

Complete Genome Sequence of Pigmentation Negative *Yersinia pestis* strain Cadman

Sean Lovett^a, Kitty Chase^b, Galina Koroleva^a, Gustavo Palacios^a, David Rozak^{b#}, Jason T.
Ladner^a

Center for Genome Sciences^a, Diagnostic Systems Division^b, U.S. Army Medical Research
Institute of Infectious Diseases, Fort Detrick, Frederick, Maryland, USA

Running head: Complete Genome Sequence of *Y. pestis* strain Cadman

Address correspondence to David Rozak, david.a.rozak2.ctr@mail.mil

Abstract

Here we report the genome sequence of *Yersinia pestis* strain Cadman, an attenuated strain lacking the *pgm* locus. *Y. pestis* is the causative agent of plague and generally must be worked with under BSL-3 conditions. However, strains lacking the *pgm* locus are considered safe to work with under BSL-2 conditions.

Genome Announcement

Yersinia pestis is the causative agent of plague. Due to its high potential for use in biological terrorism, the CDC has classified it a category A pathogen, which must be handled within a biosafety level 3 laboratory.

Here we report the complete sequence of the pigmentation negative (*pgm*-) Cadman strain of *Y. pestis*. This strain was isolated in 1965 from cerebrospinal fluid obtained from a 2½-year-old boy(1). It is unknown whether the original isolate was *pgm*- or whether this mutation occurred during laboratory passage. However, naturally occurring *pgm*- isolates have been associated with disease in humans and rodents (2). The virulence of *pgm*- strains of *Y. pestis* is attenuated (3) and therefore they can be manipulated under less restrictive laboratory conditions.

Chromosomal DNA was prepared using the Qiagen EZ1 Advanced XL system with the EZ1 DNA Tissue kit according to manufacturer's instructions. The sequencing library was prepared using the SMRTbell™ Template Prep Kit (Pacific Biosciences, Menlo Park, CA) following manufacturer's protocol. The resulting SMRTbell template was size selected on BluePippin system (Sage Science, Beverly, MA) using 0.75% dye-free agarose cassette with 4-10kb Hi-Pass

protocol and low cut set at 4 kb. Size selected template was cleaned and concentrated with AMPure PB beads. The P4 polymerase was used with the C2 sequencing kit to generate 240-minute movies.

Raw sequencing reads were filtered for a minimum subread length of 1000bp, minimum polymerase read length of 5000bp, and minimum quality of 0.80. The filtered dataset contained 101,466 reads with a mean subread length of 8,409bp.

Reads were assembled into four contigs using HGAP3 v2.3 (4). These contigs correspond to the chromosome and plasmids pCD1, pPCP1, and pMT1. Each contig was checked for redundant sequence at the ends using Gepard v1.3(5). The plasmids were determined to be complete circular molecules, so redundant sequences at the ends were trimmed to one copy and a new breakpoint was chosen for their linear representations. A large (~50kb) repeat on the chromosome was not able to be resolved, so the chromosome was not modified. Reads were then re-aligned to the trimmed and shifted draft assembly for correction using Quiver (4).

Annotation with NCBI's Prokaryotic Genome Annotation Pipeline v3.3 (6) identified 4,238 coding regions, 19 ribosomal RNAs, 68 tRNAs, 12 ncRNAs, and 2 CRISPR arrays.

Gene content was compared to *Y. pestis* CO92 (NC_003131, NC_003132, NC_003134, NC_003143). Nucleotide sequences of annotated genes were clustered at 90% identity using USearch (7). Genes without assigned orthologs were used as queries in a BLAST search against both genomes. Genes with $\geq 90\%$ nucleotide identity to a region of the CO92 genome were considered to be present, even if unannotated. All annotated genes outside of the *pgm* locus in

CO92 were present in Cadman, although some exhibit alterations that may result in loss of function. No genes present in Cadman were absent in CO92. Absence of the *pgm* locus was confirmed by aligning all reads to the *pgm*+ CO92 using BWA-mem v0.7.12(8). No significant alignments to sequence unique to the *pgm* locus were observed.

Nucleotide sequence accession numbers. Complete sequences for all molecules have been deposited at Genbank under the accession numbers CP016273 to CP016276.

Acknowledgements. This work was supported by the Defense Threat Reduction Agency, project CB10219. Opinions, interpretations, conclusions, and recommendations are those of the authors and do not necessarily reflect the official policy or position of the US Army, US Department of Defense, nor the US Government.

References

1. **Martin AR, Hurtado FP, Plessala RA, Hurtado EG, Chapman CE, Callahan EL, Brutsche RL.** 1967. Plague meningitis. A report of three cases in children and review of the problem. *Pediatrics* **40**:610-616.
2. **Williams JE, Harrison DN, Quan TJ, Mullins JL, Barnes AM, Cavanaugh DC.** 1978. Atypical plague bacilli isolated from rodents, fleas, and man. *Am J Public Health* **68**:262-264.
3. **Burrows TW, Jackson S.** 1956. The virulence-enhancing effect of iron on nonpigmented mutants of virulent strains of *Pasteurella pestis*. *Br J Exp Pathol* **37**:577-583.

93 4. **Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A,**
94 **Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J.** 2013. Nonhybrid,
95 finished microbial genome assemblies from long-read SMRT sequencing data. Nat
96 Methods **10**:563-569.

97 5. **Krumsiek J, Arnold R, Rattei T.** 2007. Gepard: a rapid and sensitive tool for creating
98 dotplots on genome scale. Bioinformatics **23**:1026-1028.

99 6. **Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L,**
100 **Lomsadze A, Pruitt KD, Borodovsky M, Ostell J.** 2016. NCBI prokaryotic genome
101 annotation pipeline. Nucleic Acids Res doi:10.1093/nar/gkw569.

102 7. **Edgar RC.** 2010. Search and clustering orders of magnitude faster than BLAST.
103 Bioinformatics **26**:2460-2461.

104 8. **Li H, Durbin R.** 2009. Fast and accurate short read alignment with Burrows-Wheeler
105 transform. Bioinformatics **25**:1754-1760.