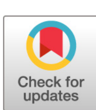
PROKARYOTES



Draft Genome Sequence of a Novel *Chitinophaga* sp. Strain, MD30, Isolated from a Biofilm in an Air Conditioner Condensate Pipe

Xuehua Wan,[a] Maxwell Darris,[b] Shaobin Hou,[a] Stuart P. Donachie[a,b]

Advanced Studies in Genomics, Proteomics and Bioinformatics, University of Hawaiʻi at Mānoa, Honolulu, Hawaii, USA[a]; Department of Microbiology, University of Hawaiʻi at Mānoa, Honolulu, Hawaii, USA[b]

ABSTRACT Most of the 24 known *Chitinophaga* species were originally isolated from soils. We report the draft genome sequence of a putatively novel *Chitinophaga* sp. from a biofilm in an air conditioner condensate pipe. The genome comprises 7,661,303 bp in one scaffold, 5,694 predicted protein-coding sequences, and a G+C content of 47.6%.

Received 15 September 2017 **Accepted** 18 September 2017 **Published** 19 October 2017

Citation Wan X, Darris M, Hou S, Donachie SP. 2017. Draft genome sequence of a novel *Chitinophaga* sp. strain, MD30, isolated from a biofilm in an air conditioner condensate pipe. Genome Announc 5:e01161-17. https://doi.org/10.1128/genomeA.01161-17.

Copyright © 2017 Wan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Stuart P. Donachie, donachie@hawaii.edu.

As part of a study of the urban microbiome in Honolulu, Hawaii, a heterotrophic bacterial strain designated MD30 was isolated from a biofilm hanging in water flowing from an air conditioning condensate pipe. A comparison of a 1,404-nucleotide (nt) fragment of the MD30 16S rRNA gene with those in the EzBioCloud database revealed the nearest neighbors to be *Chitinophaga niabensis* JS13-10^T (1) and *Chitinophaga ginsengisoli* Gsoil 052^T (2), both isolated from soil in the Republic of Korea, with 96.5% nucleotide identity over 1,397 and 1,396 nt, respectively. We sequenced the genome of MD30 to provide insights into what may distinguish this putative new *Chitinophaga* species from the type strains of the current 24 species in the genus.

Genomic DNA was isolated from *Chitinophaga* sp. strain MD30 using a phenol-chloroform method, followed by isopropanol precipitation and a 70% ethanol wash. More than 400 Mbp of sequences were generated by various platforms: 93.2 Mbp of shotgun sequences and 152.3 Mbp of 8-kb paired-end sequences by Roche 454 GS FLX+ pyrosequencing; 75.4 Mbp of Illumina paired-end sequences by the Nextera XT DNA library preparation kit and MiSeq v3 sequencing kit; and 85.9 Mbp in 14,254 long reads with an average size of 6 kbp generated by Oxford Nanopore chemistry (R9) and corrected by Canu (3). Different assembly strategies were tested, e.g., Newbler 2.8, Canu, and SPAdes 3.90 (4). Newbler assembled the Roche 454 reads into two scaffolds spanning 7,679,346 bp (N_{50} = 7,677,276) but with >160 gaps. SPAdes assembled the 454, Illumina, and Oxford Nanopore sequences, and we improved the assembled scaffolds in npScarf (5). The four scaffolds were compared to the single scaffold assembled by Newbler and then ordered using Mauve (6). After *in silico* gap closing using the BGI GapCloser (http://soap.genomics.org.cn/index.html) and BWA (7), one scaffold spanning 7,661,303 bp was assembled, with just two gaps remaining. The G+C content of the genome is 47.6%, consistent with the mean of 47.3% ± 3.6% in the genus.

The genome was annotated in the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (8), Rapid Annotation using Subsystem Technology (RAST) (9, 10), and Prokka 1.11 (11). PGAP identified 5,694 protein-coding genes and 53 tRNA-coding regions. RAST identified 6,082 protein-coding genes and 397 subsystems. Prokka identified 6,029 protein-coding genes and 53 tRNA-coding regions. CRISPRs finder identified one clustered regularly interspaced short palindromic repeat (CRISPR) region with 29 spac-

Downloaded from http://genomea.asm.org/ on November 11, 2017 by guest

ers (12, 13). Other than MotB, CheB, CheR, and CheY, most protein components of flagellar and chemotaxis systems were absent. A two-component regulatory system, FixL-FixJ, the heme-based PAS (Per-Arnt-Sim) domain-based oxygen sensor, and its cognate response regulator, were predicted (14). Further genomic analyses will help us understand the role this putatively novel species plays in a biofilm, such as that from which it was cultivated, and also how it differs from other *Chitinophaga* species.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number CP023254. The version described here is the first version.

ACKNOWLEDGMENTS

MD30 was isolated during M.D.'s participation in an NSF-funded Research Experiences for Undergraduates Site in summer 2016.

This work was supported through the NSF REU Site, "Undergraduate Research Experiences in DNA-Based Discoveries in Hawaii's Biodiversity," under award DBI1560491 to S.P.D. and Stephanie Kraft-Terry at the University of Hawai'i at Mānoa. Participation of X.W. and S.H. was provided by the Advanced Studies in Genomics, Proteomics and Bioinformatics.

REFERENCES

1. Weon HY, Yoo SH, Kim YJ, Son JA, Kim BY, Kwon SW, Koo BS. 2009. *Chitinophaga niabensis* sp. nov. and *Chitinophaga niastensis* sp. nov., isolated from soil. Int J Syst Evol Microbiol 59:1267–1271. https://doi.org/10.1099/ijs.0.004804-0.
2. Lee HG, An DS, Im WT, Liu QM, Na JR, Cho DH, Jin CW, Lee ST, Yang DC. 2007. *Chitinophaga ginsengisegetis* sp. nov. and *Chitinophaga ginsengisoli* sp. nov., isolated from soil of a ginseng field in South Korea. Int J Syst Evol Microbiol 57:1396–1401. https://doi.org/10.1099/ijs.0.64688-0.
3. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
4. 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.
5. Cao MD, Nguyen SH, Ganesamoorthy D, Elliott AG, Cooper MA, Coin LJM. 2017. Scaffolding and completing genome assemblies in real-time with nanopore sequencing. Nat Commun 8:14515. https://doi.org/10.1038/ncomms14515.
6. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147. https://doi.org/10.1371/journal.pone.0011147.
7. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler Transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
8. Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, 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.
9. 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. https://doi.org/10.1186/1471-2164-9-75.
10. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotations of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:D206–D214. https://doi.org/10.1093/nar/gkt1226.
11. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
12. Grissa I, Vergnaud G, Pourcel C. 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics 8:172. https://doi.org/10.1186/1471-2105-8-172.
13. Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a Web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res 35:W52–W57. https://doi.org/10.1093/nar/gkm360.
14. Gilles-Gonzalez MA, Gonzalez G. 2004. Signal transduction by heme-containing PAS-domain proteins. J Appl Physiol 96:774–783. https://doi.org/10.1152/japplphysiol.00941.2003.

Downloaded from http://genomea.asm.org/ on November 11, 2017 by guest