genomeA
Journals.ASM.org

Whole-Genome Sequencing and Comparative Analysis of *Yersinia pestis*, the Causative Agent of a Plague Outbreak in Northern Peru

O. Cáceres,[a] J. Montenegro,[a] C. Padilla,[a] D. Tarazona,[a] H. Bailón,[a] P. García,[b] M. Céspedes,[b] P. Valencia,[c] H. Guio[a,d]

Laboratorio de Biotecnología y Biología Molecular, Instituto Nacional de Salud, Lima, Peru[a]; Laboratorio de Zoonosis Bacteriana, Instituto Nacional de Salud, Lima, Peru[b]; Centro Nacional de Salud Pública, Instituto Nacional de Salud, Lima, Peru[c]; Asociación Latinoamericana de Biotecnología (ALBIOTEC), Lima, Peru[d]

The plague is a zoonotic disease caused by the bacterium *Yersinia pestis*. Here, we report the complete genome sequence of the *Y. pestis* strain INS, which was isolated from swollen lymph gland aspirate (bubo aspirate) of an infected patient from a pneumonic outbreak in 2010 in northern Peru.

Received 27 December 2012 **Accepted** 17 January 2013 **Published** 28 February 2013

Citation Cáceres O, Montenegro J, Padilla C, Tarazona D, Bailón H, García P, Céspedes M, Valencia P, Guio H. 2013. Whole-genome sequencing and comparative analysis of *Yersinia pestis*, the causative agent of a plague outbreak in northern Peru. Genome Announc. 1(1):e00249-12. doi:10.1128/genomeA.00249-12.

Address correspondence to H. Guio, hguio@ins.gob.pe.

Three biovars of *Yersinia pestis* have been characterized: Antiqua, Medievalis, and Orientalis (1). The Antiqua and Medievalis are ancient strains that were responsible for the plague of Justinian and the Black Death, respectively. Orientalis is responsible for the modern pandemic and has been recognized all around the world. The genomes of several strains of *Yersinia pestis* (CO92, KIM, and others) have been sequenced and annotated.

A total of 5 μg of DNA was sheared using a Covaris ultrasonicator, and a DNA library was prepared using the Fragment library kit according to the manufacturer's instructions (Applied Biosystems). The DNA library was sequenced in the SOLiD 3 Plus genome analyzer, which yielded 250-fold coverage. Bioscope software (Applied Biosystems) was used to correct dibase miscalls prior to assembly. Mapping of the reads was achieved using Bowtie 2.0 (2). The *Y. pestis* strain CO92 (NCBI accession no. AL590842.1) genome sequence was used as a reference for reassembly, providing a total of 18 contigs. Consensus sequence was obtained with SAMtools (3) and Gap5 (4) for further analysis. For the genome annotation, we used Rapid Annotations using Subsystems Technology (RAST) (5), and comparative analysis was done with the fully annotated genome of strain CO92 (6). We used SNP Finder (7) for single-nucleotide polymorphisms (SNP) and small indel detection between the reference and assembled genomes.

Genome-wide comparison of strain INS to several other genomes was performed using BLAST Ring Image Generator (BRIG) (8). The plasmids pCD1 (NCBI accession no. AL117189.1), pMT1 (NCBI accession no. AL117211.1), and pPCP1 (NCBI accession no. AL109969.1) were also used as references for the reassembly of the reads that had not been mapped.

Y. pestis INS is 4.6 Mb long and 99.93% identical to *Y. pestis* CO92. It has an average G+C content of 47.6%. Genome annotation found 4,214 protein-coding sequences, 19 rRNA genes, and 69 tRNA genes. From these, 2,316 were assigned unequivocal functions, mainly in virulence factors, like invasins ($n = 4$), adhesion ($n = 6$), toxins ($n = 17$), invasion and intracellular resistance ($n = 19$), heme and iron uptake ($n = 20$), and resistance to antibiotics ($n = 29$). We found 63 SNPs, none of which affected protein-coding sequences of the genes that were annotated. However, 43 of these 63 SNPs might influence gene expression due to their location within the 5′ or 3′ untranslated regions (UTRs) (9). Additionally, we found three pathogenic plasmids: pCD1, pMT1, and pPCP1. The first one carries a complete type IV secretory system, which increases its capacity for cell invasion and defense neutralization. The second one holds the murine toxin (MT), which allows the bacterium to survive at 37°C. The last one has a plasminogen activator gene that helps to evade the immunological response. Moreover, pCD1 had one deletion of 1 kb and 2 SNPs. This deletion was found next to the IS*100* element, which counts for significant recombination rates through horizontal gene transfer (10). On the other hand, pMT1 had 4 deletions and 2 SNPs. It is not yet clear whether these one-gene deletions correspond to highly mobile genetic elements.

Y. pestis INS has similarities to La Paz strain, mainly in the presence of a multidrug translocase gene (11) that confers increased antibiotic resistance. Our findings confirm the presence of the 1.ORI group in our country and the expansion of the biovar Orientalis (12).

Nucleotide sequence accession numbers. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. AMQL00000000. The version described in this paper is the first version, AMQL01000000.

ACKNOWLEDGMENTS

This work was supported by the Peruvian National Institute of Health. H. Guio also thanks the Science and Technology Program (FINCyT) and CONCyTEC.

REFERENCES

1. **Deng W, Burland V, Plunkett G, III, Boutin A, Mayhew GF, Liss P, Perna NT, Rose DJ, Mau B, Zhou S, Schwartz DC, Fetherston JD, Lindler LE, Brubaker RR, Plano GV, Straley SC, McDonough KA, Nilles ML, Matson JS, Blattner FR, Perry RD.** 2002. Genome sequence of *Yersinia pestis* KIM. J. Bacteriol. **184**:4601–4611.
2. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods **9**:357–359.

3. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup.** 2009. The sequence alignment/map format and SAMtools. Bioinformatics **25**:2078–2079.
4. **Bonfield JK, Smith K, Staden R.** 1995. A new DNA sequence assembly program. Nucleic Acids Res. **23**:4992–4999.
5. **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.
6. **Parkhill J, Wren BW, Thomson NR, Titball RW, Holden MT, Prentice MB, Sebaihia M, James KD, Churcher C, Mungall KL, Baker S, Basham D, Bentley SD, Brooks K, Cerdeño-Tárraga AM, Chillingworth T, Cronin A, Davies RM, Davis P, Dougan G, Feltwell T, Hamlin N, Holroyd S, Jagels K, Karlyshev AV, Leather S, Moule S, Oyston PC, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG.** 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. Nature **413**:523–527.
7. **Song J, Xu Y, White S, Miller KW, Wolinsky M.** 2005. SNPsFinder—a web-based application for genome-wide discovery of single nucleotide polymorphisms in microbial genomes. Bioinformatics **21**:2083–2084.
8. **Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA.** 2011. BLAST ring image generator (BRIG): simple prokaryote genome comparisons. BMC Genomics **12**:402.
9. **Kawano M, Reynolds AA, Miranda-Rios J, Storz G.** 2005. Detection of 5′- and 3′-UTR-derived small RNAs and cis-encoded antisense RNAs in *Escherichia coli*. Nucleic Acids Res. **33**:1040–1050.
10. **Motin VL, Georgescu AM, Elliott JM, Hu P, Worsham PL, Ott LL, Slezak TR, Sokhansanj BA, Regala WM, Brubaker RR, Garcia E.** 2002. Genetic variability of *Yersinia pestis* isolates as predicted by PCR-based IS100 genotyping and analysis of structural genes encoding glycerol-3-phosphate dehydrogenase (glpD). J. Bacteriol. **184**:1019–1027.
11. **Cummings CA, Bormann Chung CA, Fang R, Barker M, Brzoska P, Williamson PC, Beaudry J, Matthews M, Schupp J, Wagner DM, Birdsell D, Vogler AJ, Furtado MR, Keim P, Budowle B.** 2010. Accurate, rapid and high-throughput detection of strain-specific polymorphisms in *Bacillus anthracis* and *Yersinia pestis* by next-generation sequencing. Investig. Genet. **1**:5.
12. **Morelli G, Song Y, Mazzoni CJ, Eppinger M, Roumagnac P, Wagner DM, Feldkamp M, Kusecek B, Vogler AJ, Li Y, Cui Y, Thomson NR, Jombart T, Leblois R, Lichtner P, Rahalison L, Petersen JM, Balloux F, Keim P, Wirth T, Ravel J, Yang R, Carniel E, Achtman M.** 2010. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. Nat. Genet. **42**:1140–1143.