Draft Genome Assemblies of *Proteus mirabilis* ATCC 7002 and *Proteus vulgaris* ATCC 49132

T. D. Minogue,ᵃ H. E. Daligault,ᵇ K. W. Davenport,ᵇ K. A. Bishop-Lilly,ᶜ,ᵈ D. C. Bruce,ᵇ P. S. Chain,ᵇ S. R. Coyne,ᵃ O. Chertkov,ᵇ T. Freitas,ᵇ* K. G. Frey,ᶜ,ᵈ J. Jaissle,ᵇ G. I. Koroleva,ᵉ J. T. Ladner,ᵉ G. F. Palacios,ᵉ C. L. Redden,ᶜ,ᵈ Y. Xu,ᵇ S. L. Johnsonᵇ

Diagnostic Systems Division (DSD), United States Army Medical Research of Infectious Disease (USAMRIID), Fort Detrick, Maryland, USAᵃ; Los Alamos National Laboratory (LANL), Los Alamos, New Mexico, USAᵇ; Naval Medical Research Center (NMRC)-Frederick, Fort Detrick, Maryland, USAᶜ; Henry M. Jackson Foundation, Bethesda, Maryland, USAᵈ; Center for Genome Sciences (CGS), USAMRIID, Fort Detrick, Maryland, USAᵉ

* Present address: T. Freitas, Med Fusion, Lewisville, Texas, USA.

The pleomorphic swarming bacilli of the genus *Proteus* are common human gut commensal organisms but also the causative agents of recurrent urinary tract infections and bacteremia. We sequenced and assembled the 3.99-Mbp genome of *Proteus mirabilis* ATCC 7002 (accession no. JOVJ00000000) and the 3.97-Mbp genome of *Proteus vulgaris* ATCC 49132 (accession no. JPIX00000000), both of which are commonly used reference strains.

As members of the *Enterobacteriaceae* (Gram-negative bacilli), bacteria of the genus *Proteus* are commonly found in the human intestinal tract and are associated with opportunistic infections (1). The genus consists of four species (*Proteus mirabilis*, *Proteus myxofaciens*, *Proteus penneri*, and *Proteus vulgaris*), three of which cause human disease (*P. myxofaciens* is an insect pathogen) (1). Members of the genus (primarily *P. mirabilis* but also *P. penneri* and *P. vulgaris*) are the third most common causes of complicated urinary tract infections in the United States (1, 2) and are notable for their swarming abilities, which are directly linked to their ability to initiate disease (2, 3). Here, we sequenced two isolates, both originally isolated from clinical samples and both commonly used in quality control testing: *P. mirabilis* ATCC 7002 and *P. vulgaris* ATCC 49132.

High-quality genomic DNA was extracted from purified isolates of each strain using the Qiagen Genomic-tip 500 at the U.S. Army Medical Research Institute of Infectious Diseases-Diagnostic Systems Division (USAMRIID-DSD). Specifically, 100-ml bacterial cultures were grown to stationary phase and nucleic acid extracted as per the manufacturer's recommendations. The draft genomes of *P. mirabilis* ATCC 7002 and *P. vulgaris* ATCC 49132 included a combination of Illumina (4) and 454 technologies (5). For each genome, we constructed and sequenced a 100-bp Illumina library (*P. mirabilis*, 307-fold genome coverage; *P. vulgaris*, 300-fold genome coverage) and a long-insert paired-end 454 library (*P. mirabilis*, 8,695.7 ± 2,174-bp insert and 24-fold genome coverage; *P. vulgaris*, 8,041 ± 2,010-bp insert and 28-fold genome coverage). The 454 paired-end data were assembled in Newbler (6), and those consensus sequences were computationally shredded into 2-kbp overlapping fake reads (shreds). The Illumina sequencing data were assembled with Velvet (6), and the consensus sequences were computationally shredded into 1.5-kbp overlapping shreds. All data were additionally assembled in AllPaths (7), and the consensus sequences were computationally shredded into 5-kbp overlapping shreds. We integrated the 454 Newbler consensus shreds, Illumina Velvet consensus shreds, AllPaths consensus shreds, and the 454 paired-end library read pairs using parallel Phrap version (High Performance Software, LLC). Possible misassemblies were corrected using in-house scripts and manual editing in Consed (8–10).

Automatic annotation for each genome utilized an Ergatis-based (11) workflow at the Los Alamos National Laboratory (LANL), with minor manual curation. The *P. mirabilis* ATCC 7002 genome (accession no. JOVJ00000000) is 3,992,612 bp (38.8% G+C content) and contains 3,631 coding sequences (CDSs), 11 rRNAs, and 77 tRNAs in 15 contigs. The *P. vulgaris* ATCC 49132 genome (accession no. JPIX00000000) is 3,972,483 bp (37.9% G+C content) and contains 3,592 CDSs, 12 rRNAs, and 77 tRNA in 12 contigs. Each strain contains at least nine of the genes putatively implicated in swarming (2).

Nucleotide sequence accession numbers. The NCBI accession no. for *P. mirabilis* ATCC 7002 is JOVJ00000000, and for *P. vulgaris* ATCC 49132, it is JPIX00000000.

REFERENCES

1. **Rózalski A, Sidorczyk Z, Kotełko K.** 1997. Potential virulence factors of *Proteus* bacilli. Microbiol. Mol. Biol. Rev. **61**:65–89.

2. **Rather PN.** 2005. Swarmer cell differentiation in *Proteus mirabilis*. Environ. Microbiol. **7:**1065–1073. http://dx.doi.org/10.1111/j.1462-2920.2005.00806.x.

3. **Mobley HLT, Belas R.** 1995. Swarming and pathogenicity of *Proteus mirabilis* in the urinary tract. Trends Microbiol. **3:**280–284. http://dx.doi.org/10.1016/S0966-842X(00)88945-3.

4. **Bennett S.** 2004. Solexa Ltd. Pharmacogenomics **5:**433–438. http://dx.doi.org/10.1517/14622416.5.4.433.

5. **Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM, Rothberg JM.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376–380. http://dx.doi.org/10.1038/nature03959.

6. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.

7. **Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, Nusbaum C, Jaffe DB.** 2008. ALLPATHS: *de novo* assembly of whole-genome shotgun microreads. Genome Res. **18:**810–820. http://dx.doi.org/10.1101/gr.7337908.

8. **Ewing B, Green P.** 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. Genome Res. **8:**186–194.

9. **Ewing B, Hillier L, Wendl MC, Green P.** 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res. **8:**175–185. http://dx.doi.org/10.1101/gr.8.3.175.

10. **Gordon D, Abajian C, Green P.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195–202. http://dx.doi.org/10.1101/gr.8.3.195.

11. **Hemmerich C, Buechlein A, Podicheti R, Revanna KV, Dong Q.** 2010. An Ergatis-based prokaryotic genome annotation Web server. Bioinformatics **26:**1122–1124. http://dx.doi.org/10.1093/bioinformatics/btq090.