Whole-Genome Sequencing of Two *Bartonella bacilliformis* Strains

Yolanda Guillen,[a] Maria Casadellà,[a,b] Ruth García-de-la-Guarda,[c] Abraham Espinoza-Culupú,[c] Roger Paredes,[a,b,d,e] Joaquim Ruiz,[f]
Marc Noguera-Julian[a,d]

irsiCaixa AIDS Research Institute, Badalona, Catalonia, Spain[a]; Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain[b]; Laboratorio de Microbiología Molecular y Biotecnología, Facultad de Ciencias Biológicas, Universidad Nacional Mayor de San Marcos, Lima, Peru[c]; Universitat de Vic–Universitat Central de Catalunya, Vic, Catalonia, Spain[d]; Unitat VIH, Hospital Universitari Germans Trias i Pujol, Badalona, Catalonia, Spain[e]; ISGlobal, Barcelona Centre for International Health Research (CRESIB), Hospital Clínic, Universitat de Barcelona, Barcelona, Spain[f]

Bartonella bacilliformis is the causative agent of Carrion's disease, a highly endemic human bartonellosis in Peru. We performed a whole-genome assembly of two B. bacilliformis strains isolated from the blood of infected patients in the acute phase of Carrion's disease from the Cusco and Piura regions in Peru.

Address correspondence to Marc Noguera-Julian, mnoguera@irsicaixa.es.

Carrion's disease is a vector-borne neglected illness restricted to Andean valleys in Peru, Ecuador, and Colombia (1), where it remains endemic. The causal agent of Carrion's disease is *Bartonella bacilliformis*, an intracellular, nonfermentative, pleomorphic, Gram-negative microorganism, with *Lutzomyia verrucarum* being the main illness vector.

This illness has two different phases. In the acute phase, called Oroya fever, the intense destruction of erythrocytes leads to a severe hemolytic anemia and temporal immunosuppression, which facilitates both the reactivation of silent illnesses, such as tuberculosis or histoplasmosis, and the growth of opportunistic infections that may be fatal, such as bloodstream *Salmonella* infections (1, 2). In fact, in the preantibiotic era, Oroya fever was considered one of the most lethal infections, having a mortality rate of up to 85% (3); at present, a 10% mortality rate is reported in reference hospitals, even with correct antibiotic treatments (1).

The chronic phase is considered to take place months after Oroya fever, due to the development of partial immunity, but it may also be present even without a previous acute diagnosis. This phase is characterized by the proliferation of verrucous lesions, so-called verruga peruana. In this phase, the most serious complication is the presence of verrucous lesion bleeding, which in extreme cases may require blood transfusions (4).

Often, a clinical cure does not result in microbiological clearance, leading to asymptomatic carriers, with persistent *B. bacilliformis* bacteremia serving as a source for human-to-human transmissions (1). Nonetheless, their real number remains uncertain due to the lack of enough sensitive diagnosis tools (5); additional genomic data may help close these gaps.

Sequencing of the two *B. bacilliformis* strains, USM-LMMB-006 and USM-LMMB-007, collected in 2011 in southern (La Convención, Cusco) and northern Peru (Huancabamba, Piura), respectively, and identified as causal agents of Oroya fever (6), was carried out at the genomics platform of the Germans Trias i Pujol Research Institute using an Illumina MiSeq sequencer with a paired-end 300-bp sequencing kit. Using Trimmomatic (7), low-quality reads were filtered out and adapter sequences were trimmed. Quality-controlled sequence reads were assembled using A5-miseq (8) small genome assembly software (8). Protein-coding genes were predicted using the NCBI Prokaryote Genome Annotation Pipeline (PGAP) (9). The functions of the predicted protein-coding genes were annotated with the Clusters of Orthologous Groups (COG) (10) database using the WebMGA (11) interface with standard parameters. For strains USM-LMMB-006 and USM-LMMB-007, assemblies produced 15 and 11 scaffolds with median depths of coverage of 72× and 77×, total sizes of 1,401,011 and 1,405,613 bp, and GC contents of 38.0% and 38.0%, respectively. The draft genomes of *B. bacilliformis* strains USM-LMMB-006 and USM-LMMB-007 contained 1,136/1,135 coding sequences, 3/3 rRNAs, and 39/39tRNAs, respectively.

Using COG functional assignment, the 996 (88%) and 998 (88%) predicted proteins of strains USM-LMMB-006 and USM-LMMB-007, respectively, could be classified into 868 COG families and 21 COG classes. The most abundant COG classes in both strains were related to translation, ribosomal structure, and biogenesis, including 135 predicted proteins in both strains. The abundant classes in USM-LMMB-006 and USM-LMMB-007 were general function (92/92); replication, recombination, and repair (76/74); amino acid transport and metabolism (76/76); and energy production and conversion (72/72); 88/86 had unknown function.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at NCBI GenBank under the sequence accession numbers LQWW00000000 (LQWW01000001 to LQWW01000015) and LQXX00000000 (LQXX01000001 to LQXX01000011) for strains USM-LMMB-006 and USM-LMMB-007, respectively.

REFERENCES

1. **Minnick MF, Anderson BE, Lima A, Battisti JM, Lawyer PG, Birtles RJ**. 2014. Oroya fever and verruga peruana: bartonelloses unique to South America. PLoS Negl Trop Dis **8**:e2919. http://dx.doi.org/10.1371/journal.pntd.0002919.
2. **Ticona E, Huaroto L, Garcia Y, Vargas L, Madariaga MG**. 2010. The pathophysiology of the acute phase of human bartonellosis resembles AIDS. Med Hypotheses **74**:45–49. http://dx.doi.org/10.1016/j.mehy.2009.06.054.
3. **Ihler GM**. 1996. *Bartonella bacilliformis*: dangerous pathogen slowly emerging from deep background. FEMS Microbiol Lett **144**:1–11. http://dx.doi.org/10.1111/j.1574-6968.1996.tb08501.x.
4. **Pachas P**. (2000). Epidemiología de la Bartonelosis en el Perú. Oficina General de Epidemiología – Instituto Nacional de Salud. Lima (Peru).
5. **Gomes C, Martinez-Puchol S, Pons MJ, Bazán J, Tinco C, Del Valle J, Ruiz J**. 2016. Evaluation of PCR approaches for detection of *Bartonella bacilliformis* in blood samples. PLoS Negl Trop Dis **10**:e0004529. http://dx.doi.org/10.1371/journal.pntd.0004529.
6. **Espinoza-Culupú A, Quispe-Gaspar R, Jaramillo M, Icho M, Eca A, Ramírez P, Alvarado D, Guerrero JC, Vargas-Vásquez F, Córdova O, García-de-la-Guarda R**. 2014. Caracterización molecular de la región determinante de resistencia a quinolonas (QRDR) de la topoisomerasa IV de Bartonella bacilliformis en aislados clínicos. Rev Peru Biol **21**:89–98.
7. **Bolger AM, Lohse M, Usadel B**. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**:2114–2120. http://dx.doi.org/10.1093/bioinformatics/btu170.
8. **Coil D, Jospin G, Darling AE**. 2015. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics **31**:587–589. http://dx.doi.org/10.1093/bioinformatics/btu661.
9. **Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Ciufo S, Li W**. 2013. Prokaryotic genome annotation pipeline. NCBI, Bethesda, MD.
10. **Tatusov RL, Galperin MY, Natale DA, Koonin EV**. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res **28**:33–36. http://dx.doi.org/10.1093/nar/28.1.33.
11. **Wu S, Zhu Z, Fu L, Niu B, Li W**. 2011. WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics **12**:444. http://dx.doi.org/10.1186/1471-2164-12-444.