Draft Genome Sequences of Two Unclassified Bacteria, *Hydrogenophaga* sp. Strains IBVHS1 and IBVHS2, Isolated from Environmental Samples

Russell J. S. Orr,[a]* Stephane Rombauts,[b,c] Yves Van de Peer,[b,c] Kamran Shalchian-Tabrizi[a]

Section for Genetics and Evolutionary Biology (EVOGENE), Department of Biosciences, University of Oslo, Oslo, Norway[a]; Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium[b]; VIB Center for Plant Systems Biology, Ghent, Belgium[c]

ABSTRACT   We report here the draft genome sequences of *Hydrogenophaga* sp. strains IBVHS1 and IBVHS2, two bacteria assembled from the metagenomes of surface samples from freshwater lakes. The genomes are >95% complete and may represent new species within the *Hydrogenophaga* genus, indicating a larger diversity than currently identified.

Address correspondence to Russell J. S. Orr, russell_orr@hotmail.com.

* Present address: Russell J. S. Orr, Kristine Bonnevies hus, Oslo, Norway.

Advances in high-throughput sequencing, coupled with decreasing costs, have led to the number of available bacterial genomes increasing almost exponentially. Genome sequencing, however, has traditionally been limited to species that can be held and grown in culture due to the high DNA volumes needed. A predominant focus on cultivable species has led to a genome bias, where the true bacterial diversity is poorly represented. Metagenomic studies are rectifying this bias and have already revealed a large novel diversity (1). However, metagenomic studies remain limited, with many ecosystems yet to be sampled. We attempt to expand species richness in a bioproject with a goal to identify novel bacteria from various environmental samples. Here, we present the draft genomes of two unclassified *Hydrogenophaga* bacteria, isolated from the surface of freshwater lakes in Norway (Årungen, Ås) and Japan (Tsukuba, Ibaraki).

DNA was isolated using a standard phenol-chloroform protocol with ethanol precipitation and subsequent cleaning using Zymo genomic clean and concentrator. DNA was prepared and sequenced on the Illumina HiSeq2500 platform (150-bp paired-end reads; 350-bp insert size) with PacBio RS2 P6-C4 chemistry (20 kb) at the Norwegian Sequencing Centre. Metagenome drafts were assembled using SPAdes version 3.9.0 (2), single genomes were separated with MetaBAT (3), and quality was assessed with CheckM (4). Separate genomes were scaffolded using LINKS (5), and gaps were closed with Sealer (6). Genome assemblies were evaluated with PROmer (7) and REAPER (8) before being improved with Pilon (9). Genomes were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (10). Taxonomical rank was established on evaluation of CheckM (4), PhyloSift (11), and megaBLAST against the NCBI nr database.

Hydrogenophaga sp. IBVHS1 was assembled into two scaffolds, constituting seven contigs with a sequence length of 4.43 Mb and a GC content of 64.98%. The scaffold N_{50} was 3.05 Mb with Illumina coverage of 207$\times$ and PacBio coverage of 52$\times$. CheckM estimated genome completeness at 97.26% with no contamination or strain heterogeneity. The genome constitutes 4,194 genes, 46 RNAs, 40 tRNAs, 4 noncoding RNAs (ncRNAs), and 46 pseudogenes.

Hydrogenophaga sp. IBVHS2 was assembled into seven scaffolds, constituting nine contigs with a total sequence length of 3.17 Mb and a GC content of 68.73%. The

scaffold N_{50} was 0.96 Mb with Illumina coverage of 197$\times$ and PacBio coverage of 35$\times$. CheckM estimated genome completeness at 95.56% with no contamination or strain heterogeneity. The genome constitutes 2,952 genes, 43 RNAs, 39 tRNAs, 3 ncRNAs, and 23 pseudogenes.

Accession number(s). The draft genomes of *Hydrogenophaga* sp. strains IBVHS1 and IBVHS2 sequenced under this project have been deposited at DDBJ/EMBL/GenBank under the accession numbers NFUU00000000 and NFUT00000000, respectively. These biosamples (SAMN06840507 and SAMN06840508, respectively) are part of BioProject PRJNA384425.

REFERENCES

1. Nayfach S, Pollard KS. 2016. Toward accurate and quantitative comparative metagenomics. Cell 166:1103–1116. https://doi.org/10.1016/j.cell.2016.08.007.
2. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
3. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165. https://doi.org/10.7717/peerj.1165.
4. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
5. Warren RL, Vandervalk BP, Jones SJ, Birol I. 2015. LINKS: scaffolding genome assemblies with kilobase-long nanopore reads. bioRxiv https://doi.org/10.1101/016519.
6. Paulino D, Warren RL, Vandervalk BP, Raymond A, Jackman SD, Birol I. 2015. Sealer: a scalable gap-closing application for finishing draft genomes. BMC Bioinformatics 16:230. https://doi.org/10.1186/s12859-015-0663-4.
7. Delcher AL, Phillippy A, Carlton J, Salzberg SL. 2002. Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res 30:2478–2483. https://doi.org/10.1093/nar/30.11.2478.
8. Hunt M, Newbold C, Berriman M, Otto TD. 2014. A comprehensive evaluation of assembly scaffolding tools. Genome Biol 15:R42. https://doi.org/10.1186/gb-2014-15-3-r42.
9. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
10. Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity GM, Kodira CD, Kyrpides N, Madupu R, Markowitz V, Tatusova T, Thomson N, White O. 2008. Toward an online repository of standard operating procedures (SOPs) for (meta) genomic annotation. OMICS 12:137–141. https://doi.org/10.1089/omi.2008.0017.
11. Darling AE, Jospin G, Lowe E, Matsen FA, Bik HM, Eisen JA. 2014. PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ 2:e243. https://doi.org/10.7717/peerj.243.