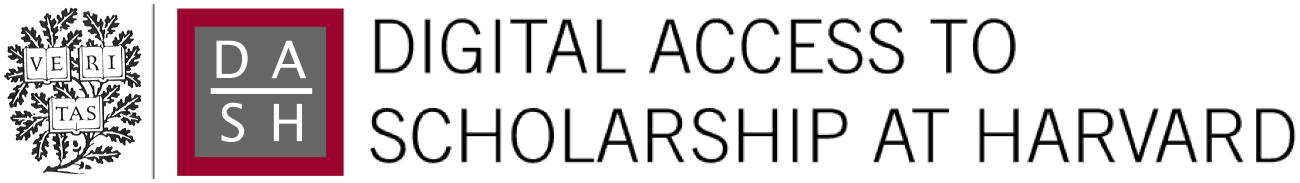

The Complete Genome Sequence of Proteus mirabilis Strain BB2000 Reveals Differences from the P. mirabilis Reference Strain

The complete genome sequence of strain BB2000 reveals differences from the *Proteus mirabilis* reference strain

Nora L. Sullivan[1], Alecia N. Septer, Andrew T. Fields, Larissa M. Wenren, and Karine A. Gibbs*

Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138

[1]current address: W.M. Keck Science Department, 925 N. Mills Ave, Claremont McKenna, Pitzer and Scripps Colleges, Claremont, CA 91711

*Corresponding author:
Karine A. Gibbs, Ph.D.
Department of Molecular and Cellular Biology
Harvard University
16 Divinity Avenue
Cambridge, MA 02138
Em: *kagibbs@mcb.harvard.edu*

Running title: Complete genome sequence of *P. mirabilis* BB2000

Abstract

We announce the complete genome for *Proteus mirabilis* strain BB2000, a model system for self recognition. This opportunistic pathogen contains a single, circular chromosome (3,846,754 base pairs). Comparisons between this genome and that of strain HI4320 reveal genetic variations corresponding to previously unknown physiological and self-recognition differences.

30 The gut commensal bacterium *Proteus mirabilis* is the primary cause of urinary
31 tract infections in patients with long-term indwelling catheters (1-4). Interestingly,
32 migrating colonies of *P. mirabilis* cells can distinguish self from non-self: a visible
33 boundary forms at the interface between two genetically distinct colonies, while two
34 genetically identical populations merge together (5). The genetic determinants of this
35 self-recognition behavior, first identified in *P. mirabilis* strain BB2000, included self-
36 identity genes containing numerous inter-strain nucleotide polymorphisms and suggested
37 that additional genetic differences between strains are likely (6). To date, only the
38 genome of *P. mirabilis* strain HI4320 (NCBI NC_010554) has been completed (7). Here
39 we report a second closed genome, that of the genetically distinct strain, BB2000 (8).
40 BB2000 genomic DNA was isolated and sequenced using standard protocols.
41 Briefly, DNA was isolated from cells cultured in modified LB broth using
42 phenol/chloroform extraction and ethanol (9). Beckman Coulter Genomics (Danvers,
43 MA) performed initial library preparation and sequencing using the Roche 454 platform.
44 Illumina sequencing was used to confirm the 454 data and resolve stretches of unknown
45 nucleotides; genomic DNA libraries were prepared according to the Illumina
46 Multiplexing Sample Preparation protocol and sequenced by Harvard FAS Systems
47 Biology Core using an Illumina HiSeq 2000. Illumina reads were assembled onto the 454
48 genomic data using Galaxy software (10). Genome closure was accomplished by
49 amplifying across gaps using polymerase chain reactions followed by Sanger sequencing
50 performed by Genewiz Corporation (South Plainfield, NJ).
51 The *P. mirabilis* BB2000 genome consists of a single chromosome (3,846,754
52 base pairs) with 38.6% G+C content. Potential coding sequences (CDSs) were identified

using the xBase annotation service, which predicted CDS regions using Glimmer (11), and assigned predicted protein products based on a direct comparison to the *P. mirabilis* HI4320 genome (12-16). CDSs absent in the HI4320 genome were assigned "hypothetical protein" as the predicted product. Twenty-eight genes related to self-recognition (6, 17) were annotated manually using blastx (12) and the HMMER web interface (18). Sequence assembly and annotation were completed using Artemis software (19). The BB2000 genome encodes 3,457 potential CDSs, of which 2,592 are assigned a putative function; the remaining 865 CDSs are classified as hypothetical proteins, with an additional 81 tRNA genes and 22 rRNA genes.

Comparison of the BB2000 genome to that of strain HI4320 (7) revealed 93% similarity between the chromosomes. The CDSs unique to each genome include genes related to phage, toxin elements, and self recognition. The HI4320 genome encodes iron acquisition proteins that are absent in BB2000. Strain HI4320 also contains a plasmid (NCBI NC_010555.1) (7), and the HI4320 chromosome encodes a complete set of *tra* genes for conjugative transfer. No plasmid was identified in BB2000, nor does its genome encode *tra* genes or any HI4320 plasmid-encoded genes. Further analysis of variations between *P. mirabilis* isolates will advance our understanding of the genetic determinants of pathogenicity and self recognition.

Nucleotide sequence accession number. The *P. mirabilis* BB2000 genome sequence has been deposited in GenBank under the accession number BankIt1590180 BB2000 CP004022.

ACKNOWLEDGMENTS

Harvard University provided funding for this research.

The authors thank Beckman Coulter Genomics, the Harvard FAS Systems Biology Core, and the Harvard Research Computing Group for insightful advice during the genome construction.

REFERENCES

1. **Mobley HLT, Belas R.** 1995. Swarming and pathogenicity of *Proteus mirabilis* in the urinary tract. Trends Microbiol. **3:**280-284.
2. **Mathur S, Sabbuba NA, Suller MTE, Stickler DJ, Feneley RCL.** 2005. Genotyping of urinary and fecal *Proteus mirabilis* isolates from individuals with long-term urinary catheters. Eur. J. Clin. Microbiol. **24:**643-644.
3. **Nicolle LE.** 2005. Catheter-related urinary tract infection. Drug Aging **22:**627-639.
4. **Armbruster CE, Mobley HL.** 2012. Merging mythology and morphology: the multifaceted lifestyle of *Proteus mirabilis*. Nat. Rev. Microbiol. **10:**743-754.
5. **Dienes L.** 1946. Reproductive processes in *Proteus* cultures. J. Bacteriol. **51:**24-24.
6. **Gibbs KA, Urbanowski ML, Greenberg EP.** 2008. Genetic determinants of self identity and social recognition in bacteria. Science **321:**256-259.
7. **Pearson MM, Sebaihia M, Churcher C, Quail MA, Seshasayee AS, Luscombe NM, Abdellah Z, Arrosmith C, Atkin B, Chillingworth T, Hauser H, Jagels K, Moule S, Mungall K, Norbertczak H, Rabbinowitsch E, Walker D, Whithead S, Thomson NR, Rather PN, Parkhill J, Mobley HLT.** 2008. Complete genome

sequence of uropathogenic *Proteus mirabilis*, a master of both adherence and motility. J. Bacteriol. **190:**4027-4037.

8. **Belas R, Erskine D, Flaherty D.** 1991. Transposon mutagenesis in *Proteus mirabilis*. J. Bacteriol. **173:**6289-6293.
9. **Sambrook J, Russell DW.** 2001. Molecular cloning a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, New York.
10. **Goecks J, Nekrutenko A, Taylor J.** 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. **11:**R86.
11. **Delcher AL, Bratke KA, Powers EC, Salzberg SL.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23:**673-679.
12. **Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389-3402.
13. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25:**955-964.
14. **Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL.** 2004. Versatile and open software for comparing large genomes. Genome Biol. **5:**R12.
15. **Chaudhuri RR, Loman NJ, Snyder LAS, Bailey CM, Stekel DJ, Pallen MJ.** 2008. xBASE2: a comprehensive resource for comparative bacterial genomics. Nucleic Acids Res. **36:**D543-D546.

16. **Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35:**3100-3108.
17. **Wenren LM, Sullivan NL, Cardarelli L, Septer AN, Gibbs KA.** 2013. Two independent pathways for self-recognition in *Proteus mirabilis* are linked by type VI-dependent export. mBio **4:**e00374-00313.
18. **Finn RD, Clements J, Eddy SR.** 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. **39:**W29-37.
19. **Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16:**944-945.