

Complete Genome Sequence of *Yersinia enterocolitica* subsp. *palaearctica* Serogroup O:3[∇]

Julia Batzilla,¹ Dirk Höper,² Uladzimir Antonenka,¹ Jürgen Heesemann,¹ and Alexander Rakin^{1*}

Max von Pettenkofer-Institute, LMU, Munich, Germany,¹ and Friedrich-Loeffler-Institute, Bundesforschungsanstalt für Tiergesundheit, Riems, Germany²

Received 9 December 2010/Accepted 27 January 2011

We report here the first finished and annotated genome sequence of a representative of the most epidemiologically successful *Yersinia* group, *Y. enterocolitica* subsp. *palaearctica* strain Y11, serotype O:3, biotype 4. This strain is a certified type strain of the German DSMZ collection (DSM no. 13030; *Yersinia enterocolitica* subsp. *palaearctica*) that was isolated from the stool of a human patient (H. Neubauer, S. Aleksic, A. Hensel, E. J. Finke, and H. Meyer. *Int. J. Med. Microbiol.* 290:61–64, 2000).

Yersinia enterocolitica is a gastrointestinal pathogen with distinctive clinical manifestations and a predilection for young children (2). It is a member of the *Enterobacteriaceae*, genus *Yersinia*, which includes three pathogenic species: *Y. pestis*, the agent of plague, and two enteropathogens, *Y. pseudotuberculosis* and *Y. enterocolitica*. *Y. enterocolitica* strains are differentiated by bio- and serotyping. Biotype 1A (BT1A) strains are mostly nonpathogenic, in contrast to the highly pathogenic BT1B (predominant in the United States) and low-pathogenic BT2 to -5 (dominating worldwide). Recently, *Y. enterocolitica* was divided into two subspecies (3): *enterocolitica*, proposed for strains with the 16S rRNA type of American origin, and *palaearctica*, proposed for strains of European origin (BT1A and BT2 to -5).

Yersiniosis is the third-most-common bacterial enteric disease in Europe, and *Yersinia enterocolitica* subsp. *palaearctica* serobiootype O:3/4 is most frequently isolated from humans and slaughter pigs (4). Thus, deciphering the *Yersinia enterocolitica* subsp. *palaearctica* O:3/4 genome will help to uncover the mechanisms of its successful worldwide dissemination.

Y. enterocolitica subspecies *palaearctica* Y11 (DSMZ no. 13030), an O:3/4 human isolate, was selected for sequencing. The genome was determined using MegaBACE (at Integrated Genomics, Jena, Germany), a 454 Genome Sequencer (GS) 20 (at 454, Branford CT), and a 454 GS FLX Titanium (at Seq-IT, Kaiserslautern, Germany). MegaBACE sequencing produced 73,848 reads with 64,366,908 bp. The 454 GS 20 run yielded 927,998 reads with 94,454,598 bp, and the 454 GS FLX Titanium run added 240,813 reads with 105,539,453 bp. These datasets were assembled into a draft genome of 103 contigs with a total length of 4,464,482 bp using DNASTar Lasergene software (version 7.2.0; Seqman Pro). Gaps were closed manually, in cooperation with LGC Genomics (Berlin, Germany),

by Sanger sequencing of PCR products. Finally, data were assembled into a complete genome sequence of 4,553,420 bp for the genome and an additional 72,460-bp contig representing the pYVO3 plasmid. For quality control, all raw data were mapped along the two contigs using 454 Reference Mapper software (version 2.3; Roche). The median sequence depth for the genome contig was 37 (1st quartile, 30; 3rd quartile, 48), and the proportion of Q40+ bases was 99.58%. For the plasmid sequence, a median depth of 12 (1st quartile, 8; 3rd quartile, 17) was achieved, with the proportion of Q40+ bases being 98.01%. The average GC content is 47%, similar to that of *Y. enterocolitica* strain 8081 (NC_008800; 47.27%). The genome sequence was annotated using the RAST server (<http://rast.nmpdr.org/>) (1). This first deciphered *Y. enterocolitica* subsp. *palaearctica* genome reflects multiple bacterial interactions and points to the structures and functions potentially involved in virulence attenuation and host adaptation.

Nucleotide sequence accession numbers. The complete Y11 genome sequence was deposited in EMBL under accession numbers FR729477 and FR745874 (plasmid).

This genome sequencing project was supported by funding from the BMBF Network FBI-Zoo (grant 01KI07122).

We are grateful for assistance from Lars Paulin in the first genome assembly projects using the Staden package. We are also thankful to Thomas Rattei, Thomas Weinmaier, and Patrick Tischler for frequent bioinformatic advice and support of the earlier Y11 draft genome with, among others, PEDANT database access and SIMAP calculations. We also appreciate the work of our former colleague A. Golubov on this genome project, as well as the helpful discussions with H. Neubauer.

REFERENCES

1. Aziz, R. K., et al. 2008. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9:75.
2. Koehler, K. M., et al. 2006. Population-based incidence of infection with selected bacterial enteric pathogens in children younger than five years of age, 1996-1998. *Pediatr. Infect. Dis. J.* Feb. 25:129-134.
3. Neubauer, H., S. Aleksic, A. Hensel, E. J. Finke, and H. Meyer. 2000. *Yersinia enterocolitica* 16S rRNA gene types belong to the same genospecies but form three homology groups. *Int. J. Med. Microbiol.* 290:61-64.
4. Ortiz Martínez, P., et al. 2010. Wide variety of bioserotypes of enteropathogenic *Yersinia* in tonsils of English pigs at slaughter. *Int. J. Food Microbiol.* 139:64-69.

* Corresponding author. Mailing address: Max von Pettenkofer-Institute, LMU, Munich, Germany. Phone: 498951605275. Fax: 498951605223. E-mail: rakin@mvp.uni-muenchen.de.

[∇] Published ahead of print on 4 February 2011.