

Whole Genome Sequencing and Comparative Analysis of *Bartonella bacilliformis* Strain INS, the Causative Agent of Carrion's Disease

D. Tarazona,^a C. Padilla,^a O. Cáceres,^a J. D. Montenegro,^a H. Bailón,^a G. Ventura,^a G. Mendoza,^b E. Anaya,^b H. Guio^{a,c}

Laboratorio de Biotecnología y Biología Molecular, Instituto Nacional de Salud, Lima, Peru^a; Laboratorio de Metaxénicas Bacterianas, Instituto Nacional de Salud, Lima, Peru^b; Asociación Latinoamericana de Biotecnología (ALBIOTEC), Lima, Peru^c

***Bartonella bacilliformis* is the etiological agent of human bartonellosis, which is highly endemic to Peru. Here, we report the first genome that was sequenced and analyzed from an isolate of *B. bacilliformis* strain INS, which originally was isolated from the blood of an infected patient with an acute phase of Carrion's disease from Jaén-Cajamarca, Peru.**

Received 18 October 2012 Accepted 18 December 2012 Published 14 February 2013

Citation Tarazona D, Padilla C, Cáceres O, Montenegro JD, Bailón H, Ventura G, Mendoza G, Anaya E, Guio H. 2013. Whole genome sequencing and comparative analysis of *Bartonella bacilliformis* strain INS, the causative agent of Carrion's disease. *Genome Announc.* 1(1):e00053-12. doi:10.1128/genomeA.00053-12.

Copyright © 2013 Tarazona 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 H. Guio, hguio@ins.gob.pe.

Carrion's disease, caused by *Bartonella bacilliformis*, is commonly localized in the Andean region of South America, particularly in the valleys lining the western side of the Andes (1). Peru has reported rates of 242 per 100,000 inhabitants in the regions of Ancash, Cajamarca, and Amazonas (<http://www.dge.gob.pe/>).

Carrion's disease has two clinical phases. During the acute phase, also called "Oroya's fever," *B. bacilliformis* invades red blood cells and produces severe hemolytic anemia (2). The chronic phase, also known as "verruca peruana," is characterized by skin nodules and mulaire lesions, which usually bleed and lead to fibrosis (3). The differences in the manifestation of the disease during the two phases of bartonellosis suggest a complex interaction between *B. bacilliformis* and the human host that may involve a multitude of both bacterial and host proteins. Here, we report the whole-genome sequence of *B. bacilliformis* INS, which was isolated from a human patient.

Genome sequencing of *B. bacilliformis* strain INS was carried out at the Peruvian National Institute of Health using the SOLiD 3 Plus genome analyzer, with 700-fold coverage. Sequence reads were aligned to a consensus sequence and were assembled by Bowtie 2 (4) to the *B. bacilliformis* KC583 (NC_008783.1), providing a total of 20 contigs. The protein-coding genes were predicted using Glimmer 3.0 (5); tRNAs and rRNAs were identified with tRNAscan-SE (6) and RNAmmer (7), respectively. The genome sequence was annotated using Rapid Annotations using Subsystem Technology (RAST) (8) and Prokaryotic Genome Annotation Pipeline (PGAAP) (9). The functions of the predicted protein-coding genes were annotated with the NCBI nonredundant (NR) (10) and Clusters of Orthologous Groups (COG) (11) databases. The *B. bacilliformis* INS draft genome sequence has a total of 1,444,107 bp with an average G+C content of 38.3%. It contains 1,244 predicted coding sequences (CDSs), 4 rRNAs, and 44 tRNAs.

Using COG functional assignment, the predicted proteins (73.2%) could be classified into 18 COG categories. There are 950 subsystems represented in the genome. The five most abundant

subsystems are related to protein metabolism ($n = 189$); cofactors, vitamins, and pigments ($n = 122$); RNA metabolism ($n = 114$); amino acids ($n = 99$); and DNA metabolism ($n = 73$). In addition, many CDSs are involved in the cell wall and capsule ($n = 70$); membrane transport ($n = 34$); and virulence, disease, and defense ($n = 32$). These findings indicate that strain INS has various genetic processes and a unique adaptation to human blood cells.

Pair-wise BLAST comparisons with Efficient Database framework for comparative Genome Analyses using BLAST score Ratios (EDGAR) (12) of the protein sequences from *B. bacilliformis* INS, *Bartonella henselae* (NC_005956.1), and *Bartonella quintana* (NC_005955.1) revealed 342 CDSs unique to *B. bacilliformis* INS, and the core had 930 genes in common. We identified genes homologous to the invasion-associated locus B family protein, which is responsible for invading human erythrocytes (13).

The INS genome contains genes encoding flagellar proteins, which are essential for erythrocyte invasion (2) and confer an immunogenic effect (14). Other proteins showed similarity to OMP25 and D15 proteins, which have been defined as immunogenic and protective antigens (15–17) and which present mutations in the *gyrA* gene correlated to quinolone resistance (18).

The genomic sequencing of *B. bacilliformis* strain INS will provide insight into the conservation of genomic sequences. The data generated herein will be important for the further understanding of the molecular mechanisms of Carrion's disease pathology and will generate support for rapid diagnosis by enzyme-linked immunosorbent assay (ELISA), PCR, and possible vaccine targets.

Nucleotide sequence accession numbers. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. [AMQK000000000](https://www.ncbi.nlm.nih.gov/nuccore/AMQK000000000). The version described in this article is the first version, [AMQK010000000](https://www.ncbi.nlm.nih.gov/nuccore/AMQK010000000).

ACKNOWLEDGMENTS

This work was supported by the Peruvian National Institute of Health. H.G. also thanks the Peruvian Science and Technology Program (FIN-CyT) and CONCyTEC.

REFERENCES

1. Maguiña C, Guerra H, Ventosilla P. 2009. Bartonellosis. *Clin. Dermatol.* 27:271–280.
2. Ihler GM. 1996. *Bartonella bacilliformis*: dangerous pathogen slowly emerging from deep background. *FEMS Microbiol. Lett.* 144:1–11.
3. Maguiña C, Gotuzzo E. 2000. Bartonellosis. New and old. *Infect. Dis. Clin. North Am.* 14:1–22, vii. doi:10.1016/S0891-5520(05)70215-4.
4. Langmead B. 2010. Aligning short sequencing reads with bowtie. *Curr. Protoc. Bioinformatics* 11:11.7.
5. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679.
6. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25: 955–964.
7. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100–3108.
8. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75.
9. Pruitt KD, Tatusova T, Brown GR, Maglott DR. 2012. NCBI reference sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* 40:D130–D135.
10. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. 2012. GenBank. *Nucleic Acids Res.* 40:D48–D53.
11. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41.
12. Blom J, Albaum SP, Doppmeier D, Pühler A, Vorhölter FJ, Zakrzewski M, Goesmann A. 2009. EDGAR: a software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinformatics* 10:154.
13. Mitchell SJ, Minnick MF. 1995. Characterization of a two-gene locus from *Bartonella bacilliformis* associated with the ability to invade human erythrocytes. *Infect. Immun.* 63:1552–1562.
14. Li Z, Dumas F, Dubreuil D, Jacques M. 1993. A species-specific periplasmic flagellar protein of *Serpulina* (*Treponema*) *hyodysenteriae*. *J. Bacteriol.* 175:8000–8007.
15. Cloeckaert A, Vizcaino N, Paquet JY, Bowden RA, Elzer PH. 2002. Major outer membrane proteins of *Brucella* spp.: past, present and future. *Vet. Microbiol.* 90:229–247.
16. Commander NJ, Spencer SA, Wren BW, MacMillan AP. 2007. The identification of two protective DNA vaccines from a panel of five plasmid constructs encoding *Brucella melitensis* 16M genes. *Vaccine* 25:43–54.
17. Adler B, Bulach D, Chung J, Doughty S, Hunt M, Rajakumar K, Serrano M, van Zanden A, Zhang Y, Ruffolo C. 1999. Candidate vaccine antigens and genes in *Pasteurella multocida*. *J. Biotechnol.* 73:83–90.
18. Minnick MF, Wilson ZR, Smitherman LS, Samuels DS. 2003. *gyrA* mutations in ciprofloxacin-resistant *Bartonella bacilliformis* strains obtained *in vitro*. *Antimicrob. Agents Chemother.* 47:383–386.