

Genome Sequences of Three hpAfrica2 Strains of *Helicobacter pylori*

Stacy S. Duncan,^a M. Teresita Bertoli,^{b*} Dangeruta Kersulyte,^b Pieter L. Valk,^a Sravya Tamma,^b Issy Segal,^c Mark S. McClain,^a Timothy L. Cover,^{a,d,e} Douglas E. Berg^{b,f*}

Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee, USA^a; Department of Molecular Microbiology, Washington University Medical School, St. Louis, Missouri, USA^b; Department of Gastroenterology, Chris Hani Baragwanath Hospital, Johannesburg, South Africa^c; Department of Pathology, Microbiology and Immunology, Vanderbilt University School of Medicine, Nashville, Tennessee, USA^d; Veterans Affairs Tennessee Valley Healthcare System, Nashville, Tennessee, USA^e; Division of Infectious Disease, Department of Medicine, University of California, San Diego, La Jolla, California, USA^f

* Present address: Douglas E. Berg, Division of Infectious Disease, Department of Medicine, University of California, San Diego, La Jolla, California, USA; M. Teresita Bertoli, Departamento de Microbiología, Universidad Dr. José Matías Delgado, Antiguo Cuscatlán, El Salvador.

S.S.D., M.T.B., and D.K. contributed equally to this work.

We present the genome sequences of three hpAfrica2 strains of *Helicobacter pylori*, which are postulated to have evolved in isolation for many millennia in people of San ethnicity. Although previously considered to be ancestral to *Helicobacter acinonychis*, the hpAfrica2 strains differ markedly from *H. acinonychis* in their gene arrangement. These data provide new insights into *Helicobacter* evolution.

Received 13 August 2013 Accepted 28 August 2013 Published 26 September 2013

Citation Duncan SS, Bertoli MT, Kersulyte D, Valk PL, Tamma S, Segal I, McClain MS, Cover TL, Berg DE. 2013. Genome sequences of three hpAfrica2 strains of *Helicobacter pylori*. *Genome Announc*. 1(5):e00729-13. doi:10.1128/genomeA.00729-13.

Copyright © 2013 Duncan et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to Timothy L. Cover, timothy.l.cover@vanderbilt.edu, or Douglas E. Berg, douglaseberg@gmail.com.

Helicobacter pylori chronically colonizes the stomachs of billions of people worldwide, sometimes leading to gastroduodenal disease, although usually without symptoms and sometimes with beneficial effects (1). *H. pylori* is very diverse genetically, with different sets of genotypes predominating in well-separated human populations (2, 3). Two superlineages have been described: one including hpNEAfrica, hpAfrica1, and all non-African *H. pylori* strains and one including hpAfrica2, which is postulated to have diverged from mainstream *H. pylori* some 100,000 years ago and evolved in isolation in southern African San (hunter-gatherer, click-speaking) societies (4). We sequenced the genomes of three strains classified by multilocus sequence typing (MLST) as being of the unusual hpAfrica2 type, found among 17 strains cultured with informed consent from patients from Soweto in Chris Hani Baragwanath Hospital, Johannesburg, South Africa, from 1997 to 1999. Notably, these patients were of Bantu, not San, ethnicity.

DNAs from single colonies were sequenced using 454 methods (read lengths, ≥ 400 bp; coverage, ≥ 25 -fold) and assembled with 454 Newbler software. All gaps between contigs were filled by PCR and capillary sequencing (5), and sequences were annotated at NCBI (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/; strain SouthAfrica7) or with the Institute for Genome Sciences (IGS) Annotation Engine (<http://ae.igs.umaryland.edu/cgi/index.cgi>) (6). The three SAfrica strain genome sizes are 1.6 to 1.7 Mb, which is typical for *H. pylori*.

A BLASTn analysis of sequential 1-kb fragments genome-wide indicated that these SAfrica strains are closely related to one another, with little if any admixture from non-hpAfrica2 strains. nWayComp analysis (7) identified 173 open reading frames (ORFs) in *Helicobacter acinonychis* (GenBank accession no. NC_008229.1, postulated to descend from hpAfrica2) (4) that are absent in the SAfrica strains and, conversely, 93 ORFs in the SA-

frica strains that are absent from *H. acinonychis*. One gene present in all three SAfrica strains (HPSA_05790, encoding the DnaJ-like heat shock protein) had no close homolog in the current NCBI database, and another (HPSA_06345, hypothetical) was absent from all but one other *H. pylori* strain. Each SAfrica strain lacks the *cag* pathogenicity island and contains a type II *hopQ* outer membrane protein gene. The SAfrica strains contain s2/m2-type *vacA* genes (although *vacA* in SouthAfrica7 is frameshifted). These features are associated with low gastroduodenal disease risk.

Mauve analysis (8) illustrated that the *H. pylori* genome organization is well conserved, with < 10 rearrangements (inversions/translocations, defined as locally colinear blocks; <http://gel.ahabs.wisc.edu/mauve/>) among the SAfrica strains and < 20 such differences between the SAfrica and non-hpAfrica2 strains, such as J99. However, > 100 rearrangements distinguished *H. acinonychis* from these putatively ancestral hpAfrica2-type strains. The SAfrica strains lack all but one *H. acinonychis* prophage gene (a gene also present in many prophage-free *H. pylori* strains and thus probably not of viral origin).

Collectively, these outcomes heighten interest in the proposed ancestry relationships between *H. pylori* and *H. acinonychis* (4) and more generally between the many helicobacters of humans and other mammals. Further analyses of hpAfrica2 genomes should provide new insights into important evolutionary questions—for example, how bacterial superlineages and distinct species (here, hpAfrica2 from ethnic Bantu Soweto residents, and *H. acinonychis*) arose and diverged from ancestral forms and remained distinct genome-wide—especially for species such as *H. pylori*, generally considered to enjoy “free recombination” (9).

Nucleotide sequence accession numbers. The genome sequences of these *H. pylori* strains have been deposited in GenBank under the following accession no.: SouthAfrica7, CP002336;

SouthAfrica20, CP006691; and SouthAfrica50, AVNI0000000. The versions described in this paper are the first versions (SouthAfrica7, CP002336.1; SouthAfrica 20, CP006691.1; and SouthAfrica 50, AVNI01000001.1 and AVNI01000002.1).

ACKNOWLEDGMENTS

This work was supported by NIH R01 AI039657, R01 AI068009, P01 CA116087, R21 AI078237, and R21 AI088337 and the Department of Veterans Affairs.

We thank Reidwaan Ally and Rajesh Mistry for help in obtaining gastric biopsy samples and initial culturing of *H. pylori* strains from Chris Hani Baragwanath Hospital patients. We also thank M0Gene LC (St. Louis, MO) for sequencing of SouthAfrica7, the Vanderbilt Technologies for Advanced Genomics facility for sequencing of SouthAfrica20 and SouthAfrica50, the NCBI Prokaryotic Genome Annotation Pipeline and the Institute for Genome Sciences Annotation Engine service at the University of Maryland School of Medicine for annotation of the SAfrica strain genome sequences, and the IGS Annotation Engine team for assistance in the submission of the annotated SouthAfrica20 and SouthAfrica50 sequences to GenBank.

REFERENCES

- Cover TL, Blaser MJ. 2009. *Helicobacter pylori* in health and disease. *Gastroenterology* 136:1863–1873.
- Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, Blaser MJ, Graham DY, Vacher S, Perez-Perez GI, Yamaoka Y, Mégraud F, Otto K, Reichard U, Katzowitzsch E, Wang X, Achtman M, Suerbaum S. 2003. Traces of human migrations in *Helicobacter pylori* populations. *Science* 299:1582–1585.
- Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, Falush D, Stamer C, Prugnolle F, van der Merwe SW, Yamaoka Y, Graham DY, Perez-Trallero E, Wadstrom T, Suerbaum S, Achtman M. 2007. An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* 445:915–918.
- Moodley Y, Linz B, Bond RP, Nieuwoudt M, Soodyall H, Schlebusch CM, Bernhöft S, Hale J, Suerbaum S, Mugisha L, van der Merwe SW, Achtman M. 2012. Age of the association between *Helicobacter pylori* and man. *PLOS Pathog.* 8:e1002693. doi:10.1371/journal.ppat.1002693.
- Kersulyte D, Kalia A, Gilman RH, Mendez M, Herrera P, Cabrera L, Velapatiño B, Balqui J, Puente de la Vega F, Rodriguez Ulloa CA, Cok J, Hooper CC, Dailide G, Tamma S, Berg DE. 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.
- Galens K, Orvis J, Daugherty S, Creasy HH, Angiuoli S, White O, Wortman J, Mahurkar A, Giglio MG. 2011. The IGS standard operating procedure for automated prokaryotic annotation. *Stand. Genomic Sci.* 4:244–251.
- Yao J, Lin H, Doddapaneni H, Civerolo EL. 2007. nWayComp: a genome-wide sequence comparison tool for multiple strains/species of phylogenetically related microorganisms. *In Silico Biol.* 7:195–200.
- Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. doi:10.1371/journal.pone.0011147.
- Suerbaum S, Smith JM, Bapumia K, Morelli G, Smith NH, Kunstmann E, Dyrek I, Achtman M. 1998. Free recombination within *Helicobacter pylori*. *Proc. Natl. Acad. Sci. U. S. A.* 95:12619–12624.