CrossMark
click for updates

Draft Genome Sequence of *Pedobacter* sp. Strain Hv1, an Isolate from Medicinal Leech Mucosal Castings

Brittany M. Ott,[a] Lidia Beka,[b] Joerg Graf,[b,c] Rita V. M. Rio[a]

Department of Biology, West Virginia University, Morgantown, West Virginia, USA[a]; Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut, USA[b]; Institute for Systems Genomics, University of Connecticut, Storrs, Connecticut, USA[c]

The *Pedobacter* sp. Hv1 strain was isolated from the medicinal leech, *Hirudo verbana,* mucosal castings. These mucosal sheds have been demonstrated to play a role in horizontal symbiont transmission. Here, we report the draft 4.9 Mbp genome sequence of *Pedobacter* sp. strain Hv1.

Address correspondence to Rita V. M. Rio, rita.rio@mail.wvu.edu.

The genus *Pedobacter* was reclassified from the genus *Sphingobacterium* in 1998 ([1]) due to various phenotypic variations. Members of the *Pedobacter* genus have been described from water and soil samples, as well as from the microbiotas of nematodes, dung beetles, and mosquitoes ([2–4]). Recently, strain Hv1 was identified as the most abundant bacterium within the mucosal casts of medicinal leeches (*Hirudo verbana*) ([5]). Here, we present a draft genome of *Pedobacter* sp. Hv1 strain.

The genome of *Pedobacter* Hv1 was sequenced using the Illumina MiSeq (1,880,901 paired-end reads). Raw reads were assembled using the CLC Genomics Workbench (Aahrus, Denmark), yielding 30 total contigs with lengths of >0.5 kb and an N_{50} of 665,494 bp. Due to low coverage, three contigs were immediately eliminated. The *oriC* was detected using Ori-Finder ([6, 7]) in Contig 12 (positions 65,484 to 65,930). An additional assembly was generated using AbySS ([8]) with varying k-mer lengths, which were aligned with the CLC assembly using Mauve ([9, 10]). However, the gaps could not be resolved. BLASTn detected 4 contigs with ~100% sequence identity and 0 gaps compared to regions in larger contigs. Subsequently, Tandem Repeats Finder ([11]) did not identify long terminal repeats typically associated with mobile elements, and the lack of obvious G+C skew eliminates recent gene duplication events. Mapping reads with BWA (Burrow-Wheeler Alignment tool) ([12]) and IGV (Integrative Genomics Viewer) ([13, 14]) eliminated these contigs due to high probability of an assembly error, resulting in 23 total contigs with lengths >0.85 kb. The updated contigs totaled 4,900,269 bp with an average coverage of 78×; N_{50} of 666,494 bp. The G+C content is 37.2%.

Functional annotation of the assembled genome was performed with RAST 2.0 ([15, 16]), and manual annotation and curation, which predicted 4,421 protein encoding genes and 38 RNAs. RNAs were further classified using tRNAscan-SE-1.23 ([17]), which predicted 35 tRNA genes (including tRNAs for all 20 amino acids, with 4 each for leucine and arginine), and the RNAmmer 1.2 server ([18]), which identified one 5S, one 23S, and one 16S rRNA gene. Genes related to the utilization and biosynthesis of carbohydrates, including N-acetylglucosamine, man-

nose, and trehalose decomposition, suggests the capability of adapting to varying metabolites. Hv1 contains a number of loci involved in osmotic and oxidative stress response, such as superoxide dismutase [Cu-Zn] precursor, cytochrome c551 peroxidase, and the universal stress protein family 4.

The Hv1 genome contains loci involved in type I, III, and IV secretions systems. Of particular interest, given the diverse microbiota from which Hv1 was originally isolated, is the presence of the conjugative transposon (e.g., various *tra* loci), associated with the type IV secretion system and the transfer of DNA from environmental sources ([19]). Interestingly, this operon is missing in the closest, fully sequenced relative, *Pedobacter heparinus* ([20]) (RAST, closest neighbor tool; similarity score, 542), while the more distantly related *P. saltans* ([21]) houses the same genetic components of the *tra* operon (with the exception of *tra*G and *tra*P).

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. LLWP00000000. The version described in this paper is version LLWP01000000.

ACKNOWLEDGMENTS

REFERENCES

1. **Steyn PL, Segers P, Vancanneyt M, Sandra P, Kersters K, Joubert JJ.** 1998. Classification of heparinolytic bacteria into a new genus, *Pedobacter*, comprising four species: *Pedobacter heparinus* comb. nov., *Pedobacter piscium* comb. nov., *Pedobacter africanus* sp. nov. and *Pedobacter saltans* sp. nov. proposal of the family Sphingobacteriaceae fam. nov. Int J Syst Bacteriol **48**:165–177. http://dx.doi.org/10.1099/00207713-48-1-165.
2. **Baquiran J-P, Thater B, Sedky S, De Ley P, Crowley D, Orwin PM.** 2013. Culture-independent investigation of the microbiome associated with the nematode *Acrobeloides maximus*. PLoS One **8**:e67425. http://dx.doi.org/10.1371/journal.pone.0067425.
3. **Estes AM, Hearn DJ, Snell-Rood EC, Feindler M, Feeser K, Abebe T, Dunning Hotopp JC, Moczek AP.** 2013. Brood ball-mediated transmission of microbiome members in the dung beetle, *Onthophagus taurus*

(Coleoptera: Scarabaeidae). PLoS One **8**:e79061. http://dx.doi.org/10.1371/journal.pone.0079061.

4. **Charan SS, Pawar KD, Severson DW, Patole MS, Shouche YS.** 2013. Comparative analysis of midgut bacterial communities of *Aedes aegypti* mosquito strains varying in vector competence to dengue virus. Parasitol Res **112**:2627–2637. http://dx.doi.org/10.1007/s00436-013-3428-x.

5. **Ott BM, Rickards A, Gehrke L, Rio RVM.** 2015. Characterization of shed medicinal leech mucus reveals a diverse microbiota. Front Microbiol **5**:757. http://dx.doi.org/10.3389/fmicb.2014.00757.

6. **Gao F, Zhang C-T.** 2008. Ori-finder: a web-based system for finding oriCs in unannotated bacterial genomes. BMC Bioinformatics **9**:79. http://dx.doi.org/10.1186/1471-2105-9-79.

7. **Luo H, Zhang C-T, Gao F.** 2014. Ori-finder 2, an integrated tool to predict replication origins in the archaeal genomes. Front Microbiol **5**:482. http://dx.doi.org/10.3389/fmicb.2014.00482.

8. **Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I.** 2009. ABySS: a parallel assembler for short read sequence data. Genome Res **19**:1117–1123. http://dx.doi.org/10.1101/gr.089532.108.

9. **Darling ACE, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14**:1394–1403. http://dx.doi.org/10.1101/gr.2289704.

10. **Darling AE, Mau B, Perna NT.** 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One **5**:e11147. http://dx.doi.org/10.1371/journal.pone.0011147.

11. **Benson G.** 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res **27**:573–580. http://dx.doi.org/10.1093/nar/27.2.573.

12. **Li H, Durbin R.** 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics **26**:589–595. http://dx.doi.org/10.1093/bioinformatics/btp698.

13. **Thorvaldsdóttir H, Robinson JT, Mesirov JP.** 2013. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform **14**:178–192. http://dx.doi.org/10.1093/bib/bbs017.

14. **Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP.** 2011. Integrative genomics viewer. Nat Biotechnol **29**:24–26. http://dx.doi.org/10.1038/nbt.1754.

15. **Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA.** 2008. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics **9**:386. http://dx.doi.org/10.1186/1471-2105-9-386.

16. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O.** 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics **9**:75. http://dx.doi.org/10.1186/1471-2164-9-75.

17. **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. http://dx.doi.org/10.1093/nar/25.5.0955.

18. **Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res **35**:3100–3108. http://dx.doi.org/10.1093/nar/gkm160.

19. **Cascales E, Christie PJ.** 2003. The versatile bacterial type IV secretion systems. Nat Rev Microbiol **1**:137–149. http://dx.doi.org/10.1038/nrmicro753.

20. **Han C, Spring S, Lapidus A, Del Rio TG, Tice H, Copeland A, Cheng J, Lucas S, Chen F, Nolan M, Bruce D, Goodwin L, Pitluck S, Ivanova N, Mavromatis K, Mikhailova N, Pati A, Chen A, Palaniappan K, Land M, Hauser L, Chang Y-J, Jeffries CC, Saunders E, Chertkov O, Brettin T, Göker M, Rohde M, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk H-P, Detter JC.** 2009. Complete genome sequence of *Pedobacter heparinus* type strain (HIM 762–3T). Stand Genomic Sci **1**:54–62. http://dx.doi.org/10.4056/sigs.22138.

21. **Liolios K, Sikorski J, Lu M, Nolan M, Lapidus A, Lucas S, Hammon N, Deshpande S, Cheng J-F, Tapia R, Han C, Goodwin L, Pitluck S, Huntemann M, Ivanova N, Pagani I, Mavromatis K, Ovchinikova G, Pati A, Chen A, Palaniappan K, Land M, Hauser L, Brambilla E-M, Kotsyurbenko O, Rohde M, Tindall BJ, Abt B, Göker M, Detter JC, Woyke T, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Klenk H-P, Kyrpides NC.** 2011. Complete genome sequence of the gliding, heparinolytic *Pedobacter saltans* type strain (113T). Stand Genomic Sci **5**:30–40. http://dx.doi.org/10.4056/sigs.2154937.