Article title

Genome sequence of *Nitrosopumilus adriaticus* CCS1 assembled from an ammonia-oxidizing enrichment culture

Authors

Muzi Li[a*], Patrick H. Thieringer[a*], Barbara Bayer[b], Matthew Kellom[c], Alyson E. Santoro[a#]

Affiliations

* These authors contributed equally.

a. Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, California, USA

b. Department of Microbiology and Ecosystem Science, Centre for Microbiology and Environmental System Science, University of Vienna, Vienna, Austria

c. Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California, USA

Running title

Genome of *Nitrosopumilus adriaticus* CCS1

Corresponding author's email address

Alyson E. Santoro, asantoro@ucsb.edu

Abstract

We report the metagenome-assembled genome of an ammonia-oxidizing archaeon that is closely related to *Nitrosopumilus adriaticus* NF5, but shows distinct genomic features compared to strain NF5.

Announcement

Ammonia-oxidizing archaea (AOA) catalyze the first step of nitrification at nearly all depths of the oceanic water column (1). *Nitrosopumilus adriaticus* CCS1 is an AOA enriched from shallow seawater along the Pacific Coast of North America.

Nitrosopumilus sp. CCS1 was enriched during the cultivation of *Ca*. Nitrosopelagicus brevis CN25 (2). *Ca*. N. brevis CN25 was cultivated in filter-sterilized (0.2 µm pore size) medium (2), made from surface seawater obtained at various locations around the southern California Current. Thus, the exact provenance of strain CCS1 is unknown. During a periodic observation of the CN25 enrichment using flow cytometry, we identified the presence of a second putative AOA population with a higher DNA content, which was eventually isolated by continued transfers in a HEPES-buffered artificial seawater medium (3) with additions of antibiotics (100 µg mL^{-1} streptomycin and kanamycin). Though CCS1 has since been isolated (3), at the time of sequencing, one heterotrophic contaminant remained in the CCS1 enrichment culture. CCS1 was the only AOA remaining in the culture at the time of sequencing.

Cultures were grown in 50 or 200 mL of HEPES-buffered artificial seawater medium (3) in polycarbonate bottles containing 50-500 µM NH_4Cl without agitation in the dark at 20-25ºC. Cultures were combined and vacuum filtered onto a single 0.22 µm pore size polyethersulfone (Pall Supor) membrane filter. DNA was extracted using the DNEasy Blood & Tissue Kit (Qiagen) with a modified lysis protocol (4). The DNA library was prepared using the Nextera XT kit and sequenced on the Illumina MiSeq platform. Paired-end reads (6,179,225 x 2) were generated with an average read length of 198 bp, resulting in 2,447,233,357 bases and a genome coverage of 1234x. All tools were used with their default parameters unless otherwise specified. Reads were assembled with SPAdes (v3.12.0) (5) (--careful -k 21,33,55,77,99,127). Scaffold binning was performed using MaxBin (v2.2.4) (6). The assembled genome was annotated via the Integrated Microbial Genomes (IMG) Annotation Pipeline (v5.2.1) (7).

The final genome has one scaffold with a total length of 1,782,213 bp and a GC content of 33.41% (Table 1). Genome completeness and contamination, estimated by CheckM2 (v1.0.1) (8), are 99.95% and 0.19%, respectively. OrthoFinder (v2.5.5) (9) was used to conduct a comparative genomic analysis against a representative set of ammonia-oxidizing archaea (2, 10). A concatenated alignment of single-copy genes in 758 orthogroups that are single-copy in at least 89.7% species was generated by OrthoFinder. The alignment was then fed into IQ-TREE (v2.3.0) (11) using the best-fit model LG+F+I+G4 and 1000 bootstrap replicates to construct a phylogenetic tree, where strain CCS1 and *Nitrosopumilus adriaticus* NF5 (12) were grouped into the same clade (Figure 1). *N. adriaticus* NF5 and strain CCS1 share an average nucleotide identity (ANI) of 97.9% as determined by FastANI (v1.32) (13), thus we designate this organism a strain of *N. adriaticus*. Despite the high genomic similarity, these two strains were derived from distant and unconnected oceans (strain NF5: Adriatic Sea, strain CCS1: Pacific Ocean). Each strain exhibits unique genomic features, for instance, strain NF5 possesses multiple Cas1 family proteins that are not found in strain CCS1. Contrastingly, strain CCS1 harbors several FkbM family methyltransferases absent in strain NF5. Moreover, strain CCS1 encodes an additional putative NO-forming nitrite reductase (JGI gene ID 2932132417) with low similarity to those found in strain NF5.

Table 1 Genome statistics of *Nitrosopumilus adriaticus* CCS1

Strain	IMG genome ID	Genome size (bp)	GC (%)	No. of scaffolds	No. of total genes	No. of protein coding	No. of rRNA operons	No. of tRNA operons	Completeness/ Contamination (%)
Nitrosopumilus adriaticus CCS1	2932130896	1,782,213	33.41	1	2,240	2,192	3	43	99.95/0.19

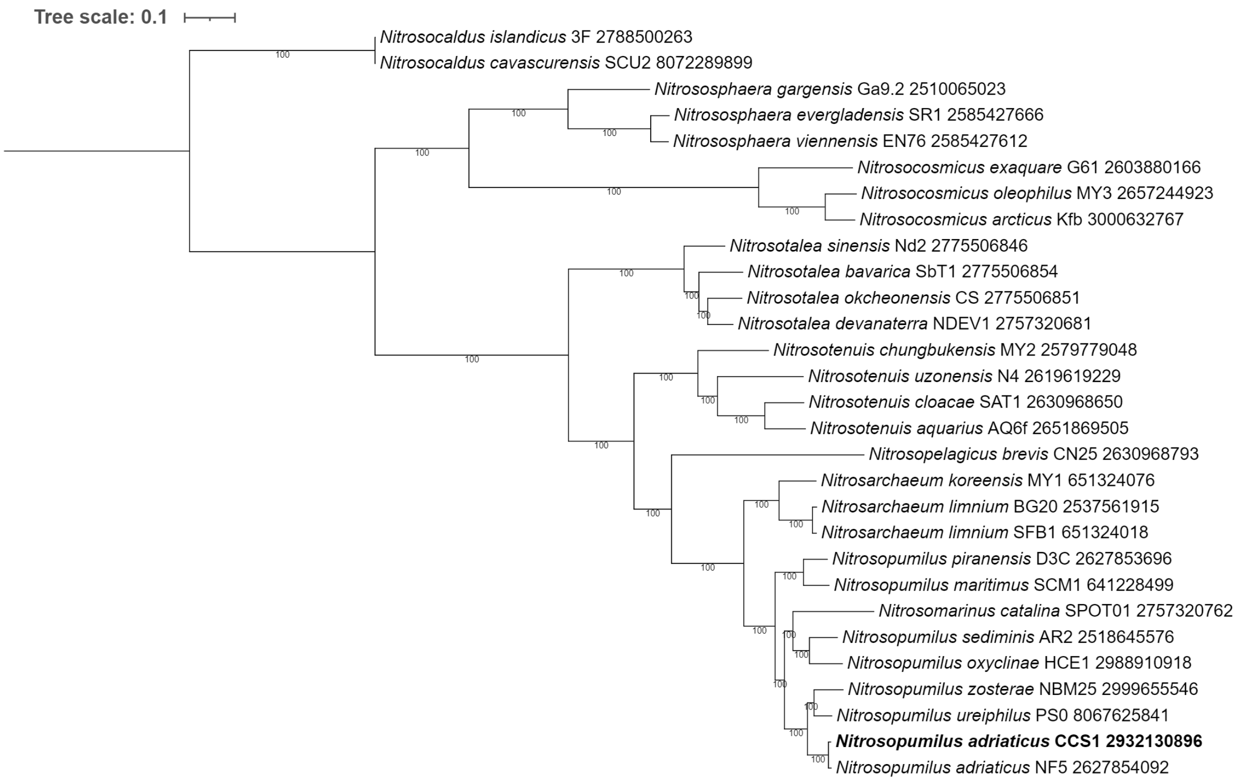

Figure 1. A maximum-likelihood phylogenetic tree of representative ammonia-oxidizing archaea genomes visualized with iTOL (v6) (15). The UFBoot support values are underneath the branches.

Data availability statement

The sequencing data were deposited at NCBI SRA under accession number SRR29007723. The genome of strain CCS1 is available in GenBank under accession number CP167059 and available at IMG under taxon ID 2932130896. The alignment file used for phylogenetic construction is accessible on figshare at https://doi.org/10.6084/m9.figshare.26515999.v1.

Acknowledgments

We thank Kelsey McBeain and Nathan Youlton for their assistance with cultivation and DNA extraction. The DNA library was prepared and sequenced at Biological Nanostructures Laboratory at the University of California, Santa Barbara. Genome sequencing and analysis was supported by a Simons Foundation Early Career Investigator in Marine Microbiology and Evolution Award (345889) and a Simons Investigator in Aquatic Microbial Ecology Award (LI-SIAME-00001560) to AES. MK was supported by Gordon and Betty Moore Foundation award MMI5541 to AES. BB was supported by the Austrian Science Fund (FWF) Project J4426-B. The work conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. Use was made of computational facilities purchased with funds from the National Science Foundation (CNS-1725797) and administered by the Center for Scientific Computing (CSC). The CSC is supported by the California NanoSystems Institute and the Materials Research Science and Engineering Center (MRSEC; NSF DMR 2308708) at UC Santa Barbara.

References

1. Santoro AE, Richter RA, Dupont CL. 2019. Planktonic Marine Archaea. Annual Review of Marine Science 11:131–158.
2. Santoro AE, Bayer B, Elling FJ, Pearson A. 2021. Candidatus Nitrosopelagicus, p. 1–13. *In* Bergey's Manual of Systematics of Archaea and Bacteria. John Wiley & Sons, Ltd.
3. Bayer B, McBeain K, Carlson CA, Santoro AE. 2023. Carbon content, carbon fixation yield and dissolved organic carbon release from diverse marine nitrifiers. Limnology and Oceanography 68:84–96.
4. Santoro AE, Casciotti KL, Francis CA. 2010. Activity, abundance and diversity of nitrifying archaea and bacteria in the central California Current. Environmental Microbiology 12:1989–2006.
5. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes De Novo Assembler. Current Protocols in Bioinformatics 70:e102.
6. Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607.
7. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, Szeto E, Pillay M, Chen I-MA, Pati A, Nielsen T, Markowitz VM, Kyrpides NC. 2015. The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). Stand in Genomic Sci 10:86.
8. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. 2023. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nat Methods 20:1203–1212.
9. Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20:238.
10. Qin W, Zheng Y, Zhao F, Wang Y, Urakawa H, Martens-Habbena W, Liu H, Huang X, Zhang X, Nakagawa T, Mende DR, Bollmann A, Wang B, Zhang Y, Amin SA, Nielsen JL, Mori K, Takahashi R, Virginia Armbrust E, Winkler M-KH, DeLong EF, Li M, Lee P-H, Zhou J, Zhang C, Zhang T, Stahl DA, Ingalls AE. 2020. Alternative strategies of nutrient acquisition and energy conservation map to the biogeography of marine ammonia-oxidizing archaea. The ISME Journal 14:2595–2609.
11. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution 37:1530–1534.
12. Bayer B, Vojvoda J, Offre P, Alves RJE, Elisabeth NH, Garcia JAL, Volland J-M, Srivastava A, Schleper C, Herndl GJ. 2016. Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation. The ISME Journal 10:1051–1063.
13. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114.