Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea


*Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195; 1Geosyntec Consultants, Seattle, WA 98101; 2Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331; 3Université de Provence Aix-Marseille I, Laboratoire de Chimie Bactérienne, Centre National de la Recherche Scientifique Unité Propre de Recherche, Marseille, 13402 France; 4Biosciences Division, Lawrence Livermore National Laboratory, Livermore, CA 94550; 5Microbial Program, Joint Genome Institute, Walnut Creek, CA 94598; 6Center for Microbial Ecology, Michigan State University, East Lansing, MI 48824; 7Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064; 8Razavi Newman Center for Bioinformatics, Salk Institute for Biological Studies, La Jolla, CA 92037; 9School of Chemical Sciences, University of Illinois, Urbana, IL 61801; 10Leibniz-Institut für Meereswissenschaften, Kiel, 24105 Germany; 11Water Technology Center, Karlsruhe, 76139 Germany; 12Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019; 13Institut für Chemie und Biologie des Meeres, Universität Oldenburg, Oldenburg, 26129 Germany; 14Departments of Biochemistry, Molecular Biology and Cell Biology, and Chemistry, Northwestern University, Evanston, IL 60208; and 15Biological Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543

Edited by David Karl, University of Hawaii, Honolulu, HI, and approved April 2, 2010 (received for review December 6, 2009)

Ammonia-oxidizing archaea are ubiquitous in marine and terrestrial environments and now thought to be significant contributors to carbon and nitrogen cycling. The isolation of Candidatus “Nitrosopumilus maritimus” strain SCM1 provided the opportunity for linking its chemolithotrophic physiology with a genomic inventory of the globally distributed archaea. Here we report the 1,645,259-bp closed genome of strain SCM1, revealing highly copper-dependent systems for ammonia oxidation and electron transport that are distinctly different from known ammonia-oxidizing bacteria. Consistent with in situ isotopic studies of marine archaea, the genome sequence indicates N. maritimus grows autotrophically using a variant of the 3-hydroxypropionate/4-hydroxybutyrate pathway for carbon assimilation, while maintaining limited capacity for assimilation of organic carbon. This unique instance of archaean biosynthesis of the osmoprotectant ectoine and an unprecedented enrichment of mult copper oxidases, thioredoxin-like proteins, and transcriptional regulators points to an organism responsive to environmental cues and adapted to handling reactive copper and nitrogen species that likely derive from its distinctive biochemistry. The conservation of N. maritimus gene content and organization within marine metagenomes indicates that the unique physiology of these specialized oligophiles may play a significant role in the biogeochemical cycles of carbon and nitrogen.

Marine Group I archaea are among the most abundant microorganisms in the global oceans (1–3). Originally discovered through ribosomal RNA gene sequencing (3, 4), recent metagenomic, biogeochemical, and microbiological studies established the capacity of these organisms to oxidize ammonia, thus linking this abundant microbial clade to one of the key steps of the global nitrogen cycle (5–9). For a century following the discovery of autotrophic ammonia oxidizers, only Bacteria were thought to catalyze this generally rate-limiting transformation in the two-step process of nitrification (10). Despite recent enrichment of mesophilic as well as thermophilic ammonia-oxidizing archaea (AOA) (6, 11, 12), only a single Group I-related strain, isolated from a gravel inoculum from a tropical marine aquarium, has thus far been successfully obtained in pure culture (7).

The isolation of Nitrosopumilus maritimus strain SCM1 ultimately confirmed an archaeal capacity for chemoaautotrophic growth on ammonia. More detailed characterization of this strain revealed cytological and physiological adaptations critical for life in an oligotrophic open ocean environment, most notably one of the highest substrate affinities yet observed (13). Among characterized ammonia oxidizers, only N. maritimus is capable growing at the extremely low concentrations of ammonia generally found in the open ocean (7, 13). This strain therefore provided an excellent opportunity to investigate the core genetic inventory for ammonia-based chemoaautotrophy by Group I crenarchaea.

The genome content and gene order of N. maritimus is highly similar to environmental populations represented in marine bacterioplankton metagenomes, confirming on a genomic level its close relationship to many oceanic crenarchaea. Thus, an evaluation of the genomic inventory of N. maritimus should offer a framework to identify features shared among ammonia-oxidizing Group I crenarchaea, resolve physiological diversity among AOA, and refine understanding of their ecology in relationship to the larger assemblage of marine archaea—not all of which are ammonia oxidizers. In support of this expectation, the physiological and genomic profiles together show that many of the “non-extreme” archaea identified in metagenomic studies, and currently assigned to the Crenarchaeota kingdom, are AOA that contribute to global carbon and nitrogen cycling, possibly determining rates of nitrification in a variety of environments (6, 8, 9, 13).

Results and Discussion

Primary Sequence Characteristics. N. maritimus strain SCM1 contains a single chromosome of 1,645,259 bp encoding 1,997 predicted genes and no extrachromosomal elements or complete prophage sequences (Table 1). No unambiguous origin of replication could be determined on the basis of local gene content or GC skew, as commonly observed for other archaeal genomes (14). Approximately 61% of the N. maritimus open-reading frames together show that many of the


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequence reported in this paper has been deposited in the NCBI database (accession no. NC_010085).

1To whom correspondence should be addressed. E-mail: dastahl@washington.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.0913533107/-/DCSupplemental.
frames (ORFs) could be assigned to clusters of orthologous groups of proteins (COGs), a lower percentage than for genomes of ammonia-oxidizing bacteria (AOB) (Table S1) but similar to *Cenarchaeum symbiosum* (15). The genome possesses a relatively high coding density (91.9%), with a larger fraction dedicated to energy production/conversation, coenzyme transport/metabolism, and translation genes than other characterized *Cenarchaeota*, but similar to two common species of phototrophic marine bacteria, *Prochlorococcus*, and *Synechococcus*.

**Energy Metabolism.** The stoichiometry of ammonia oxidation to nitrite is similar to that of characterized aerobic, obligate chemolithoautotrophic AOB (13), yet the contributing biochemistry is distinctly unique. All AOB share a common pathway where hydroxylamine, produced by an ammonia monooxygenase (AMO), is oxidized to nitrite by a heme-rich hydroxylamine oxidoreductase (HAO) complex; the oxidation of hydroxylamine supplies electrons oxidized to nitrite by a heme-rich hydroxylamine oxidoreductase. Notably, mapping the sequence encoded by *amoB* of *N. maritimus* sequences are no more similar (in either content or sequence identity) between the archaeal and bacterial versions of AMO (7, 27). The corresponding genes appear to be the result of a series of duplications within the *N. maritimus* lineage (Fig. S1).

A second family of predicted redox active periplasmic proteins, composed of 11 thioldisulfide oxidoreductases from the thioredoxin family (Nmar_0639, _0655, _0829, _0881, _1140, _1143, _1148, _1150, _1181, _1658, and _1670), show low (but recognizable) identity with the better characterized disulfide bond oxidases/isomerases found in *Bacillus subtilis* (BdbD) and *Escherichia coli* (DsbA, DsbC, and DsbG). The mean percentage of sequence identity between the *N. maritimus* proteins and BdbD is 21±3%. The significantly lower mean percentage of sequence identity to DsbA, DsbC, and DsbG (9.2, 10.7, and 10.4%, respectively) is comparable to that shared between the *E. coli* proteins (10-11%). Although functional equivalency cannot be established, all but *Nmar_0881* preserve the conserved thioredoxin-like active site FX4CXXC sequence (18–20). In *E. coli*, both DsbA and DsbC rectify nonspecific disulfide bonds catalyzed by copper (21), whereas up-regulation of dsbA by the Cpx regulon occurs during copper stress (22, 23). Eukaryotic protein disulfide isomerase (PDI) homologs sequester and/or reduce oxidized Cu(II), possibly serving as copper acceptors/donors for copper-containing proteins (24). Another described function of PDIs is the capture and transport of nitric oxide (25, 26), a possible intermediate or by-product of ammonia oxidation. The related protein family in *N. maritimus* may function in part to alleviate copper and nitric oxide toxicity.

**Pathways for Ammonia Oxidation and Electron Transfer.** The three genes (Nmar_1500, _1503, and _1502) annotated as *amoA*, *amoB*, and *amoC* and coding for a putative ammonia monooxygenase complex are the only recognizable genetic hallmarks of ammonia oxidation in the genome sequence. However, the *N. maritimus* sequences are no more similar (in either content or organizational structure) to *E. coli* genes than they are to the genes encoding bacterial particulate methane monooxygenases (pMMO), suggestive of functional differences between the archaeal and bacterial versions of AMO (7, 27). Notably, mapping the sequence encoded by *amoB* onto the pmoB crystal structure of *Methylococcus capsulatus* (Bath) (28) reveals the conservation of the ligands to the pMMO metal centers and the complete absence of both a transmembrane helix and a C-terminal cupredoxin domain predicted to be present in bacterial AMO (Fig. S2).

The structural differences in the archaeal AMO, the lack of genes encoding the hydroxylamine–ubiquinone redox module (29), and a periplasm enriched in redox active proteins together suggest significant divergence from the bacterial pathway of ammonia oxidation.
ammonia oxidation. There are two hypothetical mechanistic alternatives (Fig. 1, Table S2): either a unique biochemistry exists for the oxidation of hydroxylamine or the divergent AMO does not actually produce hydroxylamine. If the former is true, hydroxylamine oxidation may occur via one of the periplasmic MCOs (CuHAO). Given the lack of cytchrome c proteins, the four electrons would then be transferred to a quinone reductase (QRED) via small blue copper-containing plastocyanin-like electron carriers. The protein encoded by Nmar_1226, which contains four transmembrane-spanning regions and two plastocyanin-like domains, may serve as an analog of the membrane-bound cytchrome c₅₅₂ quinone reductase present in AOB (29) and is a good candidate for the QRED.

In an alternative scenario, the archaeal AMO produces not hydroxylamine, but the reactive intermediate nitroxyld (nitroxyld hydride, HNO). Nitroxyld is a highly toxic and reactive compound recently recognized as having biological significance in a number of systems (30, 31). During archaeal ammonia oxidation, nitroxyld might be formed by a unique monoxygenase function of archaeal AMO. Alternatively, the archaeal AMO may act as a dioxygenase and insert two oxygen atoms into ammonia, producing nitroxyld from the spontaneous decay of HNOHOH.

Both reaction sequences eliminate the requirement for reductant recycling during the initial oxidation reaction, a simplification offering significant ecological advantage (when compared with AOB) in nutrient poor environments. In this pathway, one of the MCO-like proteins may act as a nitroxyld oxidoreductase (NXOR) and facilitate the oxidation of nitroxyld to nitrite with the extraction of two protons and two electrons in the presence of water. The proposed NXOR would relay the two extracted electrons into the quinone pool via the QRED pathway described above.

In this proposed model, the electrons extracted by either a CuHAO or a NXOR (and transferred into the quinone pool) would generate a proton motive force (PMF) through complexes III (plastocyanin-like subunit, Nmar_1542; Rieske-type subunit, Nmar_1544; transmembrane subunit, Nmar_1543) and IV (Nmar_0182-5), driving the generation of ATP by an F₅F₄-type ATP synthase (Nmar_1688–1693). The production of reductant (i.e., NADH) would require the reverse operation of complex I (NuoABCDDHIKMLN, Nmar_0276–286) as a quinol oxidase driven by a PMF. The proposed biochemistry involving nitroxyld produces the same net gain as bacterial ammonia oxidation, providing two electrons for reduction of the quinone pool and subsequent linear electron flow and the generation of a PMF. The presence of a copper-containing (versus heme) complex III and the unique evolutionary placement of terminal oxidase (complex IV) between two of the heme–copper oxygen reductase families further distinguish this proposed ammonia oxidation pathway from that in AOB.

Carbon Fixation and Mixotrophy. N. maritimus, like all known AOB, grows chemolithoautotrophically by using inorganic carbon as the sole carbon source (7, 32). However, whereas AOB use the Calvin–Bassham–Benson cycle with the CO₂-fixing enzyme ribulose bisphosphate carboxylase/oxygenase (Rubisco) as the key enzyme, the absence of genes in N. maritimus coding for Rubisco and other enzymes of this cycle points to an alternative pathway for carbon fixation. The most likely mechanism supported by the genome sequence is the 3-hydroxypropionate/4-hydroxybutyrate pathway elucidated in the thermophilic crenarchaeote Metallosphaera sedula and suggested as a potential pathway of carbon fixation in C. symbiosum (33). The pathway has two parts: a sequence including two carboxylation reactions transforming acetyl-CoA to succinyl-CoA and a multistep sequence converting succinyl-CoA into two molecules of acetyl-CoA. Genes identified in the N. maritimus genome coding for key enzymes of the pathway (Fig. S3) include a biotin-dependent acetyl-CoA/propionyl-CoA carboxylase (Nmar_0272–0274), methylmalonyl-CoA epimerase and mutate (Nmar_0953, 0954, and 0958), and 4-hydroxybutyrate dehydratase (Nmar_0207). With the exception of one gene (Nmar_1608), all of the genes implicated in the 3-hydroxypropionate/4-hydroxybutyrate pathway for carbon assimilation in N. maritimus are present and show highest similarity to the genes of C. symbiosum (34, 35). Although N. maritimus and M. sedula most likely use the same CO₂-fixation reaction sequences, not all individual reactions appear to be catalyzed by identical enzymes. In one instance, the stepwise reductive transformation of malonyl-CoA to propionyl-CoA involves five enzymes in M. sedula (33, 36, 37). Although the N. maritimus genome lacks any close homologs of the M. sedula genes, it contains alternative alcohol dehydrogenases, aldehyde dehydrogenases, acyl-CoA synthetases, and enoyl-CoA hydratases possibly fulfilling the same functions. Similarly, M. sedula catalyzes the activation of 3-hydroxypropionate to 3-hydroxypropionyl-CoA, using an AMP-forming 3-hydroxypropionyl-CoA synthetase (37). The N. maritimus genome lacks an obvious homolog, although does code for an ADP-forming acyl-CoA synthetase (Nmar_1309) that suggests a more energy efficient alternative.

In addition to the genes coding for the 3-hydroxypropionate/4-hydroxybutyrate pathway, the genome of N. maritimus contains a number of genes encoding enzymes of the tricarboxylic acid (TCA) cycle. No homologs for genes coding for a citrate-cleaving enzyme (ATP citrate lyase or citryl-CoA lyase) were identified, permitting exclusion of the reductive TCA cycle as a pathway for carbon fixation. The lack of these genes suggests that N. maritimus utilizes either an incomplete (or horseshoe-type) TCA cycle for strictly biosynthetic purposes or possibly a complete oxidative TCA cycle.

N. maritimus grows on a completely inorganic medium, indicating the genes coding for essential biosynthetic capacity (SI Materials and Methods), yet its genomic inventory also suggests some flexibility in the utilization of organic sources of phosphorus and carbon. Two systems for phosphorus acquisition are suggested: the high-affinity, high-activity phosphate transport system (pstSCAB, Nmar_0479, Nmar_0481–0483) and a phosphonate transporter (Nmar_0873–0875). However, because the genome lacks genes encoding known C-P lyases and hydrolases (39), and phosphate limitation is not alleviated by supplementation with phosphonates common in the marine environment.

![Fig. 1. Proposed AOA respiratory pathway. Text indicates the described possible hydroxylamine (blue text and arrows) and nitroxyld (green) pathways. Red arrows indicate electron flow not involved in ammonia oxidation. Blue shading denotes blue copper-containing proteins. Pink box indicates possible alternative respiratory electron sink. Hexagons containing Q and QH2 represent the oxidized and reduced quinone pools, respectively.](image-url)
(e.g., aminoethylphosphonate), there is as yet no support for a functional phosphonate utilization pathway. Numerous organic transport functions are also evident, broadly encompassing transporters for different amino acids, dipeptides/oligopeptides, sulfonates/taurine, and glycerol. Additional physiological characterization will likely demonstrate some capacity for mixotrophic growth, as suggested by isotopic studies of natural populations (40–42). No genomic evidence exists for growth on urea, as *N. maritimus* lacks the homologs of the putative urease and urea transporter genes identified in *C. symbiosis* (35).

**Noncoding RNA Genes.** The *N. maritimus* genome contains a full complement of essential noncoding RNA (ncRNA) genes, including one copy each of 55/16S/23S ribosomal RNAs, RNase P, SRP RNA, and 44 transfer RNAs (Table S3). In addition to six normally placed canonical tRNA introns, noncanonical introns were found at different positions in six of the tRNAs (Val_{CAC}, Met, Trp, Arg_{CCT}, Leu_{TAA}, and Glu_{TTC}), a phenomenon previously observed only in thermophiles and *C. symbiosis* (34, 43). All other sequenced crenarchaea (including *C. symbiosis*) contain at least 46 tRNAs. *N. maritimus* lacks tRNA sequences coding for Pro_{TGG} or Arg_{CCG}, perhaps resulting from (or related to) the low G + C content of the genome and preference for protein codons ending in A/T. Other archaeal genomes with relatively low G + C content, such as the euryarchaeon *Haloquadratum walsbyi*, also lack these tRNAs, while possessing the exact complement of 44 tRNAs found in *N. maritimus* (43). This occurrence likely reflects a difference in posttranscriptional modification of the wobble base of tRNAs Pro_{TGG} and Arg_{CCG}, allowing more efficient decoding of the rare codons CCG and CGG, respectively (44).

Six candidates for C/D box small RNAs (sRNAs) were identified (Nmar sR1–sR6). Most C/D box sRNAs guide the precise positioning of posttranscriptional 2′-O-methyl group addition to tRNAs or tRNAs, a process also occurring in eukaryotes, but not *Bacteria*. In *N. maritimus*, predicted targets of 2′-O-methylation may include the wobble base of the Leu_{CAA} tRNA and two different positions separated by 26 nucleotides in the 23S rRNA. Before the *N. maritimus* and *C. symbiosis* genomes, multiple C/D box sRNAs were found almost exclusively in hyperthermophilic archaea (45). Detection of these conserved, syntenic guide sRNAs in the two mesophilic crenarchaeal genomes supports an RNA stabilization/chaperone function not seen in other archaeal mesophiles and possibly more similar to their predicted function(s) in eukaryotes (46).

**Regulation of Transcription.** The genome contains at least eight transcription factor B (TFB) and two TATA-box binding protein (TBP) (Table S3) genes required for starting site-specific transcription initiation, making *N. maritimus* among the densest and richest archaeal genomes for these transcription factors. TFB and TBP are thought to serve functions similar to the bacterial sigma factors (e.g., modulating cellular function in response to fluctuating environmental conditions) in *Archaea* with genomes coding for multiple copies, with optimal TFB/TBP partners (47). Although many other archaeal genomes contain multiple copies of these transcription factors, only the haloarchaea have more than five TFB genes (47). The functional significance of this exceptionally high density of regulatory factors in an apparently metabolically specialized organism will likely be informed by future transcriptional analyses of different growth states. Genes for two widely distributed types of archaeal chromatin proteins are present, an archaeal histone (Nmar 0683) and two Alba genes (Nmar 0255 and Nmar 0933). These are thought to maintain chromosomal material in a state permitting polymerase accessibility, with differential expression possibly providing for altered global transcription (48).

**Unique Cell Division Machinery and Previously Uncharacterized Instance of Archaeal Biosynthesis of Hydroxycetone.** The *N. maritimus* genome contains elements homologous with two systems of cell division: fisZ (Nmar 1262) and cdvABC (Nmar 0700–0816, and 1088). The cdvABC operon, induced at the onset of genome segregation and cell division, codes for machinery related to the euarcheotic endosomal sorting complex (49, 50). With the exception of *N. maritimus*, *C. symbiosis*, and the Thermoproteales (where the cell division machinery remains uncharacterized), all available archaeal genomes have either the FisZ or the Cdv cell division machinery, but not both. The two cell division systems may comprise a hybrid mechanism or serve two distinct processes in marine *Crenarchaeota*.

The genome of *N. maritimus* also encodes for synthesis of the compatible solute hydroxycoetone: ectoine synthase (*ectC; Nmar 1344*), a L-2,4-diaminobutyrate transaminase (*ectB; Nmar 1345*), a L-2,4-diaminobutyrate acetyltransferase (*ectA; Nmar 1346*), and an ectoine hydroxylase (*ectD; Nmar 1343*). Although widely distributed among *Bacteria* (in particular the genome sequences of marine organisms), the presence of these genes in *N. maritimus* represents a unique indication of archaeal biosynthesis.

**Relationship to C. symbiosis.** The genome of *N. maritimus* differs significantly in G + C content and size from that of the closely related sponge symbiont (∼97% 16S rRNA gene sequence identity). Despite the differences in overall genomic G + C content (34.2% for *N. maritimus* versus 57.7% for *C. symbiosis*), the G + C content

---

**Fig. 2.** Synteny plots comparing the *N. maritimus* genome with (A) the *Crenarchaeum symbiosum* A type genome, (B) crenarchaeal genome fragments, and (C) Sargasso Sea fosmids. Vertical gray bar indicates location of ribosomal RNA operon.
for the rRNAs is largely similar between both organisms (50–52%). The higher ORF density (1.19 ORF/kb) relative to C. symbiosum (0.986 ORF/kb) results principally from the 0.4 Mbp smaller genome of N. maritimus, not from a large disparity in the number of predicted ORFs. The two genomes share 1,267 genes in common (when compared via reciprocal BLAST with expectation cutoff values of $10^{-8}$), yet there is little conservation of synteny (Fig. 2C, Table S4). Most of the increased size of the C. symbiosum genome and much of the divergence in gene content are associated with discrete regions (“islands”), although no obvious functionality could be assigned to individual islands. Homologs for a majority of genes implicated in the archaeal ammonia oxidation pathway (51 of 69 listed in Table S2) appear present in N. maritimus.

**Phylogeny and Evolution.** The widely distributed Group I archaeal lineage with which N. maritimus is affiliated was earlier assigned to the hyperthermophilic Crenarchaeota (3). Questions regarding this association arose through phylogenetic analysis of C. symbiosum ribosomal proteins, indicating possible divergence before the Crenarchaeota–Euryarchaeota split and therefore deserving provisional assignment to a new archaeal kingdom, the Thaumarchaeota (51). The basal position of the Group I archaea previously inferred from protein sequences encoded by the C. symbiosum genome was reexamined by reanalysis of the combined dataset, using patterns of gene distribution (Table S3) and phylogeny inference. Maximum-likelihood analyses confirmed the basal branching with significant statistical support (bootstrap value = 90%, Fig. S4). Bayesian analysis of a selection of species from the same dataset produced results linking C. symbiosum and N. maritimus as sister taxa of Crenarchaeota, albeit with nonsignificant support (posterior probability = 0.88). Although a definitive placement within the Archaea still must be confirmed by inclusion of genomes from more distantly related lineages, both analyses strongly support a lineage distinct from all other cultivated Crenarchaeota.

**High Similarity to the Metagenome of the Globally Distributed Marine Group I Archaea.** The genome of N. maritimus shares remarkable conservation of gene content and gene order with numerous archaeal sequences previously recovered in fosmid libraries and recent oceanic surveys. The Antarctic genome fragments DeepAnt-EC39 (taken from 500 m depth) (52) and cosmid 74A4 (from surface waters) (53) both share very high synteny with portions of the N. maritimus genome (Fig. 2B and Table S6) despite significant differences in rRNA sequence identity (93 and 98%, respectively). Sixteen Sargasso Sea contigs (95% 16S rRNA gene sequence identity with N. maritimus) also have high synteny (Fig. 2C and Table S6). Retrieval of DNA fragments from the Global Ocean Sampling (GOS) database using differential protein sequence similarity showed that N. maritimus-like sequences constituted an average of 1.15% of all sequences across a wide range of temperatures (9–29 °C), salinities (freshwater to seawater, 0.1–63 practical salinity units), two open oceans, and several coastal environments (Fig. S4). The Block Island, NY, coastal site and the Lake Gatun, Panama Canal, site (neither of which share any notable physical/chemical characteristics other than sample depth) both exhibited notably large increases in density (>2.5%). The available GOS sequences provided almost complete and uniform coverage of the N. maritimus genome (Fig. 3B), although at least three significant gaps in coverage exist (possibly corresponding with unique N. maritimus genomic islands). Whereas some of the coverage may result in matches to bacterial sequences, particularly for very highly conserved proteins, the majority of recruited proteins had >50% sequence identity to N. maritimus proteins. Together, these shared genomic features suggest N. maritimus is representative of many of the globally abundant marine Group I Crenarchaeotaea and that it should provide a useful model for developing an understanding of the basic physiology of these abundant and cosmopolitan organisms.

A comparison of the sequence content and genome organization hints at functionally more divergent marine population types. N. maritimus has limited syntetic similarity to a deep-water population represented by the North Pacific fosmid 4B7 (93% 16S rRNA sequence identity), but shares proteins highly similar to most of those coded on this fosmid. Previous comparison of marine crenarchaeal genomic fragments reported changes in genomic organization with sampling depths, suggestive of depth-related habitat types (54). Coupled with recent evidence indicating varied physiological lifestyles along depth and latitudinal gradients, distinct crenarchaeal ecotypes likely exist, analogous to that observed in marine cyanobacteria (2, 55). However, no clear correlations currently exist between environmental parameters and the similarity of recruited fragments.

The genome sequence presented here also offers further insight into the ecological success of AOA. For example, using the likely more energy-efficient 3-hydroxypropionate/4-hydroxybutyrate pathway for CO₂ fixation rather than the Calvin–Bassham–Benson cycle used by AOB could provide a growth advantage. Further ecological advantage may be conferred by their potential capacity for mixotrophic growth or the use of copper as a major redox-active metal for respiration in generally iron-limited oceans. However, a deeper understanding of the remarkable success of this archaeal lineage will come only from more detailed physiological, biochemical, and genetic characterization of N. maritimus and additional environmental isolates.

**Materials and Methods**

Genome sequencing was performed on high-molecular-weight DNA extracted from two cultures of N. maritimus. Whole-genome shotgun sequencing of 3-, 8-, and 40-kb DNA libraries by the Joint Genome Institute produced at least 8× coverage. Annotation of the closed genome was performed using Department of Energy (DOE) computational support at Oak Ridge National Laboratory and The Institute for Genomic Research (TIGR) Autounnotation Service in conjunction with Manatee visualization software. Complete details describing high-molecular-weight DNA purification and sequence analysis are found in SI Materials and Methods.

**ACKNOWLEDGMENTS.** The authors thank David Bruce and Paul Richardson from the Joint Genome Institute for facilitating genome sequencing. This
work was supported by the Department of Energy Microbial Genome Program by National Science Foundation Microbial Interactions and Processes Grant MCB-0604448 (to D.A.S. and J.R.D.I.T.), by National Science Foundation Molecular and Cellular Biosciences Grant MCB-0920741 (to D.A.S.), by National Science Foundation Biological Oceanography Grants OCE-0623174 (to D.A.S.) and OCE-0623908 (to S.M.S.), by National Science Foundation Grant EF-0412129 (to M.G.K.), by incentive funds from the University of Louisville VP Research office (to M.G.K.), by the Deutsche Forschungsgemeinschaft (M.K.), by US Department of Agriculture Grant 2010-65115-20380 (to A.C.R.), and by a Salk Institute Innovation Grant (to G.M.).

19. Meima R, et al. (2002) The Walker et al. PNAS Science Foundation Biological Oceanography Grants OCE-0623174 (to D.A.S.) and OCE-0623908 (to S.M.S.), by National Science Foundation Grant EF-0412129 (to M.G.K.), by incentive funds from the University of Louisville VP Research office (to M.G.K.), by the Deutsche Forschungsgemeinschaft (M.K.), by US Department of Agriculture Grant 2010-65115-20380 (to A.C.R.), and by a Salk Institute Innovation Grant (to G.M.).
Supporting Information

Walker et al. 10.1073/pnas.0913533107

SI Materials and Methods

High-Molecular-Weight Genomic DNA Preparation. Two cultures of Nitrospumilus maritimus strain SCM1 were grown as previously described in 500 mL of media in 1-L flasks (1). Cells from both cultures were harvested in late-exponential phase using Sterivex filters for one culture and a 0.1-μm filter for the other. High-molecular-weight DNA was isolated as previously described using either agarose plugs (2) or a modified guanidinium thiocyanate protocol (3). Cells from the Sterivex filters were resuspended in 1 mL of 2x STE buffer [1 M NaCl, 0.1 M EDTA (pH 8.0), 10 mM Tris (pH 8.0)], extracted from the filter, and mixed with 1 vol of 1% molten SeaPlaque LMP agarose (FMC). The mixture was cooled to 40 °C, immediately drawn into a 1-mL syringe, and placed on ice for 10 min. The agarose plug was mixed with 10 mL of lysing buffer, incubated at 37 °C for 1 h, and then transferred to 40 mL of ESP buffer (1% Sarkosyl–1 mg of Proteinase K per mL in 0.5 M EDTA). After incubation at 55 °C for 16 h, the solution was replaced with fresh ESP buffer and incubated at 55 °C for another hour. DNA was purified using phenol:chloroform:isoamyl alcohol (24:24:1) and recovered by precipitation with isopropanol.

Cells collected on the 0.1-μm filter were resuspended in 100 μL of Tris-EDTA (pH 8.0) and 100 mg/mL lysozyme before incubating at 37 °C for 30 min. Then, 3.0 mL of a solution containing 5 M guanidinium thiocyanate, 100 mM EDTA (pH 8.0), and 0.5% (vol/vol) Sarkosyl was added. The solution was mixed gently for 15 min before being cooled on ice for 10 min. After cooling, an equal volume of cold 7.5 M ammonium acetate was added, and the solution was mixed gently and cooled on ice. Purification and precipitation of DNA were performed with chloroform:isoamyl alcohol (24:1) and isopropanol.

Genome Sequencing. A completely sequenced and closed genome of N. maritimus was obtained through collaboration with the Joint Genome Institute. Whole-genome shotgun sequencing of the lysis and precipitation of DNA were performed using the Artemis Comparison Tool (3) with a comparison library generated through WebACT (www.webact.org/WebACT/home). Orthologous genes shared between these two organisms were identified through reciprocal BLAST searches, with an expectation cutoff value of 10−7 and a minimum of 75% alignable N. maritimus sequence. Comparisons with the Global Ocean Sampling (GOS) and Sargasso Sea metagenomic datasets were performed using several single-copy universal archaeal genes to determine a count of ~15 archaeal genomes in the GOS dataset. An initial set of candidate N. maritimus-like proteins was found by BLASTP searches with each N. maritimus protein-coding gene, using a cutoff of 100 hits and a maximum expected value of e = 10. This set consisted of 125,326 peptides drawn from 107,223 scaffolds. Neighboring ORFs on any scaffold with two or more hits were added to make a total of 319,585 peptides, amounting to 5.2% of all ORFs in GOS. These sequences were filtered by BLAST alignment to four sequence datasets, containing 24 complete proteomes from diverse euryarchaeota, crenarchaeota, and bacteria. Sequences scoring more highly to N. maritimus and/or C. symbiosum proteins than any other entry in these datasets were retained, giving a filtered dataset of 21,278 ORFs. Coverage of the N. maritimus proteome was measured by bidirectional BLASTP of N. maritimus proteins against the filtered dataset to assign putative orthologs. The average coverage was ~11x, although some highly conserved genes had a much greater number of hits, probably due to recruitment of nonarchaeal homologs: 48 ORFs had >30 hits, of which most were highly conserved. Searches for CRISPR regions were performed using the Java-based CRISPR recognition tool (CRT) with least stringent settings (4).

Genome Sequence Analysis. Autoannotation of the closed genome sequence was performed by both the Computational Biology group at Oak Ridge National Laboratory (http://genome.ornl.gov/microbial/nmar/02jul07/) and the TIGR Autoannotation Service (now hosted by JCVI; details available from http://www.jcvi.org/cms/research/projects/annotation-service/overview/). The genome visualization software Manatee (release 2.4.1; latest version available from http://manatee.sourceforge.net/) was used for manual curation. Analysis of potential transporter genes was performed using the Transporter Automatic Annotation Pipeline (TransAAP) through the TransportDB genomic comparison tool (membranetransport.org).

Direct comparisons with the C. symbiosum genome and genome fragments were performed using the Artemis Comparison Tool (3) with a comparison library generated through WebACT (www.webact.org/WebACT/home). Orthologous genes shared between these two organisms were identified through reciprocal BLAST searches, with an expectation cutoff value of 10−7 and a minimum of 75% alignable N. maritimus sequence. Comparisons with the Global Ocean Sampling (GOS) and Sargasso Sea metagenomic datasets were performed using several single-copy universal archaeal genes to determine a count of ~15 archaeal genomes in the GOS dataset. An initial set of candidate N. maritimus-like proteins was found by BLASTP searches with each N. maritimus protein-coding gene, using a cutoff of 100 hits and a maximum expected value of e = 10. This set consisted of 125,326 peptides drawn from 107,223 scaffolds. Neighboring ORFs on any scaffold with two or more hits were added to make a total of 319,585 peptides, amounting to 5.2% of all ORFs in GOS. These sequences were filtered by BLAST alignment to four sequence datasets, containing 24 complete proteomes from diverse euryarchaeota, crenarchaeota, and bacteria. Sequences scoring more highly to N. maritimus and/or C. symbiosum proteins than any other entry in these datasets were retained, giving a filtered dataset of 21,278 ORFs. Coverage of the N. maritimus proteome was measured by bidirectional BLASTP of N. maritimus proteins against the filtered dataset to assign putative orthologs. The average coverage was ~11x, although some highly conserved genes had a much greater number of hits, probably due to recruitment of nonarchaeal homologs: 48 ORFs had >30 hits, of which most were highly conserved. Searches for CRISPR regions were performed using the Java-based CRISPR recognition tool (CRT) with least stringent settings (4).

Maximum-Likelihood and Bayesian Trees of the Archaeal Domain. The trees are based on the concatenation of ribosomal proteins used by Brochier-Armanet et al. (5) but including sequences from N. maritimus and from “Candidatus Korarchaeum cryptofilum” OPF8. Sequences were aligned using MUSCLE (6). Resulting alignments were visually inspected and improved with the MUST software (7). Regions where homology between sites was doubtful were removed from further phylogenetic analyses. A total of 6,142 positions were kept for the phylogenetic analyses. The maximum-likelihood tree was computed with PHYML, using the Wagner model corrected by a gamma law to take into account evolutionary rate among site variations (8). The parameter alpha of the gamma distribution as the proportion of invariable sites was estimated from the dataset. The robustness of each branch was estimated by the bootstrap procedure implemented in PHYML. A Bayesian tree analysis on a subset of 29 taxa was performed using MrBayes 3.2 (9) with a mixed model of amino acid substitution and a gamma distribution (eight discrete categories and an estimated proportion of invariant sites) to take into account among-site rate variation. MrBayes was run with four chains for 1 million generations and trees were sampled every 100 generations. To construct the consensus tree, the first 1,500 trees were discarded as “burn-in.” The reduction of the taxonomic sampling was necessary to reduce the computation time.

Walker et al. www.pnas.org/cgi/content/short/0913533107
Walker et al., www.pnas.org/cgi/content/short/0913533107

Fig. 51. Phylogeny of plastocyanin-like protein sequences. Sequences with significant matches to COG3794 (PetE: Plastocyanin [Energy production and conversion]) were used to query the non-redundant protein sequence database from NCBI. Sequences from the top 20–30 non-\( N. maritimus \) hits were retrieved and their match to the above conserved domain model verified. Sequences from experimentally characterized proteins were obtained from the available literature and included, aligned with ClustalW and then curated manually. Distance-based phylogenies were inferred in Phylip using the Neighbor-Joining algorithm and 100 bootstrap replicates. Bootstrap support values >60% are displayed. Nodes with <50% bootstrap support were collapsed.
Fig. S2. Archaeal ammonia monooxygenase AmoB sequence mapped onto the crystal structure of the particulate methane monooxygenase (PDB accession code 1YEW). The pmoA subunit is shown in dark gray, the pmoC subunit in light gray, and the pmoB subunit in red and pink. The red part represents the region of pmoB conserved in the predicted *N. maritimus* AmoB. The transmembrane helix and C-terminal cupredoxin domain shown in pink are missing in the predicted *N. maritimus* AmoB. Cyan spheres represent copper ions. The grey sphere represents a zinc ion.
Fig. S3. Proposed 3-hydroxypropionate/4-hydroxybutyrate cycle for autotrophic carbon fixation by *N. maritimus*. 

Walker et al. www.pnas.org/cgi/content/short/0913533107 4 of 5
Fig. 54. (A) Maximum-likelihood phylogeny of Group 1 Archaea. The phylogeny was inferred using an alignment of concatenated R-proteins (66 taxa, 6,142 positions). WAG+Inv+Gamma (4 classes); 100 replicates. (B) Bayesian tree of mesophilic Group 1 Archaea inferred using an alignment of concatenated R-proteins (29 taxa, 6,142 positions). Mixed model + Gamma (4 classes); 100 replicates.

Other Supporting Information Files

Table S1 (DOC)
Table S2 (DOC)
Table S3 (DOC)
Table S4 (DOC)
Table S5 (DOC)
Table S6 (DOC)