described. Herein we present genome sequences of two *H. pylori* strains, NAB47 and NAD1, from India. In this report, we briefly mention the sequencing and finishing approaches, genome assembly with downstream statistics, and important features of the two draft genomes, including their phylogenetic status. We believe that these genome sequences and the comparative genomics emanating thereupon will help us to clearly understand the ancestry and biology of the Indian *H. pylori* genotypes, and this will be helpful in solving the so-called Indian enigma, by which high infection rates do not corroborate the minuscule number of serious outcomes observed, including gastric cancer.

Helicobacter pylori's coevolution with its host (10, 11, 16) and its tight compartmentalization (13, 16, 18, 19) into several different populations and subpopulations have delivered an excellent premise to pursue the idea of geographic evolution/spread of humans and their pathogens from Africa and to gain insights into pathogen adaptation mechanisms (1, 3). Based partly on these conventions, Indian *H. pylori* isolates have shown to have European origins (9) and are widely held as mostly innocuous or only mildly pathogenic. The severity of *H. pylori*-induced gastro-duodenal diseases and their outcomes vary in different geographic regions and populations, which may be significantly attributable to different genetic compositions of the underlying bacterial strains. More data based on genome sequences from many of strains from different countries are needed to clearly establish the genetic makeup, colonization potential, and virulence characteristics of a particular strain or genotype. In view of this, genome sequence-based characterizations of strains prevalent in different locales is necessary (2).

We describe genomes of *H. pylori* strains NAB47 (Bangalore) and NAD1 (Delhi) from duodenal ulcer patients. Illumina sequencing was performed as described previously (4, 8); briefly, about 3 gigabytes and 1.8 gigabytes of data comprising 72-bp paired-end reads (insert size, 300 bp) provided genome coverages of approximately 300× and 200×, respectively. The raw reads were filtered using the FASTX tool kit (17) and assembled using Velvet (20); the reads yielded 107 (NAB47) and 103 (NAD1) contigs with a hash length set to 37. These contigs were joined into 34 (NAB47) and 48 (NAD1) scaffolds by using SSPACE (6). The scaffolds were aligned and ordered according to their closest reference genome and confirmed using BLAST (12) and Mummer (14). The draft genomes were submitted to RAST (5) for annotation, and the output was validated by using Glimmer (7) and EasyGene (15).

The draft genomes of *H. pylori* NAB47 and NAD1 had sizes of about 1,590,862 bp and 1,588,938 bp, respectively, with G+C contents of 39.17 and 39.03%, respectively. The genomes revealed coding percentages of 91.5% (NAB47) and 91.3% (NAD1) and encoded 1,572 and 1,567 proteins, respectively; each of the genomes contained 36 tRNA genes and 6 rRNA genes. The average lengths for protein-coding genes were found to be 929 bp and 922 bp, respectively. Major virulence markers, such as *cagA*, *vacA*, the whole *cag* pathogenicity island, and several outer membrane proteins of the Hop family, were annotated. In addition, NAD1 harbored two plasmids of 16 kb and 10 kb each that carried genes for transposase, IS*606*, and mobilization proteins, together with replication protein A. CagA protein in both of the strains contained EPIYA D-type motifs, which are typical of Indo-European strains. Important plasticity region genes, such as jhp0940, jhp0947, and *dupA*, were absent, and hp0986 was detected only in NAB47. Finally, whole-genome phylogeny incorporating all the available genomes reconfirmed an Indo-European ancestry (HpEurope).

We believe that the genomes described herein are likely to rekindle our knowledge of the genetic makeup and evolutionary relationships of *H. pylori* in India. Comparative genomic analyses extending out to other unexplored strains from the tribal and mainstream populations will facilitate understanding of the true pathogenic potential (amid adaptive evolution) of the Indian *H. pylori*. Furthermore, they will be immensely helpful in global epidemiological studies and also for the development of diagnostic tools tailored to a particular host population.

Received 31 July 2012 Accepted 17 August 2012
Address correspondence to Niyaz Ahmed, niyazSL@uohyd.ernet.in.
N.K. and A.K.M. contributed equally to this work.

doi:10.1128/JB.01371-12

Nucleotide sequence accession numbers. The genome sequences of *H. pylori* NAB47 and NAD1 have been deposited with GenBank and assigned accession numbers AJFA00000000 and AJGJ00000000, respectively. The updated sequences/contigs are also available for download from the International Society for Genomic and Evolutionary Microbiology (ISOGEM) server (http://isogem.org/HPNAB47.txt and http://isogem.org/HPNAD1.txt).

ACKNOWLEDGMENTS

We acknowledge support from the University of Malaya High Impact Research Grant (UM.C/625/1HIR/MOHE/CHAN-02)-Molecular Genetics. A.K.M. acknowledges support from the Department of Biotechnology (BT/PR10407/BRB/10/604/2008) and Indian Council of Medical Research. These genomes were completed under the wider umbrella of the Indo-German International Research Training Group, Internationales Graduiertenkolleg (GRK1673), Functional Molecular Infection Epidemiology, an initiative of the German Research Foundation (DFG) and the University of Hyderabad (India). N.K. would like to acknowledge a Junior Research Fellowship received from the Council of Scientific and Industrial Research (CSIR), India, and J.A. acknowledges ICMR for a Senior Research Fellowship.

We are also grateful to M/s Genotypic Technology Pvt. Ltd., Bengaluru, India, for their efforts with the Illumina sequencing. We acknowledge the Bioinformatics Facility (BIF) at the Department of Biotechnology, University of Hyderabad, for use of their computational infrastructure. Further, we thank Akash Ranjan for helpful discussions and for enabling access to the SUN Microsystems CDFD Centre of Excellence for some of our data analyses.

REFERENCES

1. **Ahmed N.** 2011. Coevolution and adaptation of *Helicobacter pylori* and the case for 'functional molecular infection epidemiology.' Med. Princ. Pract. **20**:497–503.
2. **Ahmed N.** 2009. A flood of microbial genomes: do we need more? PLoS One **4**:e5831. doi:10.1371/journal.pone.0005831.
3. **Atherton JC, Blaser MJ.** 2009. Coadaptation of *Helicobacter pylori* and humans: ancient history, modern implications. J. Clin. Invest. **119**:2475–2487.
4. **Avasthi TS, et al.** 2011. Genomes of two chronological isolates (*Helicobacter pylori* 2017 and 2018) of the West African *Helicobacter pylori* strain 908 obtained from a single patient. J. Bacteriol. **193**:3385–3386.
5. **Aziz RK, et al.** 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics **9**:75. doi:10.1186/1471-2164-9-75.
6. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W.** 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **5**:578–579.
7. **Delcher AL, et al.** 1999. Improved microbial gene identification with Glimmer. Nucleic Acids Res. **27**:4636–4641.
8. **Devi SH, et al.** 2010. Genome of *Helicobacter pylori* strain 908. J. Bacteriol. **192**:6488–6489.
9. **Devi SM, et al.** 2007. Ancestral European roots of *Helicobacter pylori* in India. BMC Genomics **8**:184. doi:10.1186/1471-2164-8-184.
10. **Devi SM, et al.** 2006. Genomes of *Helicobacter pylori* from native Peruvians suggest admixture of ancestral and modern lineages and reveal a Western type *cag*-pathogenicity island. BMC Genomics **7**:191. doi:10.1186/1471-2164-7-191.
11. **Falush D, et al.** 2003. Traces of human migrations in *Helicobacter* pylori populations. Science **299**:1582–1585.
12. **Kent WJ.** 2002. BLAT: the BLAST-like alignment tool. Genome Res. **12**:656–664.
13. **Kersulyte D, et al.** 2010. Helicobacter pylori from Peruvian Amerindians: traces of human migrations in strains from remote Amazon, and genome sequence of an Amerind strain. PLoS One **5**:e15076. doi:10.1371/journal.pone.0015076.
14. **Kurtz S, et al.** 2004. Versatile and open software for comparing large genomes. Genome Biol. **5**:R12. doi:10.1186/gb-2004-5-2-r12.
15. **Larsen TS, Krogh A.** 2003. EasyGene: a prokaryotic gene finder that ranks ORFs by statistical significance. BMC Bioinformatics **4**:21. doi:10.1186/1471-2105-4-21.
16. **Linz B, et al.** 2007. An African origin for the intimate association between humans and *Helicobacter pylori.* Nature **445**:915–918.
17. **Taylor J, Schenck I, Blankenberg D, Nekrutenko A.** 2007. Using Galaxy to perform large-scale interactive data analyses. Curr. Protoc. Bioinformatics **Chapter 10:**Unit 10.5.
18. **Wirth T, et al.** 2004. Distinguishing human ethnic groups by means of sequences from *Helicobacter pylori*: lessons from Ladakh. Proc. Natl. Acad. Sci. U. S. A. **101**:4746–4751.
19. **Yamaoka Y.** 2009. Helicobacter pylori typing as a tool for tracking human migration. Clin. Microbiol. Infect. **15**:829–834.
20. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**:821–829.