



Draft Metagenome-Assembled Genome Sequences of Three Novel Ammonia-Oxidizing *Nitrososphaera* Strains Recovered from Agricultural Soils in Western Colorado

Arsen Yerlan,^a Rebecca A. Daly,^a Reza Keshavarz Afshar,^{b*} Michael Shaffer,^{a§} Kelly C. Wrighton,^a Bridget B. McGivern^a

^aDepartment of Soil and Crop Science, Colorado State University, Fort Collins, Colorado, USA

^bWestern Colorado Research Center, Fruita, Colorado, USA

ABSTRACT Microbial nitrification is critical to nitrogen loss from agricultural soils. Here, we report three thaumarchaeotal metagenome-assembled genomes (MAGs) representing a new species of *Nitrososphaera*. These genomes expand the representation of archaeal nitrifiers recovered from arid, agricultural soils.

Ammonium- or urea-based fertilizers are the dominant form of applied nitrogen in agricultural systems (1). Ammonia-oxidizing bacteria and archaea (AOA) are recognized as partially controlling the fate of this nitrogen through nitrification, rapidly converting ammonium to nitrate, which is more mobile and can lead to substantial nitrous oxide (N₂O) production (1–3). The AOA are classified as *Thermoproteota* (formerly *Thaumarchaeota*) and have been primarily tracked by marker gene sequencing in soil (4). Here, we report three metagenome-assembled genomes (MAGs) for a novel *Nitrososphaera* species recovered from agricultural soil in western Colorado.

Soil samples were collected from the Western Colorado Research Center (Fruita, CO, USA; 39°10'47.9994", -108°42'0") in February 2021. Surface (0- to 5-cm) soil samples were taken from fallow agricultural plots managed under conventional tillage (*n* = 1) and an untilled system (*n* = 1). DNA was extracted from 0.4 g of each soil using the Zymo Quick-DNA fecal/soil microbe microprep kit, following the soil protocol. Metagenomic libraries were prepared using the Tecan Ovation Ultralow v2 system and were sequenced on the NovaSeq 6000 platform on a S4 flow cell at Genomics Shared Resource, Colorado Cancer Center (Denver, CO, USA). The untilled and tilled metagenomes comprised 37.9 Gbp and 28.5 Gbp of 150-bp paired-end reads, respectively. For each metagenome, the read quality was determined using FastQC v0.11.2 (5), the reads were trimmed using Sickle v1.33 (pe -t sanger) (6) and assembled using MEGAHIT v1.2.9 (-k-min 31 -k-max 121 -k-step 10 -mem-flag 1) (7), and the contigs were binned using MetaBat2 v2.12.1 (8). The MAG quality was assessed using

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

Copyright © 2022 Yerlan et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Kelly C. Wrighton, kelly.wrighton@colostate.edu, or Bridget B. McGivern, bridget.mcgivern@colostate.edu.

*Present address: Reza Keshavarz Afshar, Rodale Institute, Kutztown, Pennsylvania, USA.

§Present address: Michael Shaffer, Merck, Cambridge, Massachusetts, USA.

The authors declare no conflict of interest.

Received 12 April 2022

Accepted 20 July 2022

Published 8 August 2022

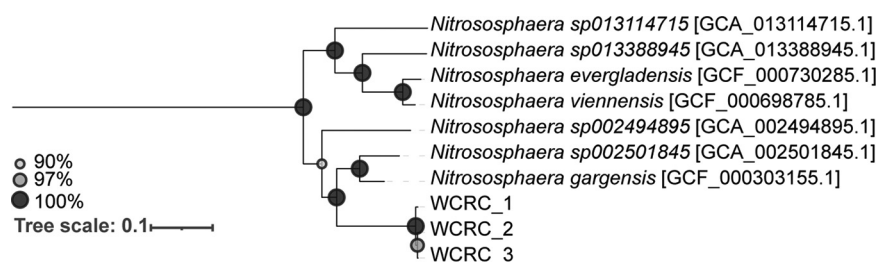


FIG 1 Phylogenetic tree based on the WCRC *Nitrososphaera* MAGs and GTDB-tk r202 *Nitrososphaera* species representatives. The tree is rooted on the species representatives of *g_Nitrosocosmicus*. The GTDB-tk *de_novo_wf* workflow was used to generate a multiple-sequence alignment (MSA) using *g_Nitrosocosmicus* as the outgroup and filtering to *g_Nitrososphaera*. The resulting MSA was used to construct a maximum likelihood phylogenetic tree using RAxML v8.2.9 (15) with the PROTGAMMAWAG model and 100 bootstraps. Bootstraps for the nodes were all greater than 90% and are sized according to the legend.

TABLE 1 Metagenome-assembled genome statistics for WCRC_1, WCRC_2, and WCRC_3

Characteristic	Data for strain:		
	WCRC_1	WCRC_2	WCRC_3
Origin soil management	Conventional till	Untilled	Conventional till
BioSample accession no.	SAMN26177291	SAMN26177292	SAMN26177293
Genome size (bp)	1,081,057	1,177,263	876,390
No. of contigs	154	181	123
GC content (%)	44.6	44.4	44.7
Longest contig (bp)	29,717	25,749	20,557
N_{50} (bp)	8,096	7,090	7,633
Completeness (%)	78.8	83.5	71.36
Contamination (%)	1.94	2.91	0.97
No. of predicted coding genes	1,275	1,395	1,050
No. of tRNAs	28	25	20
Encoded rRNA	5S		
Mean base coverage (×)	6.05	6.28	10.04

CheckM v1.1.2 (9), and the taxonomy was assigned using GTDB-tk v1.5.0 (r202) (10). MAG annotation was performed using DRAM (11) within KBase (12). Default parameters were used unless noted. Two of the *Nitrososphaera* MAGs (WCRC_1 and WCRC_3) were recovered from the conventional tilled soil metagenome and the other (WCRC_2) from the untilled soil metagenome.

The three MAGs were assigned to a new species in the genus *Nitrososphaera* using GTDB-tk, where there are currently 14 *Nitrososphaera* genomic representatives across seven species (GTDB-tk r202) (Fig. 1). The pairwise amino acid identity is 97.67% between the three MAGs, suggesting that they are members of the same species (13). The statistics of these three MAGs are presented in Table 1.

Genome annotation of the *Nitrososphaera* MAGs supported their roles as AOA. All MAGs encoded the B and C subunits of ammonia monooxygenase (EC 1.14.99.39). While the A subunit is missing across the MAGs, this is likely due to the known challenge of assembling this gene and the incomplete nature of these MAGs (Table 1). In accordance with other observed *Thaumarchaeota*, hydroxylamine oxidase was absent across the MAGs, while WCRC_1 encoded nitrite reductase (14). Collectively, these MAGs provide genomic context for a novel species of ammonia-oxidizing *Nitrososphaera* derived from agricultural soils.

Data availability. The sequencing data for this project have been deposited at GenBank under BioProject accession number [PRJNA725542](#). The MAGs have been deposited under BioSample accession numbers [SAMN26177291](#), [SAMN26177292](#), and [SAMN26177293](#). The metagenomic reads have been deposited in the Sequence Read Archive under accession numbers [SRS11831377](#) and [SRS11831378](#).

ACKNOWLEDGMENTS

We thank Katrina Deiner at the Genomics Core at the University of Colorado Anschutz Medical Campus for sequencing expertise. We also thank the students of the Soil Microbiome Research Experience course (SOCR 481A-5) for their support during data analysis. Finally, we thank the DOE Systems Biology Knowledgebase (KBase) educational team, specifically Elisha Wood-Charlson and Ellen Dow, for assistance with data ingestion into KBase.

KBase is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, under award numbers DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886. B.B.M. and K.C.W. were fully or partially supported by an Early Career Award to K.C.W. from the National Science Foundation under award number 1912915.

REFERENCES

- Wendeborn S. 2020. The chemistry, biology, and modulation of ammonium nitrification in soil. *Angew Chem Int Ed Engl* 59:2182–2202. <https://doi.org/10.1002/anie.201903014>.
- Clark IM, Hughes DJ, Fu Q, Abadie M, Hirsch PR. 2021. Metagenomic approaches reveal differences in genetic diversity and relative abundance of nitrifying bacteria and archaea in contrasting soils. *Sci Rep* 11:15905. <https://doi.org/10.1038/s41598-021-95100-9>.
- Tian H, Xu R, Canadell JG, Thompson RL, Winiwarter W, Suntharalingam P, Davidson EA, Ciais P, Jackson RB, Janssens-Maenhout G, Prather MJ, Regnier P, Pan N, Pan S, Peters GP, Shi H, Tubiello FN, Zaehele S, Zhou F, Armeth A, Battaglia

- G, Berthet S, Bopp L, Bouwman AF, Buitenhuis ET, Chang J, Chipperfield MP, Dangal SRS, Dlugokencky E, Elkins JW, Eyre BD, Fu B, Hall B, Ito A, Joos F, Krummel PB, Landolfi A, Laruelle GG, Lauerwald R, Li W, Lienert S, Maavara T, MacLeod M, Millet DB, Olin S, Patra PK, Prinn RG, Raymond PA, Ruiz DJ, van der Werf GR, et al. 2020. A comprehensive quantification of global nitrous oxide sources and sinks. *Nature* 586:248–256. <https://doi.org/10.1038/s41586-020-2780-0>.
4. Huang L, Chakrabarti S, Cooper J, Perez A, John SM, Daroub SH, Martens-Habbena W. 2021. Ammonia-oxidizing archaea are integral to nitrogen cycling in a highly fertile agricultural soil. *ISME Commun* 1:19. <https://doi.org/10.1038/s43705-021-00020-4>.
 5. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
 6. Joshi NA, Fass JN. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. <https://github.com/najoshi/sickle>.
 7. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>.
 8. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. <https://doi.org/10.7717/peerj.7359>.
 9. 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>.
 10. Chaumeil PA, 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>.
 11. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P, Narrowe AB, Rodríguez-Ramos J, Bolduc B, Gazitúa MC, Daly RA, Smith GJ, Vik DR, Pope PB, Sullivan MB, Roux S, Wrighton KC. 2020. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res* 48:8883–8900. <https://doi.org/10.1093/nar/gkaa621>.
 12. 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, 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>.
 13. Konstantinidis KT, Rosselló-Móra R, Amann R. 2017. Uncultivated microbes in need of their own taxonomy. *ISME J* 11:2399–2406. <https://doi.org/10.1038/ismej.2017.113>.
 14. Zhalnina KV, Dias R, Leonard MT, Dorr de Quadros P, Camargo FAO, Drew JC, Farmerie WG, Daroub SH, Triplett EW. 2014. Genome sequence of *Candidatus Nitrososphaera evergladensis* from group I.1b enriched from Everglades soil reveals novel genomic features of the ammonia-oxidizing archaea. *PLoS One* 9:e101648. <https://doi.org/10.1371/journal.pone.0101648>.
 15. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.