

Complete genome sequence of *Nitratifactor salsuginis* type strain (E9I37-1^T)

Iain Anderson¹, Johannes Sikorski², Ahmet Zeytun^{1,3}, Matt Nolan¹, Alla Lapidus¹, Susan Lucas¹, Nancy Hammon¹, Shweta Deshpande¹, Jan-Fang Cheng¹, Roxanne Tapia^{1,3}, Cliff Han^{1,3}, Lynne Goodwin^{1,3}, Sam Pitluck¹, Konstantinos Liolios¹, Ioanna Pagani¹, Natalia Ivanova¹, Marcel Huntemann¹, Konstantinos Mavromatis¹, Galina Ovchinnikova¹, Amrita Pati¹, Amy Chen⁴, Krishna Palaniappan⁴, Miriam Land^{1,5}, Loren Hauser^{1,5}, Evelyne-Marie Brambilla², Olivier D. Ngatchou-Djao⁶, Manfred Rohde⁶, Brian J. Tindall², Markus Göker², John C. Detter^{1,3}, Tanja Woyke¹, James Bristow¹, Jonathan A. Eisen^{1,7}, Victor Markowitz⁴, Philip Hugenholtz^{1,8}, Hans-Peter Klenk², and Nikos C. Kyrpides^{1*}

¹ DOE Joint Genome Institute, Walnut Creek, California, USA

² DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany

³ Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA

⁴ Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA

⁵ Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

⁶ HZI – Helmholtz Centre for Infection Research, Braunschweig, Germany

⁷ University of California Davis Genome Center, Davis, California, USA

⁸ Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia

*Corresponding author: Nikos C. Kyrpides

Keywords: anaerobic, microaerobic, non-motile, Gram-negative, mesophilic, strictly chemolithoautotroph, *Nautiliaceae*, GEBA

Nitratifactor salsuginis Nakagawa *et al.* 2005 is the type species of the genus *Nitratifactor*, a member of the family *Nautiliaceae*. The species is of interest because of its high capacity for nitrate reduction *via* conversion to N₂ through respiration, which is a key compound in plant nutrition. The strain is also of interest because it represents the first mesophilic and facultatively anaerobic member of the *Epsilonproteobacteria* reported to grow on molecular hydrogen. This is the first completed genome sequence of a member of the genus *Nitratifactor* and the second sequence from the family *Nautiliaceae*. The 2,101,285 bp long genome with its 2,121 protein-coding and 54 RNA genes is a part of the *Genomic Encyclopedia of Bacteria and Archaea* project.

Introduction

Strain E9I37-1^T (= DSM 16511 = JCM 12458) is the type strain of *Nitratifactor salsuginis*, which in turn is the type and currently only species of the genus *Nitratifactor* [1]. The genus name is derived from the Neo-Latin word *nitras* meaning *nitrate* and the Latin word *fractor* meaning *breaker*, yielding the Neo-Latin word *Nitratifactor* meaning *nitrate-breaker* [1]. *N. salsuginis* strain E9I37-1^T was isolated from a deep-sea hydrothermal vent chimney at the Iheya North hydrothermal field in the Mid-Okinawa Trough in Japan [1,2]. No further isolates of *N. salsuginis* have been obtained so far. Here we present a summary classification and a set of features for *N. salsuginis* E9I37-1^T, together with the

description of the complete genomic sequencing and annotation.

Classification and features

A representative genomic 16S rRNA sequence of strain E9I37-1^T was compared using NCBI BLAST under default settings (e.g., considering only the high-scoring segment pairs (HSPs) from the best 250 hits) with the most recent release of the Greengenes database [3] and the relative frequencies, weighted by BLAST scores, of taxa and keywords (reduced to their stem [4]) were determined. The four most frequent genera were *Nitratifactor*

ruptor (48.5%), *Nitratifactor* (20.7%), *Hydrogenimonas* (15.7%) and *Alvinella* (15.1%) (eleven hits in total). Regarding the single hit to sequences from members of the species, the average identity within HSPs was 100.0%, whereas the average coverage by HSPs was 95.6%. Among all other species, the one yielding the highest score was *Hydrogenimonas thermophila*, which corresponded to an identity of 88.5% and an HSP coverage of 67.2%. (Note that the Greengenes database uses the INSDC (= EMBL/NCBI/DDBJ) annotation, which is not an authoritative source for nomenclature or classification.) The highest-scoring environmental sequence was AF420348 ('hydrothermal sediment clone AF420348') [5], which showed an identity of 96.7% and an HSP coverage of 97.8%. The five most frequent keywords within the labels of environmental samples which yielded hits were 'cave' (7.2%),

'biofilm' (5.7%), 'sulfid' (5.3%), 'spring' (4.8%) and 'structur' (3.1%) (239 hits in total). The five most frequent keywords within the labels of environmental samples which yielded hits of a higher score than the highest scoring species were 'hydrotherm' (8.6%), 'vent' (7.5%), 'pacif' (4.0%), 'microbi' (3.7%) and 'mat' (3.0%) (37 hits in total). These keywords are in accordance with the origin of the strain *N. salsuginis* E9137-1^T from a deep-sea hydrothermal vent chimney at the summits of the sulfide mounds in the sediment-hosted back-arc hydrothermal system Iheya North [1,2].

The 16S rRNA based tree in Figure 1 shows the phylogenetic neighborhood of *N. salsuginis* E9137-1^T. The sequences of the two identical 16S rRNA gene copies in the genome do not differ from the previously published 16S rRNA sequence (AB175500).

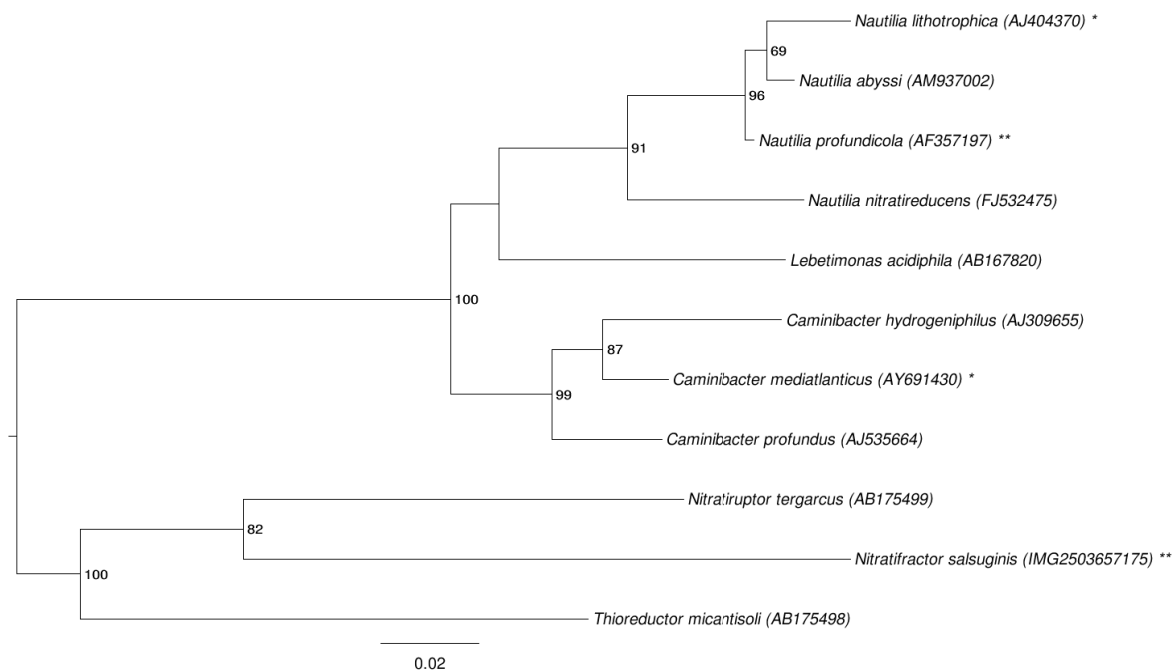


Figure 1. Phylogenetic tree highlighting the position of *N. salsuginis* strain E9137-1^T relative to the other type strains within the family *Nautiliaceae*. The tree was inferred from 1,356 aligned characters [6,7] of the 16S rRNA gene sequence under the maximum likelihood criterion [8] and rooted in accordance with the current taxonomy. The branches are scaled in terms of the expected number of substitutions per site. Numbers to the right of bifurcations are support values from 200 bootstrap replicates [9] if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [10] are labeled with an asterisk when unpublished, and with two asterisks when published [11]. The closest BLAST hit to *N. salsuginis* (see above) does not belong to *Nautiliaceae*, and this family does not appear as monophyletic in the last version of the 16S rRNA phylogeny from the All-Species-Living-Tree Project [12]. The species selection for Figure 1 was based on the current taxonomic classification (Table 1). However, an analysis including the type strains of *Nautiliaceae* and its neighboring families *Campylobacteraceae*, *Helicobacteraceae* and *Hydrogenimonaceae* (data not shown) did not provide evidence for the non-monophyly for any of these families.

The cells of strain E9I37-1^T are generally rod-shaped of 2.5 μm in length and 0.6 μm in width (Figure 2) and usually occur singly or in pairs (Figure 2) [1]. Strain E9I37-1^T is a Gram-negative, non-motile and non spore-forming bacterium (Table 1). The organism is anaerobic to microaerophilic (0.09-0.55% O₂ (v/v)) and chemolithoautotrophic, growing by respiratory nitrate reduction with H₂ as the electron donor, forming N₂ as a metabolic end product [1]. The main electron acceptors are NO₃⁻ or O₂ [1]. Strain E9I37-1^T uses S⁰ as a source of sulfur [1]. The doubling time of strain E9I37-1^T was about 2.5 h [1]. The NaCl range for growth is between 1.5% and 3.5%, with an optimum at 3%; no growth was observed below 1.0% NaCl or above 4.0% NaCl [1]. The temperature range for growth is between 28°C and 40°C, with an optimum at 37°C [1]. The pH range for growth is between 5.6 and 7.6, with an optimum at pH 7; no growth could be detected below pH 5.2 or above pH 8.1 [1]. Strain E9I37-1^T was unable to use any organic compounds as energy or carbon sources [1]. The organism was sensitive to ampicillin, rifampicin, streptomycin, chloramphenicol (each at 50 μg ml⁻¹) and kanamycin (200 μg ml⁻¹), and insensitive to approximately 150 μg ml⁻¹ kanamycin [1]. Enzymatic and genetic analyses demonstrated that strain E9I37-1^T uses the reductive TCA (rTCA) cycle for carbon assimilation [21]. This was confirmed by the presence of all genes encoding the three key rTCA cycle enzymatic ac-

tivities, namely ATP-dependent citrate lyase, pyruvate:ferredoxin oxidoreductase, and 2-oxoglutarate:ferredoxin oxidoreductase [21], but it was found to lack the gene for ribulose 1,5-bisphosphate carboxylase (RubisCO) activity, the key enzyme in the Calvin-Benson cycle [21].

Chemotaxonomy

The major cellular fatty acids of strain E9I37-1^T are C_{18:1} (42.3% of the total fatty acid), C_{16:1} (30.7%) and C_{16:0} (24.3%), C_{14:0} 3-OH (1.1%), C_{14:0} (0.9%) and C_{18:0} (0.7%) [1]. It should be noted that no information is given on the position of double bonds in the unsaturated fatty acids. No attempt has been made to examine the type strain for the presence of respiratory lipoquinones or to determine the polar lipid composition.

Genome sequencing and annotation

Genome project history

This organism was selected for sequencing on the basis of its phylogenetic position [22], and is part of the *Genomic Encyclopedia of Bacteria and Archaea* project [23]. The genome project is deposited in the Genome On Line Database [10] and the complete genome sequence is deposited in GenBank. Sequencing, finishing and annotation were performed by the DOE Joint Genome Institute (JGI). A summary of the project information is shown in Table 2.

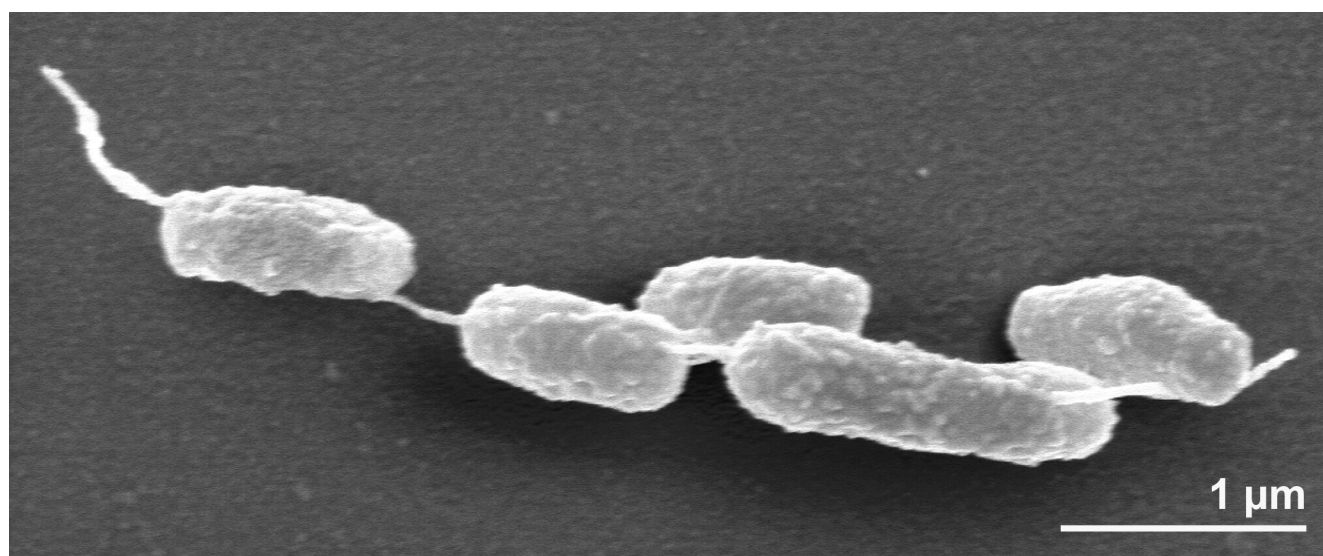


Figure 2. Scanning electron micrograph of *N. salsuginis* E9I37-1^T

Table 1. Classification and general features of *N. salsuginis* E9I37-1^T according to the MIGS recommendations [13].

MIGS ID	Property	Term	Evidence code	
		Domain <i>Bacteria</i>	TAS [14]	
		Phylum <i>Proteobacteria</i>	TAS [15]	
		Class <i>Epsilonproteobacteria</i>	TAS [16,17]	
	Current classification	Order <i>Nautiliales</i>	TAS [18]	
		Family <i>Nautiliaceae</i>	TAS [18]	
		Genus <i>Nitratifactor</i>	TAS [1]	
		Species <i>Nitratifactor salsuginis</i>	TAS [1]	
		Type strain E9I37-1	TAS [1]	
		Gram stain	negative	TAS [1]
		Cell shape	rod shaped, occurring singly or in pairs	TAS [1]
	Motility	non-motile	TAS [1]	
	Sporulation	none	TAS [1]	
	Temperature range	28-40°C	TAS [1]	
	Optimum temperature	37°C	TAS [1]	
	Salinity	1.5-3.5% NaCl	TAS [1]	
MIGS-22	Oxygen requirement	anaerobic and microaerobic	TAS [1]	
	Carbon source	probably CO ₂	NAS	
	Energy metabolism	strictly chemolithoautotrophic	TAS [1]	
MIGS-6	Habitat	deep-sea hydrothermal vent chimneys	TAS [1]	
MIGS-15	Biotic relationship	not reported	NAS	
MIGS-14	Pathogenicity	not reported	NAS	
	Biosafety level	1	TAS [19]	
	Isolation	deep-sea hydrothermal vent water of 'E9' chimney (inside part)	TAS [1,2]	
MIGS-4	Geographic location	Iheya North hydrothermal field in the Mid-Okinawa Trough in Japan	TAS [1,2]	
MIGS-5	Sample collection time	2002 or before	TAS [1,2]	
MIGS-4.1	Latitude	27.78	TAS [1,2]	
MIGS-4.2	Longitude	126.88	TAS [1,2]	
MIGS-4.3	Depth	984 m	TAS [1,2]	
MIGS-4.4	Altitude	not reported	NAS	

Evidence codes - IDA: Inferred from Direct Assay (first time in publication); TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from of the Gene Ontology project [20]. If the evidence code is IDA, the property was directly observed by one of the authors or an expert mentioned in the acknowledgements.

Growth conditions and DNA isolation

N. salsuginis E9I37-1^T, DSM 16511, was grown anaerobically in DSMZ medium 1024 (*Nitratiruptor* and *Nitratifactor* medium) [24] at 37°C. DNA was isolated from 0.5-1 g of cell paste using Jetflex Genomic DNA Purification Kit (GENOMED 600100) following the standard protocol as recommended by the manufacturer. Cell lysis was enhanced by adding 20 µl proteinase K for two hours at 58°C. DNA is available through the DNA Bank Network [25].

Genome sequencing and assembly

The genome was sequenced using a combination of Illumina and 454 sequencing platforms. All general aspects of library construction and sequencing can be found at the JGI website [26]. Pyrosequencing reads were assembled using the Newbler assembler (Roche). The initial Newbler assembly consisting of 42 contigs in five scaffolds was converted into a phrap [27] assembly by making fake reads from the consensus, to collect the read pairs in the 454 paired end library.

Illumina GAii sequencing data (158.03 Mb) was assembled with Velvet [28] and the consensus sequences were shredded into 1.5 kb overlapped fake reads and assembled together with the 454 data. The 454 draft assembly was based on 60.3 Mb 454 draft data and all of the 454 paired end data. Newbler parameters are -consed -a 50 -l 350 -g -m -ml 20. The Phred/Phrap/Consed software package [27] was used for sequence assembly and quality assessment in the subsequent finishing process. After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC). Possible mis-assemblies were corrected with gapResolution [26], Dupfinisher [29], or sequencing clones

bridging PCR fragments with subcloning. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR primer walks (J.-F. Chang, unpublished). A total of 135 additional reactions were necessary to close gaps and to raise the quality of the finished sequence. Illumina reads were also used to correct potential base errors and increase consensus quality using a software Polisher developed at JGI [30]. The error rate of the completed genome sequence is less than 1 in 100,000. Together, the combination of the Illumina and 454 sequencing platforms provided 106.7 × coverage of the genome. The final assembly contained 274,574 pyrosequence and 2,079,398 Illumina reads.

Table 2. Genome sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Finished
MIGS-28	Libraries used	Three genomic libraries: one 454 pyrosequence standard library, one 454 PE library (12 kb insert size), one Illumina library
MIGS-29	Sequencing platforms	Illumina GAii, 454 GS FLX Titanium
MIGS-31.2	Sequencing coverage	75.2 × Illumina; 31.5 × pyrosequence
MIGS-30	Assemblers	Newbler version 2.4, Velvet, phrap
MIGS-32	Gene calling method	Prodigal 1.4, GenePRIMP
	INSDC ID	CP002452
	Genbank Date of Release	January 24, 2011
	GOLD ID	Gc01594
	NCBI project ID	46883
	Database: IMG-GEBA	2503538035
MIGS-13	Source material identifier	DSM 16511
	Project relevance	Tree of Life, GEBA

Genome annotation

Genes were identified using Prodigal [31] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [32]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) non-redundant database, UniProt, TIGR-Fam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction analysis and functional annotation was performed within the Integrated Microbial Genomes - Expert Review (IMG-ER) platform [33].

Genome properties

The genome consists of a 2,101,285 bp long chromosome with a G+C content of 53.9% (Table 3 and Figure 3). Of the 2,175 genes predicted, 2,121 were protein-coding genes, and 54 RNAs; 33 pseudogenes were also identified. The majority of the protein-coding genes (66.9%) were assigned with a putative function while the remaining ones were annotated as hypothetical proteins. The distribution of genes into COGs functional categories is presented in Table 4.

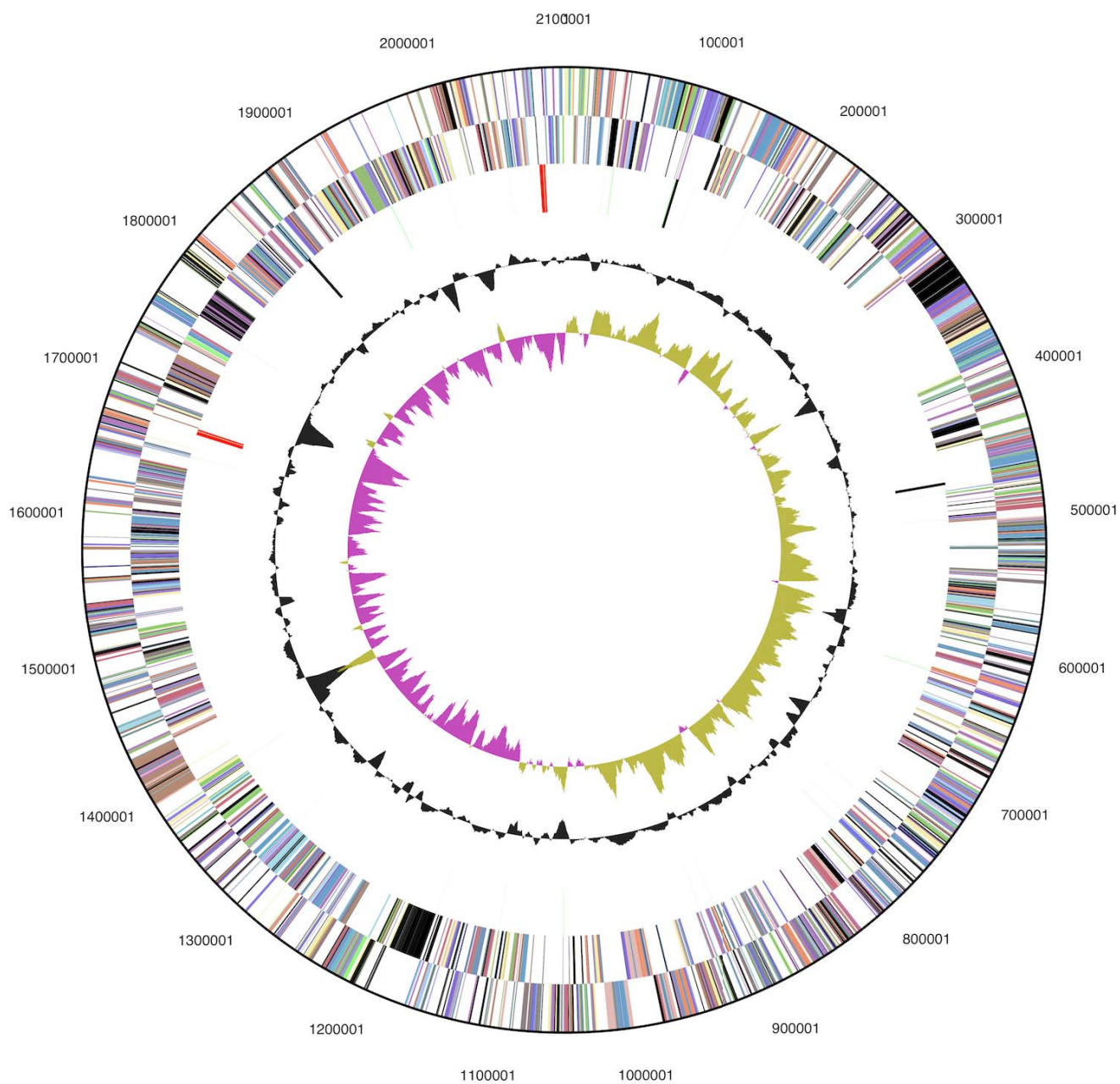


Figure 3. Graphical circular map of the chromosome; From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew.

Table 3. Genome Statistics

Attribute	Value	% of Total
Genome size (bp)	2,101,285	100.00%
DNA coding region (bp)	1,916,093	91.19%
DNA G+C content (bp)	1,132,843	53.91%
Number of replicons	1	
Extrachromosomal elements	0	
Total genes	2,175	100.00%
RNA genes	54	2.48%
rRNA operons	2	
Protein-coding genes	2,121	97.52%
Pseudo genes	33	1.52%
Genes with function prediction	1,456	66.94%
Genes in paralog clusters	144	6.62%
Genes assigned to COGs	1,525	70.11%
Genes assigned Pfam domains	1,616	74.30%
Genes with signal peptides	411	18.90%
Genes with transmembrane helices	501	23.03%
CRISPR repeats	2	

Table 4. Number of genes associated with the general COG functional categories

Code	value	%age	Description
J	149	9.0	Translation, ribosomal structure and biogenesis
A	0	0.0	RNA processing and modification
K	64	3.9	Transcription
L	114	6.9	Replication, recombination and repair
B	0	0.0	Chromatin structure and dynamics
D	20	1.2	Cell cycle control, cell division, chromosome partitioning
Y	0	0.0	Nuclear structure
V	25	1.5	Defense mechanisms
T	69	4.2	Signal transduction mechanisms
M	133	8.1	Cell wall/membrane/envelope biogenesis
N	15	0.9	Cell motility
Z	0	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	43	2.6	Intracellular trafficking, secretion, and vesicular transport
O	89	5.4	Posttranslational modification, protein turnover, chaperones
C	131	8.0	Energy production and conversion
G	58	3.5	Carbohydrate transport and metabolism
E	136	8.3	Amino acid transport and metabolism
F	50	3.0	Nucleotide transport and metabolism
H	97	5.9	Coenzyme transport and metabolism
I	37	2.3	Lipid transport and metabolism
P	81	4.9	Inorganic ion transport and metabolism
Q	18	1.1	Secondary metabolites biosynthesis, transport and catabolism
R	183	11.1	General function prediction only
S	136	8.3	Function unknown
-	650	29.9	Not in COGs

Acknowledgements

We would like to gratefully acknowledge the help of Andrea Schütze (DSMZ) for growing *N. salsuginis* cultures. This work was performed under the auspices of the US Department of Energy Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231,

Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396, UT-Battelle and Oak Ridge National Laboratory under contract DE-AC05-000R22725, as well as German Research Foundation (DFG) INST 599/1-2.

References

- Nakagawa S, Takai K, Inagaki F, Horikoshi K, Sako Y. *Nitratiruptor tergaricus* gen. nov., sp. nov. and *Nitratifactor salsuginis* gen. nov., sp. nov., nitrate-reducing chemolithoautotrophs of the epsilon-Proteobacteria isolated from a deep-sea hydrothermal system in the Mid-Okinawa Trough. *Int J Syst Evol Microbiol* 2005; **55**:925-933. [PubMed](#) [doi:10.1099/ijs.0.63480-0](https://doi.org/10.1099/ijs.0.63480-0)
- Takai K, Inagaki F, Nakagawa S, Hirayama H, Nunoura T, Sako Y, Nealson KH, Horikoshi K. Isolation and phylogenetic diversity of members of previously uncultivated ϵ -Proteobacteria in deep-sea hydrothermal fields. *FEMS Microbiol Lett* 2003; **218**:167-174. [PubMed](#)
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie E, Keller K, Huber T, Dalevi D, Hu P, Andersen G. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006; **72**:5069-5072. [PubMed](#) [doi:10.1128/AEM.03006-05](https://doi.org/10.1128/AEM.03006-05)
- Porter MF. An algorithm for suffix stripping. *Program: electronic library and information systems*. 1980; **14**:130-137.
- Teske A, Hinrichs KU, Edgcomb VP, Gomez AD, Kysela DT, Sogin ML, Jannasch HW. Microbial diversity of hydrothermal sediments in the Guaymas Basin: evidence for anaerobic methanotrophic communities. *Appl Environ Microbiol* 2002; **68**:1994-2007. [PubMed](#) [doi:10.1128/AEM.68.4.1994-2007.2002](https://doi.org/10.1128/AEM.68.4.1994-2007.2002)
- Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000; **17**:540-552. [PubMed](#)
- Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics* 2002; **18**:452-464. [PubMed](#) [doi:10.1093/bioinformatics/18.3.452](https://doi.org/10.1093/bioinformatics/18.3.452)
- Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* 2008; **57**:758-771. [PubMed](#) [doi:10.1080/10635150802429642](https://doi.org/10.1080/10635150802429642)
- Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How many bootstrap replicates are necessary? *Lect Notes Comput Sci* 2009; **5541**:184-200. [doi:10.1007/978-3-642-02008-7_13](#)
- Liolios K, Chen IM, Mavromatis K, Tavernarakis N, Hugenholtz P, Markowitz VM, Kyrpides NC. The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2009; **38**:D346-D354. [PubMed](#) [doi:10.1093/nar/gkp848](https://doi.org/10.1093/nar/gkp848)
- Campbell BJ, Smith JL, Hanson TE, Klotz MG, Stein LY, Lee CK, Wu D, Robinson JM, Khouri HM, Eisen JA, *et al.* Adaptations to submarine hydrothermal environments exemplified by the genome of *Natilia profundicola*. *PLoS Genet* 2009; **5**:e1000362. [PubMed](#) [doi:10.1371/journal.pgen.1000362](https://doi.org/10.1371/journal.pgen.1000362)
- Yarza P, Ludwig W, Euzéby J, Amman R, Schleifer KH, Glöckner FO, Rosselló-Mora R. Updates of the All-Species-Living-Tree Project based on 16S and 23S rRNA sequence analyses. *Syst Appl Microbiol* 2010; **33**:291-299. [PubMed](#) [doi:10.1016/j.syapm.2010.08.001](https://doi.org/10.1016/j.syapm.2010.08.001)
- Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, *et al.* The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. [PubMed](#) [doi:10.1038/nbt1360](https://doi.org/10.1038/nbt1360)
- Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. [PubMed](#) [doi:10.1073/pnas.87.12.4576](https://doi.org/10.1073/pnas.87.12.4576)
- Garrity GM, Bell JA, Lilburn T. Phylum XIV. Proteobacteria phyl. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), *Bergey's Manual of Sys-*

- tematic Bacteriology, Second Edition, Volume 2, Part B, Springer, New York, 2005, p. 1.
16. Validation List No. 107. List of new names and new combinations previously effectively, but not validly, published. *Int J Syst Evol Microbiol* 2006; **56**:1-6. [PubMed doi:10.1099/ijs.0.64188-0](#)
 17. Garrity GM, Bell JA, Lilburn T. Class V. *Epsilonproteobacteria* class. nov. In: Bergey's Manual of Systematic Bacteriology, 2nd edn, vol. 2 (The *Proteobacteria*), part C (The *Alpha*-, *Beta*-, *Delta*-, and *Epsilonproteobacteria*), p. 1145. Edited by D. J. Brenner, N. R. Krieg, J. T. Staley & G. M. Garrity. New York: Springer. 2005.
 18. Miroshnichenko ML, L'Haridon S, Schumann P, Spring S, Bonch-Osmolovskaya EA, Jeanthon C, Stackebrandt E. *Caminibacter profundus* sp. nov., a novel thermophile of *Nautiliales* ord. nov. within the class '*Epsilonproteobacteria*', isolated from a deep-sea hydrothermal vent. *Int J Syst Evol Microbiol* 2004; **54**:41-45. [PubMed doi:10.1099/ijs.0.02753-0](#)
 19. BAuA. Classification of bacteria and archaea in risk groups. *TRBA* 2010; **466**:150.
 20. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene Ontology: tool for the unification of biology. *Nat Genet* 2000; **25**:25-29. [PubMed doi:10.1038/75556](#)
 21. Takai K, Campbell BJ, Cary SC, Suzuki M, Oida H, Nunoura T, Hirayama H, Nakagawa S, Suzuki Y, Inagaki F, et al. Enzymatic and genetic characterization of carbon and energy metabolisms by deep-sea hydrothermal chemolithoautotrophic isolates of *Epsilonproteobacteria*. *Appl Environ Microbiol* 2005; **71**:7310-7320. [PubMed doi:10.1128/AEM.71.11.7310-7320.2005](#)
 22. Klenk HP, Göker M. En route to a genome-based classification of *Archaea* and *Bacteria*? *Syst Appl Microbiol* 2010; **33**:175-182. [PubMed doi:10.1016/j.syapm.2010.03.003](#)
 23. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, et al. A phylogeny-driven genomic encyclopaedia of *Bacteria* and *Archaea*. *Nature* 2009; **462**:1056-1060. [PubMed doi:10.1038/nature08656](#)
 24. List of growth media used at DSMZ: http://www.dsmz.de/microorganisms/media_list.php.
 25. Gemeinholzer B, Dröge G, Zetzsche H, Haszprunar G, Klenk HP, Güntsch A, Berendsohn WG, Wägele JW. The DNA Bank Network: the start from a German initiative. *Biopreservation and Biobanking* 2011; **9**:51-55. [doi:10.1089/bio.2010.0029](#)
 26. JGI website. <http://www.jgi.doe.gov>
 27. The Phred/Phrap/Consed software package. <http://www.phrap.com>
 28. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008; **18**:821-829. [PubMed doi:10.1101/gr.074492.107](#)
 29. Han C, Chain P. Finishing repeat regions automatically with Dupfinisher. In: Proceeding of the 2006 international conference on bioinformatics & computational biology. Arabnia HR, Valafar H (eds), CSREA Press. June 26-29, 2006: 141-146.
 30. Lapidus A, LaButti K, Foster B, Lowry S, Trong S, Goltsman E. POLISHER: An effective tool for using ultra short reads in microbial genome assembly and finishing. AGBT, Marco Island, FL, 2008
 31. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**:119. [PubMed doi:10.1186/1471-2105-11-119](#)
 32. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods* 2010; **7**:455-457. [PubMed doi:10.1038/nmeth.1457](#)
 33. Markowitz VM, Ivanova NN, Chen IMA, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; **25**:2271-2278. [PubMed doi:10.1093/bioinformatics/btp393](#)