



Draft Genome Sequences of *Dysgonomonas* sp. Strains BGC7 and HGC4, Isolated from the Hindgut of a Lower Termite

 Charles M. Bridges,^a  Michael C. Nelson,^{a*}  Joerg Graf,^a  Daniel J. Gage^a

^aDepartment of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut, USA

ABSTRACT *Dysgonomonas* spp. are facultative heterotrophs which colonize diverse environments, including the hindgut of the lower termite *Reticulitermes flavipes*. *Dysgonomonas* genomes are enriched for genes involving oligo- and polysaccharide utilization, enabling modification of a wide array of complex glycans. Here, we report draft genome sequences for *Dysgonomonas* sp. strains BGC7 and HGC4.

Dysgonomonas spp. are facultative anaerobic heterotrophs which grow on many oligo- and polysaccharides. *Dysgonomonas* spp. are particularly prevalent in lignocellulolytic termites and cockroaches (reference 1 and references therein), and there is interest in understanding their metabolic potential, particularly with regard to the biotransformation of complex lignocellulose- and host-derived glycans (2). Here, we present draft genome sequences for *Dysgonomonas* sp. strains BGC7 and HGC4, cultured from the hindgut of the lower termite *Reticulitermes flavipes*.

Termite hindguts were extirpated using the technique described by Matson et al. (3). Strain BGC7 was isolated on modified Eggerth-Gagnon agar (ATCC medium, 2840), and strain HGC4 was isolated on the same medium, but sterile sheep blood was replaced with a hemin-vitamin K₁ solution (25 mg and 5 mg ml⁻¹, respectively). Cultures were maintained on rich peptone-hemin-glucose (rPHG) agar (1), and plates were grown anaerobically under an atmosphere of 89% N₂, 5.5% CO₂, and 5.5% H₂ at 22°C. 16S rRNA gene sequences from isolates BGC7 and HGC4 (1) (available under GenBank accession numbers [MT340878](#) and [MT340881](#), respectively) placed both isolates within the genus *Dysgonomonas*. Genomic DNA was prepared using a Promega Wizard genomic DNA purification kit. Concentration and quality were determined by using gel electrophoresis and a Qubit fluorometer. DNA was mechanically sheared and Illumina TruSeq PCR-free libraries were prepared using DNA size selected for 550 bp. Libraries were sequenced on an Illumina MiSeq instrument using a 2 × 250-bp v2 kit. BBTools v36.38 (<https://sourceforge.net/projects/bbmap/>) was used to remove adapter sequences and quality trim and filter raw reads (threshold of Q15). Processed reads were assessed with FastQC v0.11.5 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), and draft assemblies were created using A5-miseq v20160825 (4) and SPAdes v3.12.0 (5) using default settings on the KBase platform (6). Assemblies were merged using MAC (19 July 2020) (7) with default settings. Processed reads were aligned to MAC-merged assemblies using Bowtie 2 v2.3.2 (8), and SAMtools v1.10 (9) was used for file conversion. Pilon v1.23 (10) was used to polish merged assemblies iteratively three times. Polished assemblies were manually curated and annotated by submission to the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v4.13 (11). Genomes were checked for completeness and contamination using CheckM v1.0.18 (12) on KBase. Genomic relatedness to type and non-type strain *Dysgonomonas* genomes (available from RefSeq as of 15 October 2020) was determined by digital DNA-DNA hybridization (dDDH) performed using the Type Strain Genome Server (13). Average amino acid identities (AAI) between orthologous genes were computed with CompareM v0.1.1 (<https://github.com/dparks1134/CompareM>). Carbohydrate active

Citation Bridges CM, Nelson MC, Graf J, Gage DJ. 2021. Draft genome sequences of *Dysgonomonas* sp. Strains BGC7 and HGC4, isolated from the hindgut of a lower termite. Microbiol Resour Announc 10:e01427-20. <https://doi.org/10.1128/MRA.01427-20>.

Editor Irene L. G. Newton, Indiana University, Bloomington

Copyright © 2021 Bridges et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Daniel J. Gage, daniel.gage@uconn.edu.

* Present address: Michael C. Nelson, Sema4 Genomics, Stamford, Connecticut, USA.

Received 10 December 2020

Accepted 6 January 2021

Published 28 January 2021

TABLE 1 Genome details of *Dysgonomonas* sp. strains

Parameter	Data for strain:	
	BGC7	HGC4
BioProject accession no.	PRJNA656570	PRJNA656570
BioSample accession no.	SAMN15793200	SAMN15793201
SRA accession no.	SRR12430954	SRR12430953
WGS accession no.	JACMIC000000000.1	JACMIB000000000.1
RefSeq accession	GCF_014840225.1	GCF_014840235.1
Taxonomy ID ^a	1658008	1658009
No. of raw sequencing reads	4,689,088	4,276,966
No. of filtered reads	4,265,276	3,081,766
Mean scaffold coverage ^b (×)	223	154
No. of scaffolds >100 bp	25	18
No. of contigs >100 bp	32	22
Largest scaffold (bp)	1,432,443	856,340
Genome size (bp)	4,336,043	4,502,296
<i>N</i> ₅₀ (bp)	874,238	769,148
<i>L</i> ₅₀ (bp)	2	3
G+C content (%)	37.13	35.92
No. of coding sequences	3,507	3,644
No. of protein-coding sequences	3,451	3,601
No. of tRNAs	42	42
Completeness (%)	100	100
Contamination (%)	0.55	0
Closest neighbor, 16S rRNA gene (% identity)	<i>Dysgonomonas gadei</i> ATCC BAA-286 ^c (96.42)	<i>Dysgonomonas</i> sp. HDW5A ^d (98.42)
Closest neighbor, whole genome (% dDDH ^e)	<i>Dysgonomonas capnocytophagoides</i> DSM 22835 ^f (26.5)	<i>Dysgonomonas alginatilytica</i> DSM 100214 ^g (34.1)
Closest neighbor, protein-coding sequences	<i>D. gadei</i> ATCC BAA-286 ^c	<i>D. alginatilytica</i> DSM 100214 ^g
Orthologous fraction of genes ^h (% of total)	65.43	78.09
AAI between orthologs ⁱ (%)	75.35	90.44
Total no. of CAZy domains ^j	209	282
No. of glycosyl hydrolase domains	149	199
No. of carbohydrate esterase domains	10	11
No. of polysaccharide lyase domains	14	9
No. of carbohydrate binding module domains	3	11
No. of glycosyl transferase domains	33	52
No. of protein-coding sequences with ≥1 CAZy domain (% of total)	207 (6.0)	267 (7.4)
Total no. of PULs	45	41
No. of CAZy-associated PULs (% of total)	29 (64)	32 (78)

^a ID, identifier.^b Determined using unpaired filtered reads mapped to Pilon-polished assembly using Bowtie 2.^c RefSeq assembly accession number [GCF_000213555.1](#).^d RefSeq assembly accession number [GCF_011299555.1](#).^e Digital DNA-DNA hybridization values calculated using formula *d*₄ (sum of all identities found in high-scoring pairs [HSPs] divided by overall HSP length).^f RefSeq assembly accession number [GCF_000426485.1](#).^g RefSeq assembly accession number [GCF_003201355.1](#).^h The number of orthologs divided by the minimum number of protein-coding genes from either genome.ⁱ Average amino acid identity between orthologous fraction of genes.^j Domains identified by HMMER and at least one other tool (HotPep or DIAMOND).

enzyme (CAZy) domains were identified using the dbCAN2 server (14), and polysaccharide utilization loci (PULs) were identified using PULpy (15), with both using default settings.

Genome accession details, assembly metrics, phylogenetic relatedness, and CAZy/PUL content can be found in Table 1. The genomes of *Dysgonomonas* sp. strains BGC7 and HGC4 were enriched for CAZy domain-containing proteins and PULs involved in degradation of plant polysaccharides, particularly xylans. The availability of these sequences will contribute to understanding the metabolic potential and diversity of lignocellulose-degrading organisms, particularly within the genus *Dysgonomonas*.

Data availability. Genome assemblies for isolates BGC7 and HGC4 were submitted to NCBI under BioProject accession number [PRJNA656570](#) and whole-genome sequence (WGS) accession numbers [JACMIC000000000.1](#) and [JACMIB000000000.1](#), respectively. Detailed information can be found in Table 1.

ACKNOWLEDGMENTS

We thank Kendra Maas and the UConn Microbial Analysis, Resources, and Services (MARS) facility for consultation, Illumina library preparation, and sequencing.

This work was funded by the National Science Foundation Emerging Frontiers in Research and Innovation: Multicellular and Inter-Kingdom Signaling (EFRI-MIKS) award number 1137249 to Ranjan Srivastava, Leslie Shor, William E. Mustain, Kenneth Noll, J.G., and D.J.G. The funders had no role in data collection, interpretation, or preparation of data for publication.

REFERENCES

1. Bridges CM, Gage DJ. 2021. Development and application of aerobic, chemically defined media for *Dysgonomonas*. *Anaerobe* 67:102302. <https://doi.org/10.1016/j.anaerobe.2020.102302>.
2. Lap  bie P, Lombard V, Drula E, Terrapon N, Henrissat B. 2019. Bacteroidetes use thousands of enzyme combinations to break down glycans. *Nat Commun* 10:2043. <https://doi.org/10.1038/s41467-019-10068-5>.
3. Matson E, Ottesen E, Leadbetter J. 2007. Extracting DNA from the gut microbes of the termite (*Zootermopsis nevadensis*). *J Vis Exp* e195. <https://doi.org/10.3791/195>.
4. Coil D, Jospin G, Darling AE. 2015. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 31:587–589. <https://doi.org/10.1093/bioinformatics/btu661>.
5. 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>.
6. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S, et al. 2018. KBase: the United States department of energy systems biology knowledgebase. *Nat Biotechnol* 36:566–569. <https://doi.org/10.1038/nbt.4163>.
7. Tang L, Li M, Wu F-X, Pan Y, Wang J. 2019. MAC: merging assemblies by using adjacency algebraic model and classification. *Front Genet* 10:1396. <https://doi.org/10.3389/fgene.2019.01396>.
8. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
9. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
10. 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>.
11. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>.
12. 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>.
13. Meier-Kolthoff JP, G  ker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 10:2182. <https://doi.org/10.1038/s41467-019-10210-3>.
14. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 46:W95–W101. <https://doi.org/10.1093/nar/gky418>.
15. Stewart RD, Auffret MD, Roehle R, Watson M. 2018. Open prediction of polysaccharide utilisation loci (PUL) in 5414 public Bacteroidetes genomes using PULpy. *bioRxiv* <https://doi.org/10.1101/421024>.