JOURNAL OF BACTERIOLOGY, Mar. 2009, p. 1725 Vol. 191, No. 5
0021-9193/09/$08.00+0 doi:10.1128/JB.01731-08

GENOME ANNOUNCEMENT

Genome Sequence of the *Wolbachia* Endosymbiont of *Culex quinquefasciatus* JHB▽

Steven L. Salzberg,[1]* Daniela Puiu,[1] Daniel D. Sommer,[1] Vish Nene,[2] and Norman H. Lee[3]

Center for Bioinformatics and Computational Biology, University of Maryland, College Park, Maryland 20742[1]; Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland 21201[2]; and Department of Pharmacology and Physiology, George Washington University Medical Center, Washington, DC 20037[3]

Received 10 December 2008/Accepted 18 December 2008

Wolbachia **species are endosymbionts of a wide range of invertebrates, including mosquitoes, fruit flies, and nematodes. The *w*Pip strains can cause cytoplasmic incompatibility in some strains of the *Culex* mosquito. Here we describe the genome sequence of a *Wolbachia* strain that was discovered in the whole-genome sequencing data for the mosquito *Culex quinquefasciatus* strain JHB.**

The *w*Pip *Culex quinquefasciatus* strain JHB genome was sequenced as a serendipitous artifact of the *Culex* sequencing project, just as several previous *Wolbachia* species were found in *Drosophila* sequencing data (5). We separated the bacterial data from the mosquito data and assembled a virtually complete genome of the endosymbiont. The final assembly contains 21 contiguous DNA segments containing 1,542,137 nucleotides. Our annotation found 1,378 protein-coding genes, 34 tRNA genes, and one rRNA operon.

We used all the raw sequences ("reads") from the *Culex quinquefasciatus* genome (Johannesburg strain; JHB) sequencing project as of September 2007. These comprised 7,379,314 reads, all generated by capillary sequencing. As a reference genome, we used the *Wolbachia* endosymbiont of *Culex quinquefasciatus* Pel that was recently sequenced (GenBank accession no. AY072044) (2). We aligned all whole-genome shotgun reads to the reference genome using MUMmer (minimum match of 65 bp at 80% identity) (3) in order to extract reads of bacterial origin. For each bacterial read, we also extracted its mate from the paired-end trace data. The process yielded 36,767 reads, of which 35,750 were paired.

We assembled the reads with the Celera Assembler (4) using default parameters except for the unitig error rate, which was set to 2%. The assembler generated 16 scaffolds containing 21 contigs and 92 additional contigs flagged as low quality by the assembler. We ran further analyses to eliminate nonbacterial contigs and to eliminate near-identical contigs that appeared to represent population variants within the original DNA sample, as were also observed by Klasson et al. (2). The final assembly contains 21 contigs (the largest was 478,325 bp) in a single scaffold, with an average depth of coverage of 14×. Scaffolding utilized both the original mate-pair constraints and the organization of the *w*Pip reference genome.

The JHB strain is very similar to the previously sequenced Pel strain; however, there are several areas of differences. JHB contains four unique regions not found in Pel, with lengths of 524, 565, 640, and 524 bp. One of these regions contains the gene for the DNA repair protein Radc, and another contains a transcriptional regulator gene, both with homologs in *Wolbachia dMel*. The *w*Pip Pel genome contains two regions not found in strain JHB, which are 856 bp and 120 bp in length.

There are 10 large-scale rearrangements distinguishing the genomes, in which large segments of the chromosome have been inverted. Some of these are associated with IS elements from the IS*256* family. The greater number of protein-coding genes in *w*Pip JHB—1,378 versus 1,248 in *w*Pip Pel—appears to be due to differences in annotation methods, although both projects used the Glimmer (1) gene finder.

Nucleotide sequence accession number. The genome sequence of the *Wolbachia* endosymbiont of *Culex quinquefasciatus* JHB has been deposited at GenBank/EMBL/DDBJ under accession no. ABZA00000000.

The genome of *Culex quinquefasciatus* JHB was sequenced jointly by The Broad Institute and the J. Craig Venter Institute with funding provided by the National Institute of Allergy and Infectious Diseases. S.L.S., D.P., and D.D.S. were supported in part by NIH grant no. R01-LM006845.

REFERENCES

1. **Delcher, A. L., K. A. Bratke, E. C. Powers, and S. L. Salzberg.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23:**673–679.
2. **Klasson, L., T. Walker, M. Sebaihia, M. J. Sanders, M. A. Quail, A. Lord, S. Sanders, J. Earl, S. L. O'Neill, N. Thomson, S. P. Sinkins, and J. Parkhill.** 2008. Genome evolution of *Wolbachia* strain *w*Pip from the *Culex pipiens* group. Mol. Biol. Evol. **25:**1877–1887.
3. **Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, and S. L. Salzberg.** 2004. Versatile and open software for comparing large genomes. Genome Biol. **5:**R12.
4. **Myers, E. W., G. G. Sutton, A. L. Delcher, I. M. Dew, D. P. Fasulo, M. J. Flanigan, S. A. Kravitz, C. M. Mobarry, K. H. Reinert, K. A. Remington, E. L. Anson, R. A. Bolanos, H. H. Chou, C. M. Jordan, A. L. Halpern, S. Lonardi, E. M. Beasley, R. C. Brandon, L. Chen, P. J. Dunn, Z. Lai, Y. Liang, D. R. Nusskern, M. Zhan, Q. Zhang, X. Zheng, G. M. Rubin, M. D. Adams, and J. C. Venter.** 2000. A whole-genome assembly of *Drosophila*. Science **287:**2196–2204.
5. **Salzberg, S. L., J. C. Hotopp, A. L. Delcher, M. Pop, D. R. Smith, M. B. Eisen, and W. C. Nelson.** 2005. Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. Genome Biol. **6:**R23.

* Corresponding author. Mailing address: Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD 20742. Phone: (301) 405-9611. Fax: (301) 314-1341. E-mail: salzberg@umiacs.umd.edu.

▽ Published ahead of print on 29 December 2008.