



# Complete Genome Sequence of the Novel Cellulolytic, Anaerobic, Thermophilic Bacterium *Herbivorax saccincola* Type Strain GGR1, Isolated from a Lab Scale Biogas Reactor as Established by Illumina and Nanopore MinION Sequencing

Alexander Pechtl,<sup>a</sup> Christian Rückert,<sup>b</sup> Irena Maus,<sup>b</sup> Daniela E. Koeck,<sup>a</sup> Nataliya Trushina,<sup>a</sup> Petra Kornberger,<sup>a</sup> Wolfgang H. Schwarz,<sup>a</sup> Andreas Schlüter,<sup>b</sup> Wolfgang Liebl,<sup>a</sup>  Vladimir V. Zverlov<sup>a,c</sup>

<sup>a</sup>Department of Microbiology, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany

<sup>b</sup>Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany

<sup>c</sup>Institute of Molecular Genetics, Russian Academy of Science, Moscow, Russia

**ABSTRACT** The cellulolytic bacterium *Herbivorax saccincola* strain GGR1, which represents the type strain of this species, was isolated from the *in vivo* enriched cellulose-binding community of a lab scale thermophilic biogas reactor. Here, we report the complete genome sequence of *H. saccincola* GGR1<sup>T</sup>, the first isolated member of the genus *Herbivorax*.

Biogas is a sustainable energy carrier produced in large scale from organic biomass by anaerobic fermentation in biogas reactors. Recent studies have reported on the abundance of cellulolytic *Ruminococcaceae* members within the microbial communities of thermophilic biogas-producing plants (1, 2). Members of this group were isolated from various environments, and some of them are highly efficient lignocellulose degraders that produce cellulosomes (3). The anaerobic bacterium *Herbivorax saccincola* GGR1<sup>T</sup> was isolated from a thermophilic grass silage/cow manure biogas reactor (4). Phylogenetic analysis based on 16S rRNA gene sequencing classified the new isolate GGR1<sup>T</sup> as belonging to a hitherto unknown subgroup within the family *Ruminococcaceae*. It was shown to utilize cellulose and xylan as sole carbon sources. Hydrogen, ethanol, and acetate are the major fermentation products (4). To assess its adaptive genome features, the strain GGR1<sup>T</sup> was completely sequenced using the Illumina and Nanopore MinION sequencing technologies.

For GGR1<sup>T</sup> genome sequencing, chromosomal DNA was used to generate a shotgun 8-kb mate-pair sequencing library that was sequenced on an Illumina MiSeq system (5). Sequence data were assembled using Newbler version 2.8 (Roche), resulting in 12 scaffolds containing 139 contigs. Subsequently, 2  $\mu$ g of the GGR1<sup>T</sup> genomic DNA was used to generate a second shotgun library for sequencing on the MinION system (Oxford Nanopore Technologies). DNA fragments of 5 to 50 kb were used to create a 1D<sup>2</sup> sequencing library, which was loaded to an R9.5 flowcell for a 24-h run on the MinION sequencer. Base calling and data conversion were performed using Albacore version 1.2.4 (<https://github.com/Albacore/albacore>). The assembly was performed applying Canu version 1.5 (6). After assembly, the resulting eight contigs were polished with the short Illumina reads using Pilon (7). The final assembly was done manually using Consed (8). This resulted in a circular contig of 3,604,547 bp, featuring a G+C content of 34.82%. Gene prediction and annotation were performed applying Prokka (9) and GenDB (10). This approach resulted in the detection of 3,228 coding sequences, 52 tRNAs, and 3 *rnm* operons.

Received 8 December 2017 Accepted 5 January 2018 Published 8 February 2018

**Citation** Pechtl A, Rückert C, Maus I, Koeck DE, Trushina N, Kornberger P, Schwarz WH, Schlüter A, Liebl W, Zverlov VV. 2018. Complete genome sequence of the novel cellulolytic, anaerobic, thermophilic bacterium *Herbivorax saccincola* type strain GGR1, isolated from a lab scale biogas reactor as established by Illumina and Nanopore MinION sequencing. *Genome Announc* 6:e01493-17. <https://doi.org/10.1128/genomeA.01493-17>.

**Copyright** © 2018 Pechtl et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Wolfgang Liebl, [wliebl@wzw.tum.de](mailto:wliebl@wzw.tum.de), or Vladimir V. Zverlov, [vladimir.zverlov@tum.de](mailto:vladimir.zverlov@tum.de).

A.P. and C.R. contributed equally to this work.

The GGR1<sup>T</sup> genome harbors 114 genes encoding carbohydrate-active enzymes identified by means of the carbohydrate-active-enzyme database (CAZy) annotation Web server dbCAN (11). The identified CAZy modules comprise 68 glycoside hydrolases (GHs), 20 carbohydrate esterases, 24 glycosyl transferases, and 2 polysaccharide lyases.

A total of 60 putative cellulosomal genes were identified by the presence of coding regions for type 1 or 2 dockerin or cohesin modules, among others. Fifty of these genes encode 35 different GHs and other carbohydrate-active enzymes. One corresponding primary scaffoldin comprising 11 type 1 cohesin modules and a carbohydrate-binding module (CBM3) was identified in the genome of GGR1<sup>T</sup> as well as 9 other scaffoldins and anchoring proteins. Two cellulosomal enzymes are unusual because one contains both type 1 dockerin and cohesin modules, and another one likewise harbors type 2 dockerin and cohesin modules. The availability of the *H. saccincola* GGR1<sup>T</sup> genome sequence provides the genetic basis for biotechnological exploitation of encoded biocatalysts and serves as a reference for the analysis of biomass-digesting microbial communities.

**Accession number(s).** The genome sequence of *H. saccincola* GGR1<sup>T</sup> has been deposited in the EMBL/GenBank database (EBI, NCBI) under the accession number [CP025197](https://doi.org/10.1093/nar/gks479). The strain is available from the Leibniz Institute German Collection of Microorganisms and Cell Cultures (DSMZ, Braunschweig, Germany) under strain number DSM 101079.

## ACKNOWLEDGMENTS

This project was supported by grants 03SF0440E and 031A556 from the German Federal Ministry of Education and Research to A.P. and P.K. and by grants 22017112 and 22021715 from the German Federal Ministry of Food and Agriculture to D.E.K., W.L., and V.V.Z. The work of I.M. was supported by the German Federal Ministry of Education and Research-funded project “Bielefeld-Gießen Center for Microbial Bioinformatics—BiGi” (grant number 031A533) within the German Network for Bioinformatics Infrastructure (de.NBI). This publication was supported by the German Research Foundation (DFG) and the Technical University of Munich within the Open Access Publishing Programme.

## REFERENCES

- Maus I, Koeck DE, Cibis KG, Hahnke S, Kim YS, Langer T, Kreubel J, Erhard M, Bremges A, Off S, Stolze Y, Jaenicke S, Goesmann A, Szczyrba A, Scherer P, König H, Schwarz WH, Zverlov VV, Liebl W, Pühler A, Schlüter A, Klocke M. 2016. Unraveling the microbiome of a thermophilic biogas plant by metagenome and metatranscriptome analysis complemented by characterization of bacterial and archaeal isolates. *Biotechnol Biofuels* 9:171. <https://doi.org/10.1186/s13068-016-0581-3>.
- Koeck DE, Maus I, Wibberg D, Winkler A, Zverlov VV, Liebl W, Pühler A, Schwarz WH, Schlüter A. 2015. Draft genome sequence of *Herbinix hemicellulosilytica* T3/55<sup>T</sup>, a new thermophilic cellulose degrading bacterium isolated from a thermophilic biogas reactor. *J Biotechnol* 214: 59–60. <https://doi.org/10.1016/j.jbiotec.2015.07.022>.
- Koeck DE, Pechtl A, Zverlov VV, Schwarz WH. 2014. Genomics of cellulolytic bacteria. *Curr Opin Biotechnol* 29:171–183. <https://doi.org/10.1016/j.copbio.2014.07.002>.
- Koeck DE, Mechelke M, Zverlov VV, Liebl W, Schwarz WH. 2016. *Herbinorax saccincola* gen. nov., sp. nov., a cellulolytic, anaerobic, thermophilic bacterium isolated via *in sacco* enrichments from a lab-scale biogas reactor. *Int J Syst Evol Microbiol* 66:4458–4463. <https://doi.org/10.1099/ijsem.0.001374>.
- Maus I, Cibis KG, Wibberg D, Winkler A, Stolze Y, König H, Pühler A, Schlüter A. 2015. Complete genome sequence of the strain *Defluviitoga tunisiensis* L3, isolated from a thermophilic, production-scale biogas plant. *J Biotechnol* 203:17–18. <https://doi.org/10.1016/j.jbiotec.2015.03.006>.
- 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>.
- 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>.
- Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res* 8:195–202. <https://doi.org/10.1101/gr.8.3.195>.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
- Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J, Kalinowski J, Linke B, Rupp O, Giegerich R, Pühler A. 2003. GenDB—an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res* 31:2187–2195. <https://doi.org/10.1093/nar/gkg312>.
- Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a Web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 40:W445–W451. <https://doi.org/10.1093/nar/gks479>.