

Revised Genome Sequence of *Brucella suis* 1330

Hongseok Tae,¹ Shamira Shallom,¹ Robert Settlege,¹ Dale Preston,²
L. Garry Adams,³ and Harold R. Garner^{1*}

Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, Virginia¹; Texas Animal Health Commission, State-Federal Diagnostic Laboratory, Austin, Texas²; and Department of Veterinary Pathobiology, College of Veterinary Medicine, Texas A&M University, College Station, Texas³

Received 13 September 2011/Accepted 14 September 2011

***Brucella suis* is a causative agent of porcine brucellosis. We report the resequencing of the original sample upon which the published sequence of *Brucella suis* 1330 is based and describe the differences between the published assembly and our assembly at 12 loci.**

Brucella species are Gram-negative pathogenic bacteria causing zoonotic disease and have been designated category B priority pathogens (1). To fully enable the *Brucella* research community, it is important to construct genome sequence assemblies of the highest quality, since they are the foundation for developing detection and monitoring assays and understanding the biology that may lead to mitigation of *Brucella* bacteria as both potential bioterrorism threats and infectious agents. Hereinafter, we report the resequencing of the 2.1-Mbp chromosome 1 and 1.2-Mbp chromosome 2 of *Brucella suis* 1330, which were fully sequenced in 2002 (5) and used as a reference for previous studies (2, 7), and we describe the sequence differences between the published assembly and our assembly.

The original *Brucella suis* 1330 sample was obtained from BEI Resources (<http://www.beiresources.org>). Genomic DNA was sequenced with an Illumina GAIx sequencer with a 101-cycle paired-end protocol generating about 26,000,000 sequencing read pairs (52,000,000 reads). We trimmed all low-quality bases (<0.99 quality score) from the sequencing reads and used BWA (4) to map the reads to the published reference sequence, resulting in an average 1,559-times sequence coverage. Contig sequences, assembled by two *de novo* assemblers, Abyss (6) and CLCbio Genomics Workbench, were also aligned to the sequence by BWA-SW (3) to search for long indels which were not detected by mapping of sequencing reads.

Through this hybrid approach of mapping and assembly, we identified with a very high confidence a total of 12 sequence differences, including 10 indels and two substitutions between the original and revised reference sequences. These sequence differences in the revised sequence are consistent with other reference sequences of completely sequenced *Brucella* species, including *Brucella abortus* biovar 1, *Brucella canis* ATCC 23365, *Brucella melitensis* 16 M, *Brucella microti* CCM 4915, *Brucella ovis* ATCC 25840, and *Brucella suis* ATCC 23445, suggesting that they are more likely to represent assembly errors in the original reference sequence than mutation of the original sample. Among them, six indels caused frameshifts of protein-coding loci, including BS1330_I0247, BS1330_I1176, BS1330_I1348, BS1330_I1367,

BS1330_I2082, and BS1330_II0838, of which BS1330_I2082 is annotated as a peptidyl-tRNA hydrolase domain protein while the other loci are annotated as hypothetical proteins (BR1180, BR1353, BR1372, BR1373, and BR2088 of the original annotation were excluded due to the frameshifts). Only three indels were detected at the intergenic regions, and two of them were in 8-mer tandem repeat loci. The other sequence differences include a 3-base insertion in BS1330_I2121 (BR2127) and two nonsynonymous substitutions in BS1330_I0614 (BR0618) and BS1330_I2082 (BR2088).

The quality of sequence analysis using a resequencing strategy for an organism (or new isolate) relies on the condition of the reference sequence that is thought to represent the general genome sequence of the strain to which it belongs. In this study, we revised the genome sequence of the original *Brucella suis* 1330 sample to improve the accuracy of downstream studies.

Nucleotide sequence accession numbers. The revised genome sequences of *B. suis* 1330 are available in GenBank under accession numbers CP002997 and CP002998.

This work was funded by the Department of Homeland Security through FAZD Center Grant 570636 (National Center of Excellence for Foreign Animal and Zoonotic Disease Defense) at Texas A&M University and by the Virginia Bioinformatics Institute director's funds.

B. suis 1330 genomic DNA was kindly provided by BEI resources and sequenced by the Virginia Bioinformatics Institute's Core Laboratory Facility.

REFERENCES

1. Glynn, M. K., and T. V. Lynn. 2008. Brucellosis. *J. Am. Vet. Med. Assoc.* **223**:900–908.
2. Lavigne, J.-P., A. C. Vergunst, G. Bourg, and D. O'Callaghan. 2005. The IncP island in the genome of *Brucella suis* 1330 was acquired by site-specific integration. *Infect. Immun.* **73**:7779–7783.
3. Li, H., and R. Durbin. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**:589–595.
4. Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760.
5. Paulsen, I. T., et al. 2002. The *Brucella suis* genome reveals fundamental similarities between animal and plant pathogens and symbionts. *Proc. Natl. Acad. Sci. U. S. A.* **99**:13148–13153.
6. Simpson, J. T., et al. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res.* **19**:1117–1123.
7. Wattam, A. R., et al. 2009. Analysis of ten *Brucella* genomes reveals evidence for horizontal gene transfer despite a preferred intracellular lifestyle. *J. Bacteriol.* **191**:3569–3579.

* Corresponding author. Mailing address: Virginia Bioinformatics Institute, Virginia Tech, Washington Street, MC0477, Blacksburg, VA 24061-0477. Phone: (540) 231-2582. Fax: (540) 231-2606. E-mail: garner@vbi.vt.edu.