AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

OMICS DATA SETS

Metagenome Assembly and Metagenome-Assembled Genome Sequences from a Historical Oil Field Located in Wietze, Germany

Michael O. Eze,[a,b,c] Stephan A. Lütgert,[d] Hannes Neubauer,[a,b] Angeliki Balouri,[a,b] Alina A. Kraft,[a,b] Anja Sieven,[a,b] Rolf Daniel,[a,b] Bernd Wemheuer[a,b]

[a]Genomic and Applied Microbiology, Georg-August University of Göttingen, Göttingen, Germany
[b]Göttingen Genomics Laboratory, Georg-August University of Göttingen, Göttingen, Germany
[c]Department of Earth and Environmental Sciences, Macquarie University, Sydney, NSW, Australia
[d]German Oil Museum, Wietze, Germany

ABSTRACT Crude oil-polluted sites are a global threat, raising the demand for remediation worldwide. Here, we investigated a crude oil metagenome from a former borehole in Wietze, Germany, and reconstructed 42 metagenome-assembled genomes, many of which contained genes involved in crude oil degradation with a high potential for bioremediation purposes.

Bioremediation of crude oil-contaminated sites is highly investigated due to severe pollution levels in various ecosystems worldwide. It can be enhanced by the application of microorganisms, and thus it is important to discover novel microbes capable of crude oil degradation (1).

Three crude oil-contaminated samples were taken on 11 October 2016 from a former borehole (52.6592N, 9.8323E) located at a historical oil field in Wietze, Germany (https://www.erdoelmuseum.de). Approximately 5 g of contaminated soil was taken per sample, transported to the laboratory on ice, and stored at −20°C. Environmental DNA was extracted from 100 mg of soil using the PowerSoil DNA extraction kit as recommended by the manufacturer (Qiagen, Hilden, Germany). Paired-end sequencing libraries were constructed using the Nextera DNA sample preparation kit (Illumina, San Diego, CA, USA) and the following Nextera DNA indices: N708/N508 (sample 1), N709/N508 (sample 2), and N710/N508 (sample 3). Paired-end sequencing was performed using a HiSeq 2500 instrument (rapid run mode, 500 cycles), as recommended by the manufacturer (Illumina), and resulted in 46,673,322 paired-end reads (sample 1, 16,094,584 reads; sample 2, 17,883,658 reads; sample 3, 12,695,080 reads). Reads were processed with Trimmomatic version 0.36 (2). Processing included the removal of adapter sequences and low-quality regions. Default parameters were used for all software unless otherwise specified. The quality of the processing was confirmed using FastQC version 0.91. A total of 42,049,950 paired-end reads and 1,147,707 unpaired reads were retained and assembled using metaSPAdes version 3.13.2 (3). Assembly resulted in 1,544,944 scaffolds; of these, 22,257 were larger than 2,500 bp. Coverage information for each scaffold was determined using Bowtie 2 version 2.3.2 (4) and SAMtools version 1.7 (5). The average sequencing depth was approximately 7×. Metagenome-assembled genomes (MAGs) were reconstructed with MetaBAT version 2.12.1 (6). MAG quality was determined using CheckM version 1.0.13 (7). Only MAGs with a completeness minus contamination of more than 50% and a contamination rate of less than 7% were considered for further analysis. MAGs were classified taxonomically using GTDB-Tk version 1.0.2 and the Genome Taxonomy Database (GTDB) (release 86) (8, 9), resulting in 6 archaeal MAGs and 36 bacterial MAGs. Archaeal MAGs were

Citation Eze MO, Lütgert SA, Neubauer H, Balouri A, Kraft AA, Sieven A, Daniel R, Wemheuer B. 2020. Metagenome assembly and metagenome-assembled genome sequences from a historical oil field located in Wietze, Germany. Microbiol Resour Announc 9:e00333-20. https://doi.org/10.1128/MRA.00333-20.

Editor Catherine Putonti, Loyola University Chicago

Address correspondence to Rolf Daniel, rdaniel@gwdg.de.

Received 30 March 2020
Accepted 3 May 2020
Published 21 May 2020

classified as members of the *Euryarchaeota* (1 MAG), *Halobacterota* (3 MAGs), and *Thermoplasmatota* (2 MAGs). Bacterial MAGs belonged to *Actinobacteriota* (4 MAGs), *Bacteroidota* (5 MAGs), *Chloroflexota* (5 MAGs), *Desulfobacterota* (4 MAGs), *Firmicutes* (2 MAGs), *Omnitrophota* (1 MAG), *Patescibacteria* (1 MAG), *Proteobacteria* (10 MAGs), *Spirochaetota* (1 MAG), *Synergistota* (1 MAG), and *Thermotogota* (1 MAG). One bacterial MAG was assigned to an unclassified taxon associated with *Nitrospirae*. After annotation with Prodigal version 2.6.3 (10), functional annotation was performed with DIAMOND version 0.9.29 (11) and the KEGG database (October 2018 release) (12). Functional analysis revealed that all MAGs obtained possess genes involved in xenobiotic degradation. One MAG assigned to *Rugosibacter*, a genus of known xenobiotic degraders (13), showed the highest abundance of pathways associated with xenobiotic degradation (11.8%).

Data availability. Raw sequencing data are available at the NCBI Sequence Read Archive (SRA) under accession numbers SRR10568503, SRR10568510, and SRR10568511. The metagenome assembly and the MAGs are available at GenBank under accession numbers WOYI00000000 and WOYJ00000000 to WOZY00000000, respectively. Further genome characteristics and the functional annotation are publicly available at the Göttingen Research Online Database (https://doi.org/10.25625/VX8836).

ACKNOWLEDGMENTS

We thank Anja Poehlein and Melanie Heinemann for their assistance during sequencing. We are grateful to the staff of the oil museum in Wietze (https://www.erdoelmuseum.de) for help during sampling.

REFERENCES

1. Das N, Chandran P. 2011. Microbial degradation of petroleum hydrocarbon contaminants: an overview. Biotechnol Res Int 2011:941810. https://doi.org/10.4061/2011/941810.
2. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
3. 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.
4. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
5. Li H, 1000 Genome Project Data Processing Subgroup, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
6. 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.
7. 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.
8. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P. 2019. Selection of representative genomes for 24,706 bacterial and archaeal species clusters provide a complete genome-based taxonomy. bioRxiv https://doi.org/10.1101/771964.
9. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927. https://doi.org/10.1093/bioinformatics/btz848.
10. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
11. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.
12. Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28:27–30. https://doi.org/10.1093/nar/28.1.27.
13. Corteselli EM, Aitken MD, Singleton DR. 2017. *Rugosibacter aromaticivorans* gen. nov., sp. nov., a bacterium within the family *Rhodocyclaceae*, isolated from contaminated soil, capable of degrading aromatic compounds. Int J Syst Evol Microbiol 67:311–318. https://doi.org/10.1099/ijsem.0.001622.