



Genome Sequence of *Candidatus Nitrososphaera evergladensis* from Group I.1b Enriched from Everglades Soil Reveals Novel Genomic Features of the Ammonia-Oxidizing Archaea

Kateryna V. Zhalnina¹, Raquel Dias¹, Michael T. Leonard¹, Patricia Dorr de Quadros², Flavio A. O. Camargo², Jennifer C. Drew¹, William G. Farmerie³, Samira H. Daroub⁴, Eric W. Triplett^{1*}

1 Microbiology and Cell Science Department, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, Florida, United States of America, **2** Soil Science Department, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil, **3** Genome Sequencing Services Laboratory, Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, Florida, United States of America, **4** Everglades Research and Education Center, University of Florida, Belle Glade, Florida, United States of America

Abstract

The activity of ammonia-oxidizing archaea (AOA) leads to the loss of nitrogen from soil, pollution of water sources and elevated emissions of greenhouse gas. To date, eight AOA genomes are available in the public databases, seven are from the group I.1a of the Thaumarchaeota and only one is from the group I.1b, isolated from hot springs. Many soils are dominated by AOA from the group I.1b, but the genomes of soil representatives of this group have not been sequenced and functionally characterized. The lack of knowledge of metabolic pathways of soil AOA presents a critical gap in understanding their role in biogeochemical cycles. Here, we describe the first complete genome of soil archaeon *Candidatus Nitrososphaera evergladensis*, which has been reconstructed from metagenomic sequencing of a highly enriched culture obtained from an agricultural soil. The AOA enrichment was sequenced with the high throughput next generation sequencing platforms from Pacific Biosciences and Ion Torrent. The *de novo* assembly of sequences resulted in one 2.95 Mb contig. Annotation of the reconstructed genome revealed many similarities of the basic metabolism with the rest of sequenced AOA. *Ca. N. evergladensis* belongs to the group I.1b and shares only 40% of whole-genome homology with the closest sequenced relative *Ca. N. gargensis*. Detailed analysis of the genome revealed coding sequences that were completely absent from the group I.1a. These unique sequences code for proteins involved in control of DNA integrity, transporters, two-component systems and versatile CRISPR defense system. Notably, genomes from the group I.1b have more gene duplications compared to the genomes from the group I.1a. We suggest that the presence of these unique genes and gene duplications may be associated with the environmental versatility of this group.

Citation: Zhalnina KV, Dias R, Leonard MT, Dorr de Quadros P, Camargo FAO, et al. (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(7): e101648. doi:10.1371/journal.pone.0101648

Editor: Mark R. Liles, Auburn University, United States of America

Received: January 15, 2014; **Accepted:** June 9, 2014; **Published:** July 7, 2014

Copyright: © 2014 Zhalnina et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the National Science Foundation (grant number MCB-0454030); and the United States Department of Agriculture (grant numbers 2005-35319-16300, 00067345). University of Florida, Interdisciplinary Center for Biotechnology Research, Electron Microscopy and Bio-Imaging lab for assistance with SEM and TEM. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: ewt@ufl.edu

Introduction

The ammonia-oxidizing archaea (AOA) are an abundant group of nitrifiers that plays important environmental roles in the open oceans, soils, the arctic, hot springs and marine sponges [1–8]. AOA oxidize ammonia (NH₃) to nitrite (NO₂⁻) with further oxidation to nitrate (NO₃⁻) by nitrite-oxidizing bacteria [9,10]. In soils, nitrification can increase mobility of inorganic N, hence it may cause NO₃⁻ leaching from soils, pollution of ground and surface waters, and an increased cost of applied N fertilizers in agricultural areas [11–13].

Another possible negative consequence of AOA activity in marine environments and soil, particularly in agricultural areas, is the increased pollution of the atmosphere by nitrous oxide (N₂O). Nitrous oxide is one of the most stable greenhouse gases, and

agricultural soil management is the largest source of N₂O emissions in the United States (69% of total U.S. N₂O emissions) [14]. Several studies demonstrate that AOA produce N₂O [15–18]. However, the underlying pathways for biogeoproduction of N₂O remain unknown.

AOA are difficult to culture. Only a few AOA have been cultured and sequenced from either pure or enrichment cultures [2–6,19,20]. When AOA were first discovered, major AOA groups (I.1a, I.1b and Hot spring cluster) were proposed by 16S rRNA gene and *amoA* gene identities [10]. Group I.1b (or *Nitrososphaera* cluster) is mostly represented by AOA from soil and some other habitats, including hot springs, freshwater, and freshwater sediments [8,21–23]. Group I.1a (or *Nitrosopumilus* cluster) is mainly represented by marine archaea. However it has also been found in other environments including soil, hot springs, and

freshwater [1–6,19,22,24]. Seven genome sequences from group I.1a (*Nitrosopumilus maritimus*, *Candidatus Nitrosopumilus sediminis*, *Candidatus Nitrosopumilus salaria*, *Candidatus Nitrosoarchaeum limnia*, *Candidatus Nitrosopumilus koreensis*, *Cenarchaeum symbiosum*, and *Candidatus Nitrosotenuis uzonensis*) and only one from group I.1b (*Candidatus Nitrososphaera gargensis*) are available in public databases.

The lack of genomic information limits our understanding of the physiology and biochemistry of AOA, particularly Thaumarchaeota from group I.1b. Furthermore, the recently sequenced genome of moderate thermophile *Ca. Nitrososphaera gargensis* from the group I.1b revealed some differences between I.1b and I.1a groups of Thaumarchaeota [20]. For example, the genome size of *Ca. Nitrososphaera gargensis* is much bigger than other sequenced genomes from group I.1a. Additionally, the sequence analysis indicated higher G+C content, more thermosome genes, and a different chemical structure of membrane lipids [20,25]. However, *Ca. Nitrososphaera gargensis* was isolated from hot springs, and it is unknown whether these features are specific only to thermophilic archaeon, or if mesophilic AOA, widely distributed in soils, from group I.1b also share these features.

From a previous study we found that AOA closely related to *Nitrososphaera* genus are highly abundant in the Everglades Agricultural Area, and their abundance significantly increases with agricultural management [26]. In this paper, we present (a) the preparation of an enriched culture of AOA from an Everglades histosol soil; (b) the sequencing and genome reconstruction of the first mesophilic AOA from the group I.1b enriched from the soil; (c) the genome annotation and analysis of main physiological features; and (d) the major metabolic differences between group I.1b and group I.1a.

For the first time, we report genome analysis of AOA group I.1b isolated from soil. This genome provides insight into genomic features present in *Ca. N. evergladensis* but not in other sequenced AOA. The genome analysis reveals features that distinguish AOA from I.1b and I.1a groups. This study provides important insight to guide our understanding of the role of AOA in terrestrial and marine environments.

Results and Discussion

Preparation of ammonia-oxidizing enrichment culture

An AOA enrichment culture was prepared from soil collected from the Everglades Agricultural Area using AOA medium and culture conditions described previously [21]. However, the addition of antibiotics to the enrichment culture did not result in pure culture as some other microorganisms remained. Preliminary genetic analysis of the AOA enrichment was performed by 16S rRNA amplification and Sanger sequencing of the clone library. Approximately 50% of all 16S rRNA clones were assigned to *Nitrososphaera* genus.

Enrichment was tested for the presence of ammonia-oxidizing bacteria (AOB) by PCR-amplification of the bacterial *amoA* genes. No amplification of the bacterial *amoA* was observed. In addition, sequence search of the bacterial *amoA* and 16S rRNA of AOB in the metagenomic sequences of the enrichment was performed against a customized database of bacterial *amoA* sequences and the reference Ribosomal Database Project (RDP) 16S SSU rRNA database [27]. This search did not reveal either bacterial *amoA* or known 16S rRNA genes affiliated with AOB. Further metagenomic analysis of the AOA enrichment showed that all present archaeal *amoA* (12 gene copies) corresponded to *amoA* from the *Ca. N. evergladensis* genome (NTE_00961) at the level of amino-acid identity 99.1–100%. Consistent with these results, all identified

archaeal 16S rRNA (9 gene copies) in the enrichment displayed 99.1–100% of nucleotide identity with *Ca. N. evergladensis* 16S rRNA (NTE_02406).

Ammonia consumption and nitrite (NO_2^-) production, as well as archaeal *amoA* gene copy number, were measured every three days after inoculation (Figure 1A and 1B). Ammonia was converted to NO_2^- over a period of about 21 days (Figure 1A). Simultaneous oxidation of NH_3 and production of NO_2^- was accompanied by the increase of archaeal *amoA* gene copy number (Figure 1B).

Gene prediction and annotation

The genome of the enriched AOA was reconstructed from assembled reads generated using data from the Pacific Biosciences (PacBio) platform (Figure S1A in File S1). Genome assembly was verified by PCR of selected regions in the assembled genome and by alignment of the genome with the contigs obtained from sequencing results from an Ion Torrent platform (Figure S1B in File S1). The 2.95 Mb genome sequence included 3555 genes, 50% G+C content, and 43 RNA genes. Over eighty percent (83.6%) of the assembled bases were predicted to code for proteins. Only 52% of protein coding sequences had functional assignment. Moreover, 60.6% of identified genes were in paralog clusters.

Phylogeny and general genome features of *Ca. N. evergladensis*

Based on 16S rRNA and *amoA* classification, the mesophilic AOA *Ca. N. evergladensis* is phylogenetically affiliated with Thaumarchaeota from group I.1b (Figure 2, Figure S2 in File S1). The closest cultured relatives from the group I.1b are *Ca. N. gargensis*, *N. viennensis*, and *Nitrososphaera* sp. JG1 (Figure 2). *Ca. N. evergladensis* shares 97% and 85% 16S rRNA identity with *Ca. N. gargensis* and the AOA from group I.1a, respectively (Table A in File S1). Nucleotide identity of *amoA* genes were less conserved. *Ca. N. evergladensis amoA* was 87% and 71–74% identical to *Ca. N. gargensis* and group I.1a, respectively. AOA from the group I.1b have larger genomes and almost twice the number of protein coding sequences (CDS) compared to the group I.1a (Table A in File S1). Sixty-four percent of CDS from the *Ca. N. evergladensis* genome share 35% identity with *Ca. N. gargensis* and less than 34% of CDS were found in common with *N. maritimus* (Figure 3A). Overall, I.1a and I.1b groups shared about 30% CDS (Figure 3B). Whole-genome alignment of *Ca. N. evergladensis* to *Ca. N. gargensis* revealed 40% of conserved sites between two genomes. *Ca. N. evergladensis* shared a much smaller degree of genome synteny with *N. maritimus* than *Ca. N. gargensis* (Figures 4A, 4B). An average nucleotide identity of 82.9% between both *Nitrososphaera* genomes confirmed that both genomes represent different species.

Carbon metabolism

The *Ca. N. evergladensis* genome codes for the key enzymes of a 3-hydroxypropionate/4-hydroxybutyrate pathway of CO_2 fixation, enzymes for forward and reverse tricarboxylic acid cycle (TCA) cycle, gluconeogenesis, a modified glycolytic pathway, and hexose monophosphate pathway (Table S1 in File S2, Figures S3, S4, S5, S6 in File S1).

3-hydroxypropionate/4-hydroxybutyrate carbon fixation pathway. *Ca. N. evergladensis*, much like all known chemolithotrophic AOA, is predicted to fix inorganic carbon via a modified 3-hydroxypropionate/4-hydroxybutyrate cycle [3]. Despite the fact that key enzymes for this pathway were found in other sequenced AOA genomes, some steps of the AOA 3-hydroxypropionate/4-hydroxybutyrate cycle remain undescribed

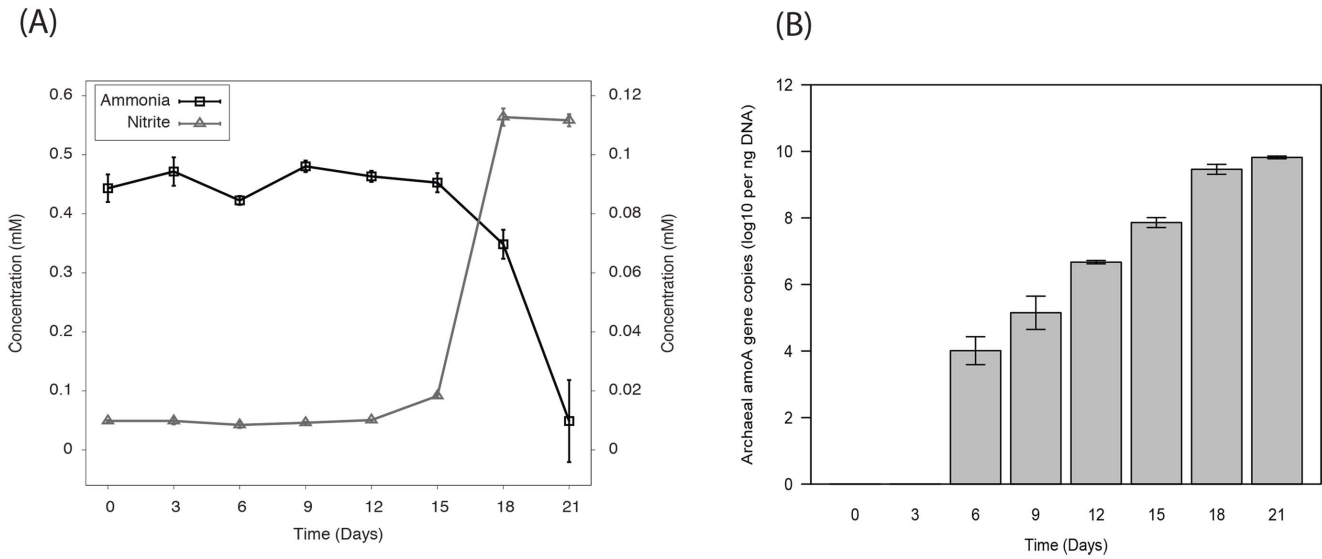


Figure 1. Correlation of ammonia oxidation with the growth of *Ca. Nitrososphaera evergladensis* enrichment culture incubated for 21 days. (A) Concentrations of ammonia and nitrite were determined spectrophotometrically. (B) Archaeal *amoA* gene copies were measured by quantitative PCR. doi:10.1371/journal.pone.0101648.g001

[1,2,20,28]. Genes for the key enzymes for this pathway are found in the *Ca. N. evergladensis* genome (Figure S3 in File S1, Table S1 in File S2). These genes include alpha and beta subunits of acetyl-CoA carboxylase, acetyl/propionyl-CoA carboxylase, methylmalonyl-CoA epimerase and two domains of mutase. Also, biotin-(acetyl-CoA-carboxylase) ligase, which is responsible for assembly

of carboxylase subunits, was identified in *Ca. N. evergladensis*. Candidates for missing enzymes that catalyze reactions of malonyl-CoA to propionyl-CoA were suggested by functional similarity and gene clustering (Table S1 in File S2).

The Thaumarchaeota were recently shown to possess a very efficient, aerobic pathway for CO₂ fixation that differs from that

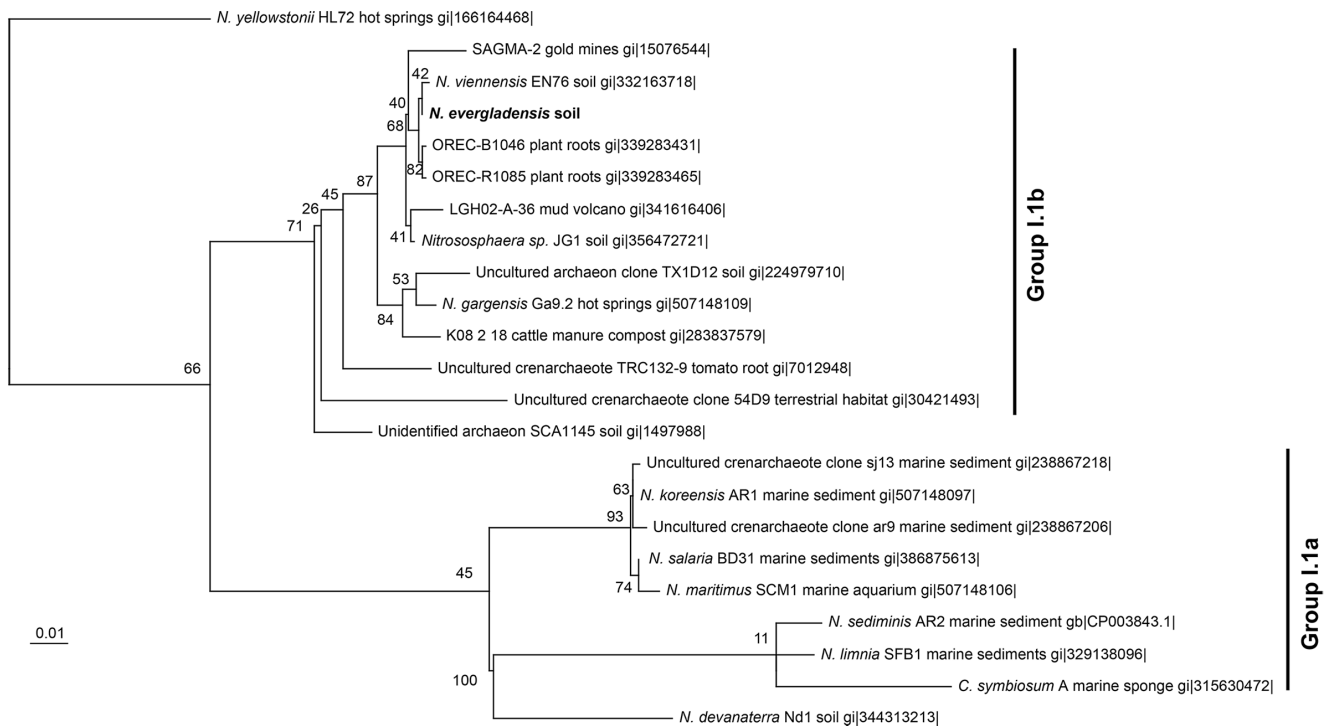


Figure 2. A phylogenetic tree of ammonia-oxidizing archaea 16S rRNA gene sequences (about 1.4 kb). 23 16S rRNA sequences of AOA were randomly selected from the National Center for Biotechnology Information databases. Conservative sites (1.08 kb) were selected using Gblocks. The branching patterns in the maximum-likelihood tree are denoted by their respective bootstrap values (1000 iterations). doi:10.1371/journal.pone.0101648.g002

Table 1. Two component systems (TS) identified in the genomes of AOA representatives from the group I.1a, I.1b and AOB using Conserved Domain Search and TIGRFam database.

Stimulus or related function	TS gene	Group I.1b			Group I.1a			AOB	
		<i>N. everg.</i>	<i>N. garg.</i>	<i>N. kor.</i>	<i>N. lim.</i>	<i>N. marit.</i>	<i>N. europaea</i>		
Cell-envelope stress	<i>BaeS</i>	17	20	8	11	4	6		
	<i>BaeR</i>	0	0	0	0	0	0	0	
Levels of phosphate	<i>PhoR</i>	0(12)*	0(14)	0(6)	0(6)	0(1)	0(5)		
	<i>PhoP(PhoB)</i>	18(3)	21(7)	8(5)	9(9)	11(4)	4(8)		
	<i>PhoU</i>	18(4)	13(5)	9(3)	10(3)	7(2)	4(1)		
Chemotaxis	<i>CheA</i>	2	2	0	0	0	3		
	<i>CheB</i>	2	1	0	1	0	2		
	<i>CheC</i>	2	1	0	1	0	0		
	<i>CheY</i>	5	3	0	1	1	3		
	<i>CheR</i>	0	2	0	3	0	0		
	<i>CheX</i>	1	0	0	1	0	0		
Nitrogen levels	<i>NtrC</i>	1(14)	0(24)	0(8)	0(17)	0(6)	1(7)		
	<i>GlnK</i>	4	4	2	2	2	0		
	<i>NtrB</i>	1(2)	0(2)	0(1)	0	0(1)	1(3)		
Osmotic stress	<i>EnvZ</i>	0	0	0	0	0	0		
	<i>OmpR</i>	1	1	1	0	0	5		
Osmotic stress	<i>MtrB</i>	0	0	0	0	0	0		
	<i>MtrA</i>	0(1)	0(1)	0(1)	0(1)	0(1)	0		
Trimethylamine N-oxide as an alternate electron acceptor	<i>TorS</i>	0(6)	0(7)	0	0(4)	0(4)	0(2)		
	<i>TorR</i>	0	0	0	0	0	0		
Oxidative, weak acid, and osmotic stress	<i>BarA</i>	0	0	0	0	0	0		
	<i>SirA</i>	1	1	1	1	1	2		
Copper resistance (CoprS)	<i>CopS</i>	0	0(1)	0(1)	0(5)	0(6)	0(5)		
	<i>CopR</i>	0(2)	0(3)	0(2)	0(2)	0	0(4)		
Quorum sensing	<i>ComP</i>	0	0	0	0	0	0		
	<i>ComA</i>	0(2)	0(2)	0	0	0	0		
Pyruvate response	<i>YpdA</i>	0(4)	0(6)	0(3)	0(3)	0(3)	0(10)		
	<i>YpdB</i>	0	0	0	0	0	0		
Total		69(52)	70(72)	27(30)	26(50)	26(28)	24(45)		

* components of TS found by TIGRFam database. doi:10.1371/journal.pone.0101648.t001

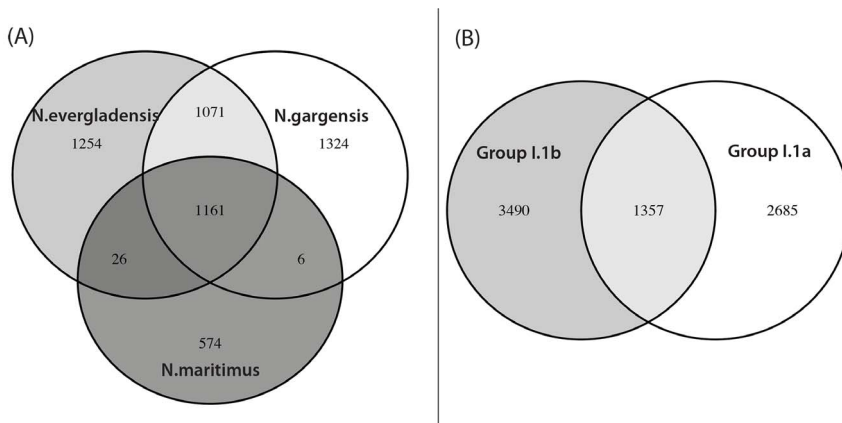


Figure 3. Comparison of protein coding sequences (CDS) of *Ca. Nitrososphaera evergladensis* with CDS of other ammonia-oxidizing archaea. (A) CDS of *Ca. Nitrososphaera evergladensis* were compared to CDS of *Ca. N. gargensis*. (B) CDS of the group I.1a (*N. maritimus*, *Ca. N. sediminis*, *C. symbiosum*, *Ca. N. limnia*, *Ca. N. korensis*) were compared to CDS of the group I.1b (*Ca. N. evergladensis* and *Ca. N. gargensis*). Overlapping regions represent CDS with amino acid sequence identity 35% and higher. doi:10.1371/journal.pone.0101648.g003

found in the Crenarchaeota [29]. The *Ca. N. evergladensis* genome has all eleven of the genes identified to date in this pathway [29].

Tricarboxylic acid cycle. The reductive tricarboxylic acid cycle (TCA) is another potential pathway through which AOA may fix CO₂ autotrophically [30]. Enzymes for both oxidative and reductive TCA cycles are predicted for *Ca. N. evergladensis* (Figure S4 in File S1, Table S1 in File S2). Two genes coding for the two subunits of 2-oxoglutarate oxidoreductase, proposed to catalyze interconversion of α -ketoglutarate and succinyl-CoA in the TCA cycle of the hyperthermophilic crenarchaeote *Thermoproteus tenax* [31,32], were found adjacent to the gene coding for aconitase. Gene homologs for four subunits of the reversible succinate dehydrogenase/fumarate reductase were also detected in *Ca. N. evergladensis*. By contrast, the marine AOA *N. maritimus*

lacks genes coding for citrate lyase and is thus thought to have an incomplete reductive TCA, or only the oxidative TCA cycle [2]. Spang *et al.* (2012) demonstrated the presence of all candidate enzymes for oxidative TCA in the hot spring AOA *Ca. N. gargensis*. Finding of a gene homolog of isocitrate lyase in the *Ca. N. gargensis* genome is evidence of possible usage of the glyoxylate bypass. Replenishment of the TCA intermediates in the *Ca. N. evergladensis* is mediated by either a pyruvate carboxylase or possibly via glyoxylate bypass. Isocitrate lyase as key enzyme of the glyoxylate bypass was identified in the sequenced genome, but malate synthase was not identified.

AOA have a genomic potential to uptake small organic molecules, and the addition of pyruvate stimulates growth of the soil archaeon *N. viennensis* [21]. However, the question of whether AOA are autotrophs or mixotrophs remains unanswered. Green

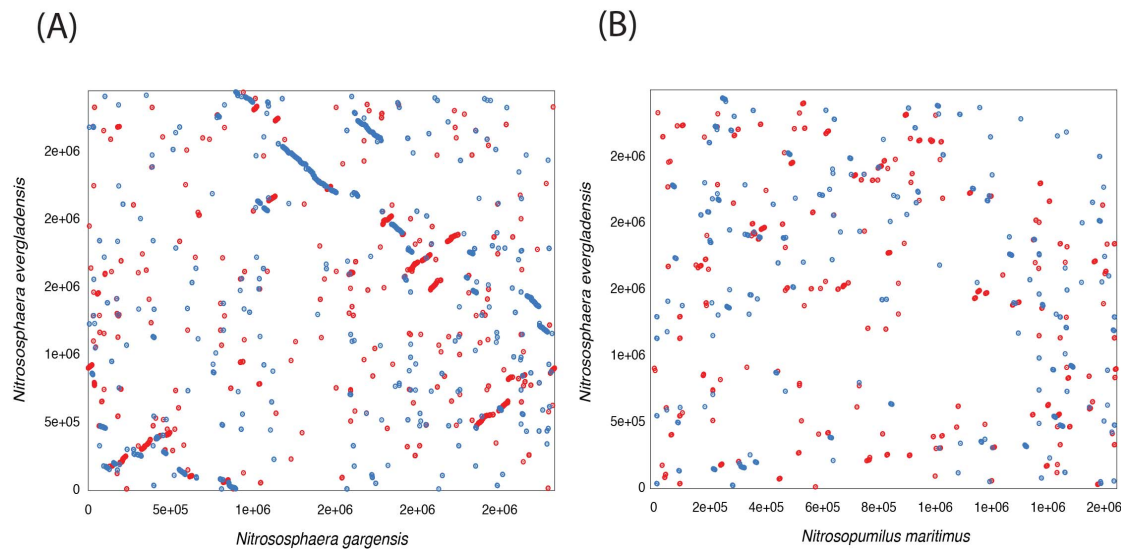


Figure 4. Genome synteny alignments of *Ca. N. evergladensis*, (A) *Ca. N. gargensis* and (B) *N. maritimus*. Axes X and Y represent topology of coding sequences in the comparing genomes. Entire genomes were compared by MUMmer 3.0 package using Promer tool [114]. Each dot represents a match of at least six amino acids from compared genomes. Forward matching amino acid sequences are plotted as red lines/dots while reverse are plotted as blue lines/dots. A line of dots with slope = 1 represents an undisturbed segment of conservation between the two sequences, while a line of slope = -1 represents an inverted segment of conservation between the two sequences. doi:10.1371/journal.pone.0101648.g004

sulfur bacteria can operate the TCA cycle in both directions [33], but compared to autotrophic growth (reductive TCA), green sulfur bacteria prefer mixotrophic growth (oxidative TCA) enhanced with pyruvate. The presence of the gene homologs of the complete oxidative TCA cycle, and encoded amino acid and di- and tricarboxylate transporters in the genome of *Ca. N. evergladensis* suggest the capacity of this AOA to metabolize small organic compounds via this pathway. The presence of 3-hydroxypropionate/4-hydroxybutyrate and reductive TCA cycles in the genome of *Ca. N. evergladensis* may give an advantage to the AOA to survive under different oxygen concentrations. Under the limited oxygen reductive TCA may be used to fix CO₂ and generate NADH. Reductive TCA is an efficient means to fix CO₂ (four ATPs per one molecule of pyruvate) but it is an oxygen sensitive pathway. It may operate under restricted oxygen, where NO₂⁻ produced during aerobic ammonia oxidation may be used as a terminal electron acceptor [34,35]. Conversely, under high oxygen availability AOA may shift to the less efficient but oxygen insensitive 3-hydroxypropionate pathway (five-nine ATPs per one molecule of pyruvate) [34]. Another evidence that suggests potential of AOA to live in the low-oxygen conditions, where they may operate reductive TCA cycle, is high AOA affinities to oxygen determined in the AOA cultures [9,16,36]. If TCA cycle present in *Ca. N. evergladensis* is solely utilized for the biosynthetic purposes, than 3-hydroxypropionate cycle will be the only pathway used for CO₂ fixation in this archaeon. Herein lies the significance of the results of that Thaumarchaeota can fix CO₂ very efficiently under aerobic conditions [29].

Gluconeogenesis and glycolysis. *Ca. N. evergladensis* has a complete gluconeogenic pathway (Figure S5 in File S1, Table S1 in File S2). Archaea operate a variety of modified Embden-Meyerhof-Parnas (EMP) pathways, which differ from the classic glycolytic pathway [37]. Unusual enzymes for glycolysis were found in the genome of *Ca. N. evergladensis* such as multiple kinases (NTE_03124, NTE_00636, NTE_01922) from the ribokinase superfamily that have broad substrate specificity (e. g. glucose, fructose and mannose) and can be candidates for hexokinase and phosphofructokinase enzymes for glycolytic pathway in AOA. The isomerization of glucose-6P/fructose-6P in *Nitrososphaera* may be catalyzed by either metal-dependent phosphoglucose isomerase (NTE_01540), which belongs to the cupin superfamily and found in the Euryarchaeota or with bifunctional phosphoglucose/phosphomannose isomerase from the sugar isomerases family (NTE_02296). A homolog of pyruvate dikinase was detected in the genome (NTE_02861). Pyruvate dikinase catalyzes reversible interconversion of PEP and pyruvate in Thermoprotei [33,37]. Genes encoding glucose-6-phosphate isomerase, sugar kinases and phosphoglycerate kinase were found only in the *Nitrososphaera* species and did not show any close similarity with other Thaumarchaeota.

Hexose monophosphate pathway. All enzyme homologs of the non-oxidative phase of the hexose monophosphate pathway were identified in *Ca. N. evergladensis* except 6-phosphogluconate dehydrogenase is missing. This enzyme is one of the key enzymes of the oxidative phase (Figure S6 in File S1, Table S1 in File S2). This enzyme was found in the genome of *Ca. N. gargensis* but not in other Thaumarchaeota. Two other enzymes of the oxidative branch of HMP, F420-dependent oxidoreductase and gluconolactonase, were identified in the genome. F420-dependent oxidoreductase, G6PDH family has been found in archaeal methanogens, *Streptomyces*, and *Mycobacteria* [38,39]. Similar to other AOA, *Ca. N. evergladensis* does not use the Entner-Doudoroff pathway [3].

Energy metabolism

An autotrophic lifestyle of AOA, in which NH₃ and O₂ are used to generate energy, was demonstrated in multiple studies [1,30]. Recently it was indicated that a group of polar Thaumarchaeota had the genomic potential to use urea to fuel a key step of nitrification [40,41]. *Ca. N. gargensis* [20] and *N. viennensis* [21] showed potential to utilize urea as source of NH₃. Coding sequences for multiple subunits of urease (*ureA*, *ureB*, *ureC*, *ureG*, *ureH*, *ureF*, *ureE*) were found clustered together in the genome with passive and electrochemically-driven urea transporters (Figure 5A). These gene homologs provide evidence that *Ca. N. evergladensis* can use urea as NH₃ source. Moreover, all subunits of urease have two copies in the genome and were identified only in the *Ca. Nitrososphaera* genomes. Some of AOA from the group I.1a (*Ca. Nitrosopumilus* sp. AR2 and *Cenarchaeum symbiosum*) showed similarity with *Nitrososphaera* ureases [1,5,6]. However, the majority of sequenced AOA from the group I.1a (*N. maritimus*, *Ca. N. limnia*, *Ca. N. salaria*, *Ca. N. koreensis*, *Ca. N. uzonensis*) do not have any signatures of urea degradation [1,4,16,19].

Ammonia is oxidized to NO₂⁻ in a two-step reaction. The first reaction is likely catalyzed by archaeal ammonia monooxygenase (AMO). Several genes (*amoA*, *amoB*, *amoC* and *amoX*-like) are predicted to encode subunits of AMO in the *Ca. N. evergladensis* genome (Figure 5B, Table S1 in File S2). The *amoA* gene is the most conserved of the *amo* genes and shares 99% amino acid identity with *amo* genes of *N. viennensis* and 95% with *Ca. N. gargensis* (Table S1 in File S2). Similar to ammonia-oxidizing bacteria (AOB), archaea of group I.1b encode several *amoC* subunits of AMO [42]. However, majority representatives of group I.1a have only one copy of *amoC* (Figure S7 in File S1). Little is known regarding the function of *amoC*. Previous studies have revealed that *amoC* may stabilize the AMO under stress conditions such as starvation and heat shock [43]. It is noteworthy that multiple copies of *amoC* appear more often in AOA and AOB associated with soil environments, which harbor more diverse stressors than marine environments and require more adaptations for organisms to survive and successfully compete [42]. For example, ammonia oxidizers from soil have up to seven *amoC* copies (Figure S7 in File S1), while marine AOA and AOB usually encode only one *amoC* [20,42]. The amino acid alignment identity between seven *amoC* copies ranges between 72% and 97% in the *Ca. N. evergladensis* genome.

Bacterial AMO oxidizes NH₃ to hydroxylamine (NH₂OH) and it is further oxidized to NO₂⁻ by hydroxylamine oxidoreductase (HAO) [44,45]. However, no homologs of the bacterial HAO are found in AOA genomes [1,2,20,28,46]. Two hypothetical pathways of NH₃ oxidation to NO₂⁻ were proposed [2]. The first suggests nitroxyl (HNO) is produced a reactive intermediate that is further oxidized to NO₂⁻ by nitroxyl oxidoreductase. In latter pathway, NH₂OH is a possible intermediate in the reaction and is oxidized to NO₂⁻ by periplasmic multicopper oxidases. Recently, Vajjala et al. [47] provided direct evidence that *N. maritimus* oxidizes NH₂OH to NO₂⁻. Therefore, the alternative pathway with NH₂OH as the intermediate is possible. Similar to other AOA, the *Ca. N. evergladensis* genome encodes genes for six periplasmic multicopper oxidase proteins that may be candidates for HAO (Figure S8 in File S1). Moreover, two of these oxidases are dissimilatory copper-containing nitrite reductases (NirK).

AOB channel two electrons from HAO through cytochromes c₅₅₄ to c_{m552} [44]. Similarly to other AOA no homologs for cytochromes c₅₅₄ and c_{m552} were predicted for *Ca. N. evergladensis*. Instead multiple copper-containing plastocyanin-like electron carriers are candidates for transferring electrons to O₂ (Table S1 in File S2, Figure S8 in File S1). NAD-quinone

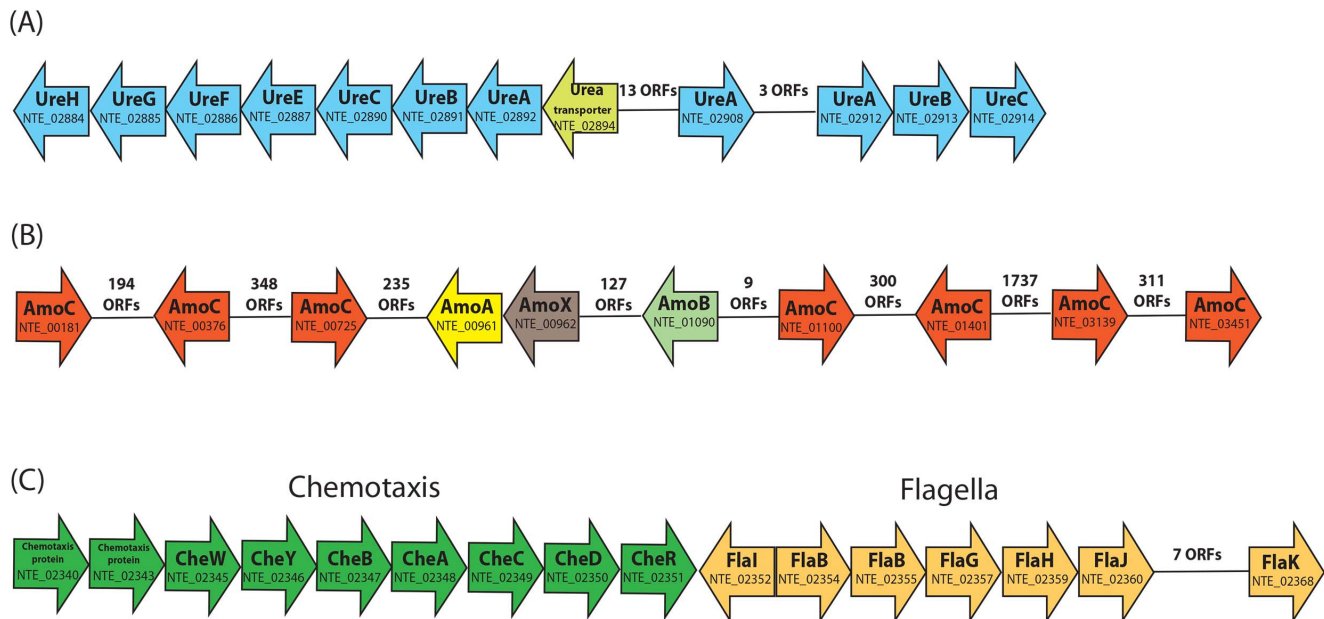


Figure 5. Organization of (A) urease and urea transporter genes; (B) ammonia monooxygenase gene order of *Ca. N. evergladensis*; (C) flagella and chemotaxis genes in the present genome.
doi:10.1371/journal.pone.0101648.g005

oxidoreductase (Complex I) catalyzes the transfer of electrons from NADH to ubiquinone. *Ca. N. evergladensis* has 11 genes encoding subunits of NAD-quinone oxidoreductase, but it is missing genes that encode the E, F and G subunits. Further, a proton motive force (PMF) may be generated through complexes III (Rieske Fe-S proteins, plastocyanines), complex IV (proton-pumping oxygen-reducing plastocyanin-copper oxidases), and by complex V (Archaeal/vacuolar-type H^+ -ATPase). Copper-containing nitrite reductases may reduce NO_2^- to nitric oxide (NO). NO was shown to have a stimulating effect on ammonia oxidation in the AOB [48]. Hence, NO may be involved in the regulation of the AMO activity in AOA.

Triacylglycerols and Polyhydroxyalkanoates as lipid reserve materials

Although many archaea store carbon in the form of polyhydroxyalkanoates (PHAs), some archaea and other organisms preserve carbon in the form of triacylglycerols (TAGs) [49]. *Ca. N. evergladensis* and *Ca. N. gargensis* possess lipases (lysophospholipase, monoglyceride lipase) that may hydrolyze ester bonds in triacylglycerides of long chain fatty acids. Extracellular lipases may also be involved in utilization of monoglycerides from the soil. These lipase homologs are lacking in the group I.1a.

Other lipophilic compounds that likely accumulated in Thaumarchaeota as a reserve material are PHAs. Polyhydroxyalkanoate synthase was found almost in all representatives of Thaumarchaeota [20,50]. The *Ca. N. evergladensis* genome encodes for class III PHA synthase (*phaC*, *phaE*) (Table S1 in File S2). Gene homolog for subunit PhaE of PHA synthase shares some similarity with *Ca. N. gargensis*, but it is very distantly related to representatives of group I.1a.

Isoprenoids as biomarkers for Thaumarchaeota. Archaea use isoprenoids to make phospholipids. The hydrophobic tails of the phospholipids are isoprenoid alcohols ether-linked to glycerophosphate to form monoglycerol-tetraether. Thaumarchaeota have a specific cyclopentane ring-containing dibiphytanyl glycerol tetraether

membrane lipid (crenarchaeol) [51,52]. Damsté et al. [51] hypothesized that formation of cyclohexane ring in crenarchaeol may be an adaptation to cold temperatures in the marine water. However, crenarchaeol was also identified in AOA from thermophilic environments [53]. In addition to crenarchaeol, high concentrations of crenarchaeol regioisomer have been determined in *Ca. N. gargensis*, but this regioisomer is either absent or present in very low amounts in other analyzed AOA from I.1a and ThAOA groups [54]. *Ca. N. evergladensis* has the mevalonate pathway, which operates in archaea and eukaryotes [55]. This pathway is used to synthesize isopentenyl diphosphate (IPP), which is converted to different isoprenoids in the cell (quinones, hydrophobic tails of the phospholipids), using a set of enzymes present in the *Ca. N. evergladensis* genome: farnesyl pyrophosphate synthetase, and octaprenyl pyrophosphate synthetase, undecaprenyl pyrophosphate synthetase (Table S1 in File S2).

Stress response

The sequenced genome revealed the presence of multiple adaptations to survive osmotic and oxidative stress, high concentrations of heavy metals, and elevated temperatures. Moreover, it contains more diverse mechanisms than AOA from the group I.1a to resist high concentrations of heavy metals.

Osmotic stress. One of the strategies to cope with high salinity, and in some cases temperature stress, in archaea is accumulation of compatible solutes, small soluble organic molecules [56]. These solutes can be either transported into the cell or synthesized *de novo*. Several aquaporins that transport water and small uncharged molecules, and belong to the major intrinsic protein family [57], were found in the *Ca. N. evergladensis* genome as well as in other AOA genomes (Table S1 in File S2). Aquaporins identified in *Ca. N. evergladensis* are related to glycerol uptake facilitators. Glycerol is one of the uncharged compatible solutes, which can be used by AOA for osmoadaptation. Mannosyl-3-phosphoglycerate and myo-inositol-1-phosphate synthases were found in the genomes of group I.1b but not in the genomes of group I.1a. These enzymes are involved in the

biosynthesis of di-myo-inositol phosphate and mannosylglycerate, two main prokaryotic compatible solutes. These compatible solutes are commonly represented in thermophilic and hyperthermophilic bacteria and archaea [58].

To mitigate oxidative damage, *Ca. N. evergladensis* encodes superoxide dismutase, peroxiredoxins, and ferritin-like proteins. The majority of enzymes involved in oxidative stress response shared similarity with *Ca. N. gargensis* and other AOA. However, some of the peroxiredoxins and ferritin-like proteins (NTE_01148, NTE_01225, NTE_01156) shared sequence similarity to Euryarchaeota and Bacteria but not to Thaumarchaeota (Table S1 in File S2).

Resistance to heavy metals. Archaea have been found in extreme environments such as mining sites with high concentration of heavy metals [59,60]. *Ca. N. evergladensis* developed mechanisms that would help it resist high external concentrations of metals with at least 21 putative heavy metal resistance proteins. Nine of these homologs were encoded only in *Nitrososphaera* genus, and not in other AOA, and six were found only in the *Ca. N. evergladensis* genome (Table S1 in File S2). Both *Ca. N. evergladensis* and *Ca. N. gargensis* are predicted to have broad tolerance to a variety of heavy metals: copper, zinc, cobalt, cadmium, arsenic, and mercury. However, AOA from group I.1a are more limited in their adaptations to high concentrations of heavy metals.

A higher tolerance of AOA than AOB to copper was shown in soil [61]. Ettema *et al.* (2006) [62] suggested a potential copper resistance gene cluster, which consists of a putative methallochaperone and P-type cation transporting ATPase. This mechanism was identified in thermoacidophilic archaeon *Sulfolobus metallicus*, *Sulfolobus solfataricus*, and *Ferroplasma acidimanus* [63]. A similar gene cluster with encoded P-type ATPases and copper chaperones was found in *Ca. N. evergladensis* (Table S1 in File S2). This mechanism of copper tolerance was also present in *Ca. N. gargensis* genome but not in other sequenced AOA. An alternative putative mechanism of copper detoxification in AOA may involve multicopper oxidases. Multicopper oxidases play an important role in copper resistance in many bacteria [64]. Multiple putative multicopper oxidases are encoded in all known AOA, however, their role in copper tolerance remains unclear (Table S1 in File S2). Periplasmic divalent cation tolerance protein (NTE_02314) is widely represented in AOA, and may also transport copper outside the cell. Copper tolerance may also involve an inorganic polyphosphate transport system [64]. Polyphosphate kinase (PPK), which catalyzes the reversible conversion of the terminal phosphate of ATP into polyphosphates (polyP), and exopolyphosphatase (PPX), is known to hydrolyze polyP. This mechanism was described in other archaea [65]. The enzymes supporting polyP transport encoded in *Ca. N. evergladensis* showed homology to *Methanosarcina*, *Methanoregula*, and *Methanomassiliicoccus* from Euryarchaeota, but this transport is absent from other known Thaumarchaeota.

Ca. N. evergladensis has three putative nickel transporter genes, and one of these high affinity permeases (NTE_02909) is specific only for this thaumarchaeon.

Other metal resistance proteins include cobalt-zinc-cadmium resistance proteins (one is unique for *Ca. N. evergladensis*), putative tellurium resistance membrane protein, and arsenic efflux proteins. Notably, an arsenic pump was identified only in the genomes of *Nitrososphaera* species but not in AOA related to *Nitrosopumilus*.

Heat shock. Besides the altered composition of lipids in the membranes that are used to survive at elevated temperatures, Thaumarchaeota encode an entire set of proteins to cope with

temperature stress [66]. The *Ca. N. evergladensis* genome harbors gene homologs of heat-shock proteins (HSP) such as small HSP, HSP60 (GroEL and Thermosomes), and chaperones such as DnaJ, DnaK and GrpE. Moreover, the copy number of these gene homologs is higher in group I.1b than in group I.1a.

Nitrogen metabolism

Ammonia assimilation. Group I.1b maybe more adapted to high concentrations of NH_3 than group I.1a. The majority of group I.1a AOA were identified in marine environments (*N. maritimus*, *Ca. N. korensis*, *Ca. N. limnia*, *Ca. N. sediminis*, *Ca. N. salaria*) where the ammonium concentration was as low as 0.017 mg L^{-1} [23,67]. Group I.1b is mainly found in the environments at much higher ammonium concentrations ($0.1\text{--}9 \text{ mg L}^{-1}$), such as soil, and representatives of this group are more tolerant of high ammonium levels compared to the majority of AOA isolated from the marine environments or soils with low pH [3,9,16,21,23,36].

NH_3 can be used by *Ca. N. evergladensis* not only as energy source, but also as a N source. The full set of enzymes involved in NH_3 assimilation is present (Figure 6), including glutamate dehydrogenase, glutamine synthetase, and glutamate synthase. Glutamate synthase (NTE_01407) was found only in the *Ca. N. evergladensis* genome, and not other sequenced AOA genomes.

Putative pathway for Nitrous oxide production. A cluster of genes that encode a putative multicopper oxidase related to nitrite reductase (*nirK*) and the gene homologs of the nitric oxide reductase subunits (*norD* and *norQ*) are present in the genome of *Ca. N. evergladensis*. However genes coding for the catalytic subunits (*norB* and *norC*) were not identified. Proximity of the *nirK* and *norD,Q* homologs in the genome may suggest that these genes code for the proteins involved in the same metabolic pathway [68]. A similar set of genes was found in both AOA groups (Table S1 in File S2). Nitrite reductase and nitric oxide reductase were shown to be involved in cell tolerance to NO_2^- and NO [69,70].

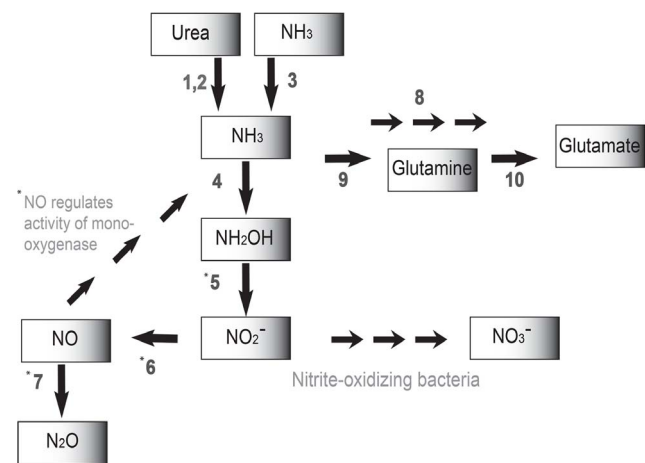


Figure 6. Components of the nitrogen metabolism of *Ca. N. evergladensis*: ammonia oxidation (4, 5), ammonia assimilation (8, 9, 10), nitrite reduction (6), nitrous oxide production (7). Reactions are mediated by the following transporters and enzymes: urea transporters, urease (1, 2), ammonia transporters (3), archaeal ammonia monooxygenase (AMO) (4), candidate enzyme: multicopper oxidase (5), nitrite reductase (*NirK*) (6), nitric oxide reductase (*NorD*, *NorQ*), catalytic subunit (*NorB*) is missing (7), glutamate dehydrogenase (8), glutamine synthetase (9), glutamate synthase (10). NO may upregulate activity of AMO. * - experimental evidences are needed. doi:10.1371/journal.pone.0101648.g006

Alternatively, as in *Nitrosomonas europaea*, *Ca. N. evergladensis* may use NO_2^- and NO as terminal electron acceptors via a putative denitrification pathway [71]. Several studies have shown that AOA cultures are able to emit N_2O and several potential pathways were suggested [15]. However, missing intermediates and missing catalytic subunits of enzymes result in an incomplete pathway for N_2O production, and further experiments are required to determine functional enzymes and intermediates for the N_2O production pathway [17,18].

Transport

In the *Ca. N. evergladensis* genome, 141 transport proteins were identified. (Table S2 in File S2), which is larger than the 89–108 found in group I.1a. Of these, 43 encoded an ATP binding cassette, 17 likely code for pores and channels, and the rest are electrochemical-potential-driven transporters, including the Twin Arginine Translocation system. Twelve of these transporters were not found in other Thaumarchaeota. Di- and tricarboxylate transporters found only in *Ca. N. evergladensis* can be involved in transport of TCA cycle intermediates, such as citrate, malate, and succinate. Thirty-one transporter genes were specific for group I.1b. They include genes coding for mechanosensitive ion channels, urea transporters, a symporter from the major facilitator superfamily (MFS), members of the cation diffusion facilitator family for transport of divalent metals, and members of solute carrier families 5 and 6-like superfamily for co-transport of Na^+ with sugars, amino acids, inorganic ions, or vitamins. Also unique for group I.1b, were proteins from the sodium bile acid symporter family. Transporters of this family were shown to be involved in sodium-dependent transport of a variety of organic molecules in plants and humans [72,73].

Motility, chemotaxis and two-component regulatory systems

At least 69 protein-coding genes related to two-component regulatory systems (TS) were found in *Ca. N. evergladensis* and 70 genes related to TS were found in *Ca. N. gargensis* (Table 1). Notably, group I.1b encodes two and a half times the number of TS genes than group I.1a (Table 1).

Motility-associated genes involved in archaeal flagella and pili assembly were clustered together with protein-coding genes for chemotaxis (Figure 5C). The operon includes genes encoding flagellins (two copies of *flaB*) followed by *fla*-associated genes (*flaG*, *flaH*, *flaJ*, *flaI*, *flaK*, and *flaD*). Structure and assembly of found flagella are closely related to type IV pili. Proteins involved in motility encoded in the *Ca. N. evergladensis* genome were observed in three other AOA, *Ca. N. limnia*, *Ca. N. uzonensis* and *Ca. N. gargensis*. These flagella-coding genes also share identity with other Thermoprotei (*Desulfurococcus kamchatkensis*, *Ignisphaera aggregans*, *Ferriplasma acidophilum*, and *Sulfolobus acidocaldarius*). Adjacent to the *fla*-operon is a set of genes involved in chemotaxis.

Ammonia is the main energy and nitrogen source for AOA. The NtrB/NtrC TS, involved in the response to different NH_3 levels is present in the *Ca. N. evergladensis* genome (Table 1). NtrB senses the nitrogen levels and, under NH_3 limitation, activates NtrC by phosphorylation. NtrC activates expression of glutamine synthetase (GlnA) and it allows cells to grow under nitrogen-limited conditions.

The PhoR/PhoP (PhoB) TS is found in the *Ca. N. evergladensis* genome and in other Archaea and Bacteria [74,75]. It plays an important role for sensing inorganic phosphate levels and under phosphate-limited conditions, PhoR/PhoP activates alkaline phosphatase (Table 1).

The *Ca. N. evergladensis* genome, and other ammonia oxidizers, encode components of TS that provide a response to different environmental stresses, such as cell-envelope stress (BaeS/BaeR) [76], osmotic pressure (EnvZ/OmpR, MtrAB, BarA/SirA), and copper resistance (CopSR) [77].

Ca. N. evergladensis encodes the sensor kinase (YpdA) of TS (YpdA/YpdB) that responds to extracellular pyruvate as a stimulus [78]. This finding supports the hypothesis that pyruvate promotes AOA growth. The response regulator of the TS ComP/ComA, which controls competence in *Bacillus subtilis* via a quorum-sensing mechanism [79], was also found in the group I.1b, but not in the group I.1a.

Information processing machinery

The information processing machinery of *Ca. N. evergladensis* is similar to other Thaumarchaeota, and it shares more homology with eukaryotes than with bacteria [52,80,81] (Table S3 in File S2). The *Ca. N. evergladensis* genome has 61 ribosomal proteins that show a phylum-specific pattern (Table S3 in File S2). One of the specific signatures of the known Thaumarchaeota is that their genomes, including *Ca. N. evergladensis*, are missing gene homologs for r-protein family LXa that is solely present in Archaea. Also, r-proteins L14e and L34e found in other archaeal phyla but not in Thaumarchaeota, are missing from *Ca. N. evergladensis*.

DNA-dependent RNA polymerase II (RNAP) is composed of 12 subunits in *Ca. N. evergladensis* as in *Ca. N. gargensis* [20]. Most of the subunits are homologous to other Archaea [80]. However, in contrast to Euryarchaeota, Crenarchaeota and Nanoarchaeota that have two genes encoding A subunit of RNAP, *Ca. N. evergladensis* and other Thaumarchaeota contain a single *rpoA* gene. This unsplit *rpoA* is common for Eukarya and it was suggested that other archaeal lineages that possess split *rpoA* branched off later in evolution than the Thaumarchaeota [52].

Archaeal RNAP requires two accessory factors: transcription factor B (TFB) (an ortholog of TFIIB), and TATA-box binding protein (TBP) [80]. *Ca. N. evergladensis* has at least nine transcription factors B (TFB), and one TATA-box-binding protein. Other representatives of Thaumarchaeota have a similar number of TFB. For example, *Ca. N. gargensis* and *Ca. N. koreensis* have 11 TFBs, *Ca. N. limnia* encodes at least 9 TFBs, *Ca. N. sediminis* and *N. maritimus* have 10 and 8 TFBs, respectively [2,20,28].

Multiprotein bridging factor 1 (MBF1) is a transcriptional cofactor that bridges the TATA box-binding protein (TBP) and regulatory DNA-binding proteins [81]. MBF1 is a conserved protein present in all eukaryotes and archaea, with exception of the *N. maritimus* and *C. symbiosum* [82]. This protein is found within the soil archaea, *Ca. N. gargensis* [20] and *Ca. N. evergladensis*. Apparently, group I.1b branched off evolutionarily earlier than group I.1a, which lost MBF1 over time.

DNA replication, repair, cell cycle. The *Ca. N. evergladensis* genome contains three *orc1/cdc6* orthologues and one of the *cdc6* orthologs is found only in group I.1b (Table S3 in File S2). *Ca. N. evergladensis* encodes small and large subunits of archaeal DNA polymerase II (pol D) and DNA polymerase type B. Also, it carries genes for the large subunit of replication factor C, both subunits of DNA primase, one copy of archaeal DNA polymerase sliding clamp, DNA ligase, RNase HIII, and flap endonuclease. The gene encoding topoisomerase IB, a signature marker for Thaumarchaeota, but not for other archaeal phyla, is present in the *Ca. N. evergladensis* genome.

Cell division. Among all archaeal phyla only Thaumarchaeota shares two systems of cell division [52,82]. One is FtsZ-

based, present in bacteria and most archaea, and another is CdvABC-based, which is present in Crenarchaeota, and homologous to the eukaryotic ESCRT system [10]. *Ca. N. evergladensis* codes for homologs of CdvA, CdvC (Vps4) and several homologs of CdvB-like proteins (ESCRT-III). Also, the genome has the FtsZ-based division system. *ftsZ* gene homolog encoded in the genome shares 40–63% amino-acid identity with other AOA, and less than 29% with other archaea and bacteria. Another cell division feature shared with eukaryotes and other archaea that is found in *Ca. N. evergladensis* is a homolog of pelota proteins required for meiotic cell division [83]. Pelota homologs are widely represented in archaea [83]. In *Ca. N. evergladensis* pelota homologs have 70% amino acid identity with *Ca. N. gargensis* and 42–50% identity with AOA from the group I.1a (Figure S9 in File S1). In archaea, this protein was suggested to play role in translational elongation, termination, and quality control of mRNA (mRNA surveillance) [84,85]. Archaeal pelota may be involved in the release of the stalled ribosomes and degradation of damaged mRNA [85].

DNA folding and repair. Similar to other Thaumarchaeota, *Ca. N. evergladensis* possesses the genes needed to compress and methylate DNA. The DNA repair system of *Ca. N. evergladensis* includes UvrABC endonuclease, which is common in mesophilic archaea [86]. The soil archaea possess other genes in DNA repair, such as ERCC4-type nuclease and helicase, DNA repair helicase RAD25, nucleotidyltransferase/DNA polymerase involved in DNA repair, and photolyase.

CRISPR-based system

CRISPRfinder determined only one CRISPR locus 7220 bp in the genome of *Ca. N. evergladensis*. The CRISPR region in *Ca. N. evergladensis*, which is longer than that in *Ca. N. gargensis* and many other AOA, was likely a result of more exposure to viruses in its environment [87]. The CRISPR locus consists of 99 repeat/spacer sequences, which is almost 3 times larger than that of *Ca. N. gargensis* [20]. The CRISPR spacers are 34–38 bp and equally separated with identical 37 bp direct repeats. The two genomes of soil group I.1b had higher CRISPR length and longer repeat length than the marine AOA. The five most common CRISPR-associated proteins (*cas1-4*, *cas7*) are adjacent to repeat/spacer sequences. Variable sequences or spacers mostly correspond to segments of captured viral sequences [88]. However, only one CRISPR spacer had significant homology to any virus, the *Helicobacter* phage phiHP33. In bacteria CRISPR-Cas system provides resistance to exogenous genetic elements and provides acquired immunity for the cell [88,89]. Most likely, *Ca. N. evergladensis* utilizes CRISPR in the similar way to maintain genome integrity.

Phylogeny and adaptations

Do the detected differences between genomes of the AOA from the group I.1a and I.1b may give us clues as to how these organisms have adapted to their environments? Many soil surveys revealed that the analyzed soils across the globe were dominated by AOA from the group I.1b (or *Nitrososphaera* cluster), while marine environments were represented by AOA from the group I.1a (or *Nitrosopumilus* cluster) [7,23]. Auguet et al. [23] studied global ecological patterns of Archaea and found that habitat classification was a strong structuring factor of the archaeal communities. Cells of isolated AOA from group I.1a are typically straight rods [3,19,90,91], whereas cultured I.1b archaea are spherically shaped [16,20,21]. Another important physiological difference between two lineages is preference of ammonia concentrations discussed above. Analysis of genomic features of

Nitrososphaera genus may point to other physiological signatures of this group. For example, sequenced AOA representatives from the group I.1b have a larger genome size and higher G+C content than the archaea from the group I.1a.

More than 3,000 CDS are exclusively present in the genomes of Thaumarchaeota from the group I.1b but are absent in the genomes from the group I.1a AOA (Figure 3B). Coding sequences unique to the I.1b archaea included DNA repair proteins, transporters, two-component systems, and information processing machinery (Table S4 in File S2). Enzymes involved in DNA repair unique to *Ca. Nitrososphaera* included DNA repair photolyase, predicted DNA alkylation repair enzyme, an uncharacterized protein predicted to be involved in DNA repair, and replicative and repair DNA polymerase IV (family X). Also, group I.1b uniquely possessed some proteins involved in information processing machinery, such as DNA topoisomerase IA, ribosomal protein L6p, and transcriptional regulators.

The central metabolism of the AOA from the group I.1b is functionally more diverse than that of I.1a group AOA. AOA from the group I.1b have complete TCA cycle and HMP pathways. Also, unlike the I.1a group, the I.1b group seems to be capable of utilizing complex carbohydrates such as glycogen, chitin, and triacylglycerides, as suggested by the presence of genes coding for the glycogen debranching enzyme (NTE_01977), multiple chitin deacetylases, gene homologs for chitinases (NTE_00025, NTE_01408), and monoglyceride lipases. This appears appropriate given the oligotrophic nature of the environments where representatives of I.1a were isolated.

The I.1b group AOA also possess more transporters than I.1a group AOA, such as an ATPase P-type transporter, a urea transporter, a putative hydroxymethylpyrimidine transporter CytX, Di- and tricarboxylate transporters, and transporters from the solute carrier families 5 and 6-like superfamily. These transporters suggest that the compounds transported by these systems are available in the soil environment, but absent or rare in the marine environment.

Several transposases from family IS605 were encoded in the AOA genomes from the group I.1b, but not in the group I.1a. Transposable elements are widely distributed in archaeal genomes, and play an important role in the genome plasticity and response to environmental stimuli [92].

The genomes from the group I.1b also have a higher number of gene duplications compared to the genomes from the group I.1a (Figure 7). Most of the duplicated genes in the group I.1b are involved in adjusting to different environmental conditions, responses to environmental stresses, or efficient nutrient utilization.

Provisional classification and conclusion

In this study we sequenced and analyzed genome of the mesophilic AOA from the group I.1b enriched from the soil. We propose the following *Candidatus* status for this microorganism:

“*Nitrososphaera evergladensis*” sp. nov.

Ethymology. Nitrosus (Latin masculine adjective), nitrous, produces nitrite; sphaera (Latin feminine. n.), spherically shaped; evergladensis (Latin neutrum genitive), isolated from the Everglades.

Locality. Histosol from the Everglades agricultural area.

Diagnosis. Ammonia-oxidizing archaea phylogenetically related to the Thaumarchaeota group I.1b (*Nitrososphaera* cluster) [93]; not isolated; enriched from the agricultural soil.

Analysis of the *Ca. N. evergladensis* genome revealed many similarities of basic metabolism with the rest of AOA, including genes coding for ammonia transporters and genes for AMO subunits, genes for CO₂ fixation via modified hydroxypropionate

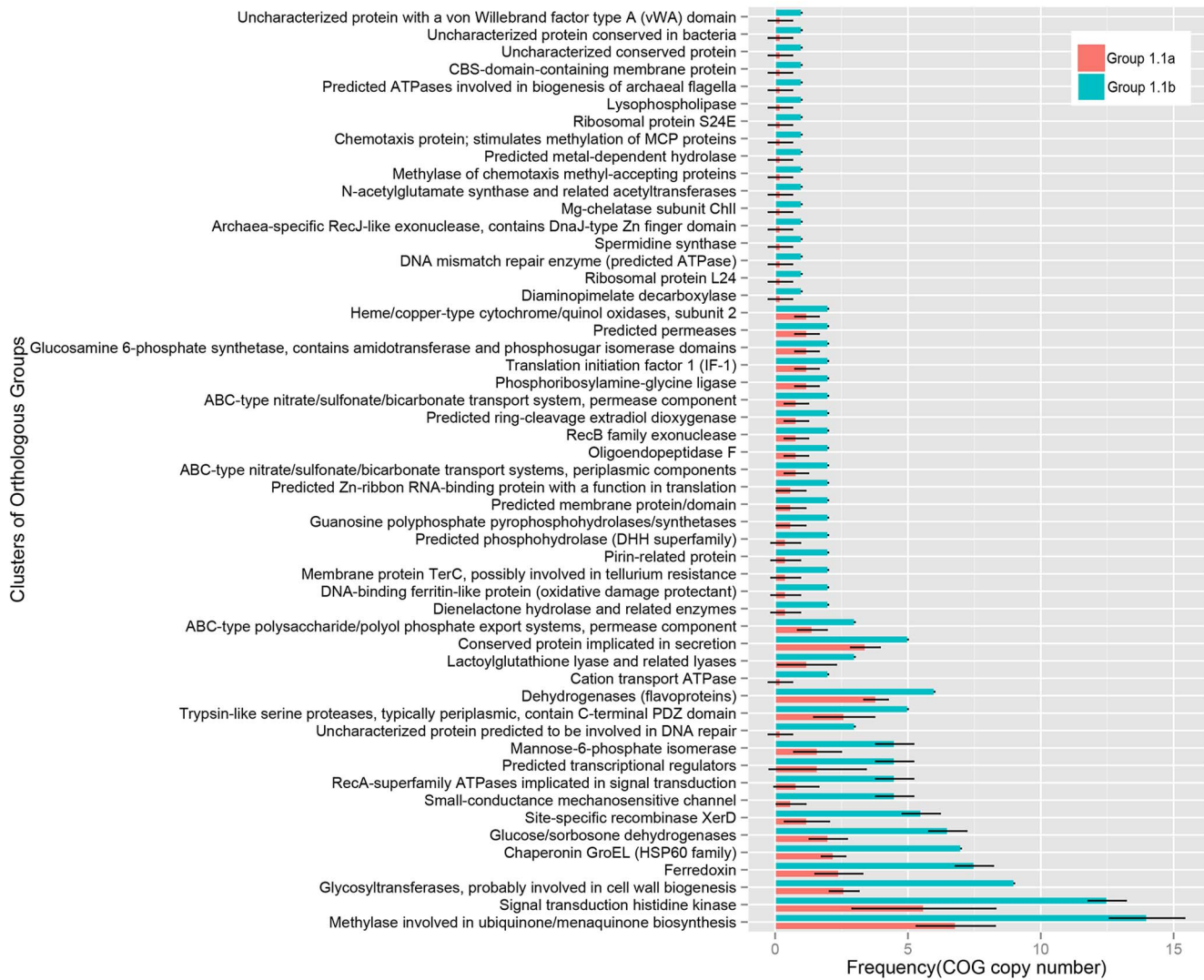


Figure 7. Distribution of Clusters of Orthologous Groups (COGs) in the AOA genomes from soil group I.1b and I.1a. COGs represented with higher copy number in the group I.1b than in the group I.1a were selected by Ttest ($P_{\text{value}} < 0.05$). Error bars display standard deviation. doi:10.1371/journal.pone.0101648.g007

cycle, as well as the HMP, TCA and gluconeogenic pathways. This organism belongs to the group I.1b of the Thaumarchaeota, and shares most of its coding sequences with the closest sequenced relative, *Ca. N. gargensis*, isolated from hot springs. Despite the fact that *Ca. N. evergladensis* is phylogenetically closely related to *Ca. N. gargensis*, they have only 40% of whole genome homology revealing significant differences in the metabolic potential of these organisms. The majority of CDS present in *Ca. N. evergladensis*, but absent in *Ca. N. gargensis*, are hypothetical proteins (Table S5 in File S2). *Ca. N. evergladensis* is also distinct from its closest relative, *Ca. N. gargensis* in that it has a much larger CRISPR region, CRISPR-associated genes, transporters for inorganic and small organic molecules, electron carriers, steroid isomerases, chitin deacetylases, and transcriptional regulators that are completely absent in the *Ca. N. gargensis* genome.

When we compared the genetic potential of the archaeal ammonia oxidizers from group I.1b and group I.1a, the AOA from the group I.1b demonstrated a higher potential to adapt to changes in the environment, and to utilize a broad array of carbon sources compared to the AOA representatives from the group I.1a.

About half of all identified proteins were not assigned to functions and may encode completely novel pathways. Further experiments must be conducted to link novel genes to their specific functions, and determine their ecological role.

Materials and Methods

These soil samples were not collected at a national park or private land. The land is owned by the University of Florida and is within the Everglades Agricultural Area, not the Everglades National Park. No permits were required to collect the soil samples uses in this work. The field studies did not involved endangered or protected species. The GPS coordinates of the research site are: 26.667863, -80.633039.

Enrichment culture

Soil samples for the enrichment were collected from agricultural plots in the Everglades Agricultural Area planted with sugarcane. The soil from this location in the Everglades Agricultural Area is classified as a histosol with pH ~8, moisture ~123%, organic

matter ~70%, nitrate concentration ~54 mg per kg soil, and ammonium concentration ~9 mg per kg soil. To enrich for *Ca. N. evergladensis*, 10 g of soil were resuspended in 0.5 L of the medium for culturing of ammonia-oxidizing archaea (AOA) [21]. The medium contained 0.5 mM NH₄Cl, and 2 ml NaHCO₃ (1M). The headspace above the non-shaking culture was air. One fifth of the enrichment culture was transferred to fresh medium every four weeks for over one year. To further enrich the medium, several antibiotics, including gentamicin (50 µg/ml), tetracycline (5 µg/ml), and erythromycin (10 µg/ml), were applied in order to suppress growth of co-cultured bacteria. However, the addition of antibiotics also affected archaea and did not produce a pure AOA culture. The concentrations of NH₃ and NO₂⁻ were determined by Griess Reagent Kit for Nitrite Determination (G-7921) (Molecular Probes, Eugene, OR, USA), and by the Ammonia Assay Kit (Sigma, St. Louis, MO, USA).

Extraction of DNA

Cells were collected by filtering 1L of culture onto 0.1 µm polycarbonate membrane (Millipore; Billerica, MA, USA). DNA was isolated from the membrane using the PowerSoil DNA Isolation Kit (MO BIO; Carlsbad, CA, USA). Extractions were performed according to the manufacturer's protocol. All genomic DNA concentration and purity were determined by NanoDrop spectrophotometry (Thermo Scientific; Wilmington, DE, USA) and by Qubit 1.27 Fluorometer (Invitrogen; Grand Island, NY, USA).

Quantification of archaeal 16S rRNA and *amoA* genes

Bacterial and archaeal 16S rRNA genes were amplified using universal prokaryotic primers 515F (5'-GTGCCAG-CAGCCGCGGTAA-3') and 806R (5'-GGACTACVSGGG-TATCTAAT-3') [94], cloned into pCR4-TOPO vector and sequenced with M13f and M13r vector primers using Sanger sequencing standard protocol. The archaeal *amoA* copy number in the culture was measured by quantitative PCR (qPCR). Primer sets Arch-amoAf and Arch-amoAr were used [95] (File S1 and S2). Bacterial *amoA* detection was carried out using primer set AmoA1f and AmoA2r [96].

DNA sequencing

Enrichment culture was sequenced using an Ion Torrent Personal Genome Machine (PGM) (Life Technologies; Grand Island, NY, USA), and the Pacific Biosciences platform (Pacific Biosciences; Menlo Park, CA, USA), according to the manufacturer's protocols. Ion Torrent sequencing resulted in 2,389,864 reads with average read length 241 bp (~127X coverage). PacBio platform produced 197,138 reads with an average length 4,117 bp (~179X coverage) (Table B in File S1).

Genome assembly and annotation

Sequenced Ion Torrent reads were imported into CLC Genomics Workbench v.4.0.3 (CLC bio; Aarhus, Denmark), and quality trimmed using a minimum phred score of 20 (with a limit of 5% of low quality bases per read) and a minimum read length of 80 bp. PacBio reads were processed with BLASR mapper (<http://www.pacbiodevnet.com/SMRT-Analysis/Algorithms/BLASR>), and filtered by size. Ion Torrent reads were independently assembled with two *de novo* assemblers IDBA-UD [97] and Mira 3.9 [98], which resulted in 212 contigs with a length up to 41,248 bp, and 24 contigs with a maximum length 418,142 bp, respectively (Table C in File S1). PacBio reads were assembled using Mira 3.9 [98] and Celera from

SMRT portal (<http://www.pacbiodevnet.com/SMRT-Analysis/Software/SMRT-Pipe>) assemblers, which yielded 21 contigs with maximum length 15,072 bp, and one contig 2,954,373 bp, respectively. In addition, all assembly results were compared and verified for errors using Vista [99] and Mauve [100] tools. Custom primers were designed to experimentally confirm complete genome assembly. Random regions with high fluctuations of G+C content, non-coding regions between operons, and regions with contig overlaps were verified by PCR amplifications. The assembled genome was annotated by the Rapid Annotations using Subsystems Technology (RAST) [101] and Expert Review version of the Integrated Microbial Genomes system (IMG ER) [102]. Limited inspection and clean up of coding sequences was done by comparison with the publicly available databases GenBank [103], TIGRfam [104], the database of Clusters of Orthologous Groups of proteins (COGs) [105], and Conserved Domain Database (CDD) [106]. CRISPRFinder was used to identify CRISPR loci [107]. The Conserved Domain Search tool was used for the annotation of two-component systems. The results from two different databases: TIGRfam and CDD were compared and merged together. Detailed information on the annotated genes can be found in Tables S1–S3 in File S1.

Phylogenetic analyses

Amino acid sequences of *amo* genes and nucleotide sequences of 16S rRNA were aligned using MUSCLE 3.8.31 [108]. GBLOCKS were used to select conserved sites and remove poorly aligned regions [109]. Likelihood trees were built using PhyML [110]. The optimized parameters for 16S rRNA and for AMO protein sequences are described in File S1 and S2.

Genome synteny, average nucleotide identity and whole-genome homology

The genome synteny plots were generated from pairwise alignments between the present genome and *Ca. N. gargensis*, and *N. maritimus* genomes obtained from GenBank database. The alignments were based on the six-frame amino acid translations of the compared genomes using Promer tool from MUMmer 3.0 system [111]. The JSpecies software was used to calculate average nucleotide identity between genomes based on the MUMmer ultra-rapid aligning tool [112]. Whole-genome homology was determined from alignment of whole genomes by VISTA servers [99].

Identification of unique coding sequences for group I.1a and I.1b (Venn Diagrams)

Coding sequences (CDS) of six sequenced AOA genomes were downloaded from GenBank. All CDS from the group I.1a were merged together and CDS from the group I.1b were also merged together. Redundant CDS were removed by clustering sequences from the group I.1a and I.1b at 50% identity using UCLUST v1.2.22q and choosing only unique sequence from each cluster [113]. A protein BLAST of sequences from group I.1b versus group I.1a was performed to determine shared, and unique CDS for both groups at ≥35% identity.

Sequence deposition

The genome sequence of "*Ca. Nitrososphaera evergladensis*" has been submitted to GenBank under Accession Number CP007174.

Supporting Information

File S1 Combined file containing Figures S1–S9 and Tables A–C. **Figure S1. Circular representation of the *Ca. Nitrososphaera evergladensis* genome** (A). 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. Alignment between Mira contigs generated from Ion Torrent reads and Celera contig generated from PacBio reads (B). Vertical colored lines indicate a high alignment score and white lines indicate a low score. **Figure S2. A phylogenetic tree of ammonia-oxidizing archaea *amoA*, *amoB*, *amoC*, and *amoX* subunits of ammonia monooxygenase.** Amino-acid sequences of *amo* subunits of AOA were randomly selected from the National Center for Biotechnology Information databases. The multiple sequence alignment of the amino-acid sequences was used for building maximum-likelihood trees. The branching patterns are denoted by their respective bootstrap values (100 iterations). Topology is colored by the metabolic group (blue represents marine group I.1a, green represents group I.1b, red is ThAOA). **Figure S3. 3-Hydroxypropionate cycle.** Identified enzymes in *Ca. N. evergladensis* genome are in green color; missing enzymes are in red color. **Figure S4. TCA cycle.** Identified enzymes in *Ca. N. evergladensis* genome are in green color; missing enzymes are in red color. **Figure S5. Gluconeogenesis/Glycolysis.** Identified enzymes in *Ca. N. evergladensis* genome are in green color; candidates for enzymes are in red color. **Figure S6. Hexose monophosphate pathway (HMP).** Identified enzymes in *Ca. N. evergladensis* genome are in green color; missing enzymes are in red color. **Figure S7. Clustering of the *amo* genes coding for subunits of Ammonia monooxygenase (*AmoA*, *AmoB*, *AmoC*, *AmoX*) in the genomes of ammonia-oxidizing archaea (AOA) and ammonia-oxidizing bacteria (AOB). Figure S8. Electron transport chain of *Ca. N. evergladensis*.** AMO – ammonia monooxygenase; CuHAO – hydroxylamine oxidore-

ductase; NIR- nitrite reductase; NOR – nitric oxide reductase; PC- small blue copper-containing plastocyanin-like electron carriers; Q and QH₂ – oxidized and reduced quinone pools. Complex I - Quinone reductase; Complex III – Riske Fe-S proteins, cytochromes; Complex IV – Heme/copper-type cytochrome/quinol oxidases. * - Suggested candidate enzymes: CuHAO (multicopper oxidase), NOR (catalytic subunits NorD, Q are not found). **Figure S9. A phylogenetic tree of archaeal pelota gene homologs.** Amino-acid sequences of pelota were randomly selected from the National Center for Biotechnology Information databases. The multiple sequence alignment of the amino-acid sequences was used for building maximum-likelihood trees. **Table A.** Comparison of *Ca. N. evergladensis* with other AOA genomes that are available in the public databases. CDS were compared at amino-acid identity $\geq 35\%$. **Table B.** Sequencing reports from the Ion Torrent platform and Pacific Biosciences platform. **Table C.** Comparative results of different assembly methods and sequencing technologies. (PDF)

File S2 Combined file containing Tables S1–S5. **Table S1.** Protein coding sequences of central carbon, nitrogen, lipid metabolism and genes involved in the stress response of the archaeon. **Table S2.** Transporters encoded in *Ca. N. evergladensis* genome. **Table S3.** Information processing machinery of *Ca. N. evergladensis*. **Table S4.** Protein coding sequences (COGs and TIGRfams) present only in the AOA group I.1b. **Table S5.** Coding sequences present only in *Ca. N. evergladensis* genome but missing from the genome of *Ca. N. gargensis*. (XLSX)

Author Contributions

Conceived and designed the experiments: KVZ RD MTL PDDQ SHD EWT. Performed the experiments: KVZ RD MTL PDDQ. Analyzed the data: KVZ RD MTL PDDQ WGF. Contributed reagents/materials/analysis tools: FAOC WGF SHD EWT. Wrote the paper: KVZ RD MTL JCD EWT.

References

- Hallam SJ, Mincer TJ, Schleper C, Preston CM, Roberts K, et al. (2006) Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine *Crenarchaeota*. *PLOS Biol* 4: 0520–0536.
- Walker CB, de la Torre JR, Klotz MG, Urakawa H, Pinel N, et al. (2010) *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc Natl Acad Sci U S A* 107: 8818–8823.
- Jung MY, Park SJ, Min D, Kim JS, Rijpstra WI, et al. (2011) Enrichment and characterization of an autotrophic ammonia-oxidizing archaeon of mesophilic crenarchaeal group I.1a from an agricultural soil. *Appl Environ Microbiol* 77: 8635–8647.
- Mosier AC, Allen EE, Kim M, Ferreira S, Francis CA (2012) Genome sequence of "*Candidatus Nitrosoarchaeum limnia*" BG20, a low-salinity ammonia-oxidizing archaeon from the San Francisco Bay estuary. *J Bacteriol* 194: 2119–2120.
- Park SJ, Kim JG, Jung MY, Kim SJ, Cha IT, et al. (2012a) Draft genome sequence of an ammonia-oxidizing archaeon, "*Candidatus Nitrosopumilus sediminis*" AR2, from Svalbard in the Arctic Circle. *J Bacteriol* 194: 6948–699.
- Park SJ, Kim JG, Jung MY, Kim SJ, Cha IT, et al. (2012b) Draft genome sequence of an ammonia-oxidizing archaeon, "*Candidatus Nitrosopumilus koreensis*" AR1, from marine sediment. *J Bacteriol* 194: 6940–6941.
- Pester M, Rattei T, Fletchl S, Gröngroft A, Richter A, et al. (2012) amoA-based consensus phylogeny of ammonia-oxidizing archaea and deep sequencing of amoA genes from soils of four different geographic regions. *Environ Microbiol* 14: 525–539.
- Radax R, Hoffmann F, Rapp H, Leininger S, Schleper C (2012) Ammonia-oxidizing archaea as main drivers of nitrification in cold-water sponges. *Environ Microbiol* 14: 909–23.
- Martens-Habbena W, Berube PM, Urakawa H, Torre JR, Stahl DA (2009) Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature* 461: 976–979.
- Stahl DA, de la Torre JR (2012) Physiology and diversity of ammonia-oxidizing archaea. *Annu Rev Microbiol* 66: 83–101.
- Raun WR, Johnson GV (1999) Improving nitrogen use efficiency for cereal production. *Agron J* 91: 357–363.
- Kowalchuk GA, Stephen JA (2001) Ammonia-Oxidizing Bacteria: A Model for Molecular Microbial Ecology. *Annu Rev Microbiol* 55: 485–529.
- Zhalnina K, de Quadros PD, Camargo FA, Triplett EW (2012) Drivers of archaeal ammonia-oxidizing communities in soil. *Front Microbiol* 3: 210.
- Inventory of US Greenhouse Gas Emissions and Sinks: 1990–2011 Available: <http://epagov/climatechange/Downloads/ghgemissions/US-GHG-Inventory-2013-Main-Text.pdf>.
- Stügelmeier M, Mooshammer M, Kitzler B, Wanek W, Zechmeister-Boltenstern S, et al. (2014) Aerobic nitrous oxide production through N-nitrosating hybrid formation in ammonia-oxidizing archaea. *ISME J* doi:10.1038/ismej.2013.220
- Kim JG, Jung MY, Park SJ, Rijpstra WIC, Damsté JS, et al. (2012) Cultivation of a highly enriched ammonia-oxidizing archaeon of thaumarchaeotal group I.1b from an agricultural soil. *Environ Microbiol* 14: 1528–1543.
- Loscher CR, Kock A, Konneke M, LaRoche J, Bange HW, et al. (2012) Production of oceanic nitrous oxide by ammonia-oxidizing archaea. *Biogeochemistry* 9: 2419–2429.
- Santoro AE, Buchwald C, McIlvin MR, Casciotti KL (2011) Isotopic Signature of N₂O produced by marine ammonia-oxidizing archaea. *Science* 333: 1282–1285.
- Lebedeva E, Hatzepichler R, Pelletier E, Schuster N, Hauzmayr S, et al. (2012) Enrichment and genome sequence of the group I.1a ammonia-oxidizing Archaeon "*Ca. Nitrosotenuis uzonensis*" representing a clade globally distributed in thermal habitats. *PLOS ONE* 8: e80835.
- Spang A, Phoehlein A, Offre P, Zumbagel S, Haider S, et al. (2012) The genome of the ammonia-oxidizing *Candidatus Nitrososphaera gargensis*: insights into metabolic versatility and environmental adaptations. *Environ Microbiol* 14: 3122–3145.

21. Tourna M, Stieglmeier M, Spang A, Könneke M, Schintmeier A, et al. (2011) *Nitrososphaera viemensis*, an ammonia oxidizing archaeon from soil. Proc Natl Acad Sci U S A 108: 8420–8425.
22. Ochseneiter T, Selezi D, Quaiser A, Bonch-Osmolovskaya L, Schleper C (2003) Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. Environ Microbiol 5: 787–797.
23. Auguet J-C, Barberan A, Casamayor E (2010) Global ecological patterns in uncultured Archaea. ISME J 4: 182–90.
24. Mosier A, Allen E, Kim M, Ferreira S, Francis C (2012) Genome sequence of “*Candidatus Nitrosopumilus salaria*” BD31, an ammonia-oxidizing archaeon from the San Francisco Bay estuary. J Bacteriol 194: 2121–2.
25. Damsté JS, Rijpstra WIC, Hopmans EC, Jung MY, Kim JG, et al. (2012) Intact polar and core glycerol dibiphytanyl glycerol tetraether lipids of Group I.1a and I.1b Thaumarchaeota in soil. Appl Environ Microbiol 78: 6866–6874.
26. Zhalnina K, de Quadros PD, Gano KA, Davis-Richardson A, Fagen JR, et al. (2013) *Ca. Nitrososphaera* and *Bradyrhizobium* are inversely correlated and related to agricultural practices in long-term field experiments. Front Microbiol 4: 104.
27. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, et al. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 37: D141–5.
28. Blainey PC, Mosier AC, Potanina A, Francis CA, Quake SR (2011) Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. PLOS ONE 6: 1–12.
29. Könneke M, Schubert DM, Brown PC, Hügl M, Standfest S, et al. (2014) Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO₂ fixation. Proc Natl Acad Sci (USA) early edition, doi:10.1073/pnas.1402028111
30. Zhang L, Offire PR, He JZ, Verhamme DT, Nicol GW, et al. (2010) Autotrophic ammonia oxidation by soil thaumarchaea. Proc Natl Acad Sci U S A 107: 17240–17245.
31. Zaparty M, Zaigler A, Stamme C, Soppa J, Hensel R, et al. (2008) DNA microarray analysis of central carbohydrate metabolism: glycolytic/glucconeogenic carbon switch in the hyperthermophilic crenarchaeum *Thermoproteus tenax*. J Bacteriol 190: 2231–8.
32. Siebers B, Tjaden B, Michalke K, Dörr C, Ahmed H, et al. (2004) Reconstruction of the central carbohydrate metabolism of *Thermoproteus tenax* by use of genomic and biochemical data. J Bacteriol 186: 2179–94.
33. Tang KH, Blankenship RE (2010) Both forward and reverse TCA cycles operate in green sulfur bacteria. J Biol Chem 285: 35848–35854.
34. Berg I, Kockelkorn D, Ramos-Vera W, Say RF, Zarzycki J, et al. (2010) Autotrophic carbon fixation in archaea. Nat Reviews Microbiol 8: 447–60.
35. Shan Y, Lai Y, Yan A (2012) Metabolic reprogramming under microaerobic and anaerobic conditions in bacteria. Subcell Biochem 64: 159–79.
36. Park BJ, Park SJ, Yoon DN, Schouten S, Sinninghe Damsté JS, et al. (2010) Cultivation of autotrophic ammonia-oxidizing archaea from marine sediments in coculture with sulfur-oxidizing bacteria. Appl Environ Microbiol 76: 7575–87.
37. Siebers B, Schönheit P (2005) Unusual pathways and enzymes of central carbohydrate metabolism in Archaea. Curr Opin Microbiol 8: 695–705.
38. Cheeseman P, Toms-Wood A, Wolfe RS (1972) Isolation and properties of a fluorescent compound, factor 420, from *Methanobacterium* strain MoH. J Bacteriol 112: 527–531.
39. Taylor M, Scott C, Grogan G (2013) F420-dependent enzymes - potential for applications in biotechnology. Trends Biotechnol 31: 63–4.
40. Alonso-Sáez L, Waller AS, Mende DR, Bakker K, Farnelid H, et al. (2012) Role for urea in nitrification by polar marine Archaea. Proc Natl Acad Sci U S A 109: 17989–17994.
41. Kirchman DL (2012) Marine archaea take a short cut in the nitrogen cycle. Proc Natl Acad Sci U S A 109: 17732–17733.
42. Norton JM, Alzerreca JJ, Suwa Y, Klotz MG (2002) Diversity of ammonia monooxygenase operon in autotrophic ammonia-oxidizing archaea. Arch Microbiol 177: 139–149.
43. Berube PM, Stahl DA (2012) The divergent AmoC subunit of ammonia monooxygenase functions as part of a stress response system in *Nitrosomonas europaea*. J Bacteriol 194: 3448–3456.
44. Arp D, Sayavedra-Soto L, Hommes N. (2002) Molecular biology and biochemistry of ammonia oxidation by *Nitrosomonas europaea*. Arch Microbiol 178: 250–255.
45. Klotz MG, Stein LY (2008) Nitrifier genomics and evolution of the nitrogen cycle. FEMS Microbiol Lett 278: 146–156.
46. Schleper C, Nicol GW (2010) Ammonia-oxidizing archaea-physiology, ecology and evolution. Adv Microb Physiol 57: 1–41.
47. Vajrala N, Martens-Habben W, Sayavedra-Soto LA, Schauer A, Bottomley PJ, et al. (2012) Hydroxylamine as an intermediate in ammonia oxidation by globally abundant marine archaea. Proc Natl Acad Sci U S A 110: 1006–1011.
48. Zart D, Schmidt I, Bock E. (2000) Significance of gaseous NO for ammonia oxidation by *Nitrosomonas europaea*. Antonie Van Leeuwenhoek 77: 49–55.
49. Murphy DJ (2012) The dynamic roles of intracellular lipid droplets: from archaea to mammals. Protoplasma 3: 541–585.
50. Quillaguaman J, Guzman H, Van-Thuoc D, Hatti-Kaul R (2010) Synthesis and production of polyhydroxyalkanoates by halophiles: current potential and future prospects. Appl Microbiol Biotechnol 85: 1687–1696.
51. Damsté JS, Schouten S, Hopmans EC, van Duin AC, Geenevasen JA (2002) Crenarchaeol: the characteristic core glycerol dibiphytanyl glycerol tetraether membrane lipid of cosmopolitan pelagic crenarchaeota. J Lipid Res 43: 1641–1651.
52. Spang A, Hatzepichler R, Brochier-Armanet C, Rattai T, Tischler P, et al. (2010) Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. Trends Microbiol 18: 331–340.
53. de la Torre J, Walker C, Ingalls A, Könneke M, Stahl D (2008) Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. Environ Microbiol 10: 810–8.
54. Pitcher A, Rychlik N, Hopmans E, Spieck E, Rijpstra W, et al. (2010) Crenarchaeol dominates the membrane lipids of *Candidatus Nitrososphaera gargensis*, a thermophilic group I.1b Archaeon. ISME J 4: 542–52.
55. Smit A, Mushegian A (2000) Biosynthesis of isoprenoids via mevalonate in archaea: The Lost Pathway. Genome Res 10: 1468–1484.
56. Muller V, Spanheimer R, Santos H (2005) Stress response by solute accumulation in archaea. Curr Opin Microbiol 8: 729–736.
57. Tanghe A, Van Dijk P, Thevelein JM (2006) Why do microorganisms have aquaporins? Trends Microbiol 14: 78–85.
58. Roesser M, Müller V (2001) Osmoadaptation in bacteria and archaea: common principles and differences. Environ Microbiol 3: 743–754.
59. Maezato Y, Blum P (2012) Survival of the fittest: overcoming oxidative stress at the extremes of acid, heat and metal. Life 2: 229–242.
60. Pikuta EV, Hoover RB, Tang J (2007) Microbial extremophiles at the limits of life. Crit Rev Microbiol 33: 183–209.
61. Li X, Zhu YG, Cavagnaro TR, Chen M, Sun J, et al. (2009) Do ammonia-oxidizing archaea respond to soil Cu contamination similarly to ammonia-oxidizing bacteria? Plant Soil 324: 209–217.
62. Ettema TJ, Brinkman AB, Lamers PP, Kornet NG, de Vos WM, et al. (2006) Molecular characterization of a conserved resistance (*cop*) gene cluster and its copper-responsive regulator in *Sulfolobus solfataricus* P2. Microbiology 152: 1969–1979.
63. Orell A, Remonsellez F, Arancibia R, Jerez CA (2013) Molecular characterization of copper and cadmium resistance determinants in the biomining thermoacidophilic archaeon *Sulfolobus metallicus*. Archaea doi:10.1155/2013/289236
64. Rowland JL, Niederweis M (2013) A multicopper oxidase is required for copper resistance in *Mycobacterium tuberculosis*. J Bacteriol 195: 3724–3733.
65. Remonsellez F, Orell A, Jerez CA (2006) Copper tolerance of the thermoacidophilic archaeon *Sulfolobus metallicus*: possible role of polyphosphate metabolism. Microbiology 152: 59–66.
66. Petitjean C, Moreira D, López-García P, Brochier-Armanet C (2011) Horizontal gene transfer of a chloroplast DnaJ-Fer protein to Thaumarchaeota and the evolutionary history of the DnaK chaperone system in Archaea. BMC Evol Biol 12: 226.
67. Rees A, Woodward M, Joint I (1999) Measurement of nitrate and ammonium uptake at ambient concentrations in oligotrophic waters of the North-East Atlantic Ocean. Mar Ecol Prog Ser 187: 295–300.
68. Yin Y, Zhang H, Olman V, Xu Y (2010) Genomic arrangement of bacterial operons is constrained by biological pathways encoded in the genome Proc Natl Acad Sci USA 107: 6310–5.
69. Beaumont HJE, Hommes NG, Sayavedra-Soto LA, Arp DJ, Arciero DM, et al. (2002) Nitrite reductase of *Nitrosomonas europaea* is not essential for production of gaseous nitrogen oxides and confers tolerance to nitrite. J Bacteriol 184: 2557–2560.
70. Beaumont HJE, van Schooten B, Lens SI, Westerhoff HV, van Spanning RJ (2004) *Nitrosomonas europaea* expresses a nitric oxide reductase during nitrification. J Bacteriol 186: 4417–4421.
71. Whittaker M, Bergmann D, Arciero D, Hooper AB (2000) Electron transfer during the oxidation of ammonia by the chemolithotrophic bacterium *Nitrosomonas europaea*. Biochim Biophys Acta 1459: 346–355.
72. Döring B, Lütteke T, Geyer J, Petzinger E (2012) The SLC10 carrier family: transport functions and molecular structure. Curr Top Membr 70: 105–168.
73. Furumoto T, Yamaguchi T, Ohshima-Ichii Y, Nakamura M, Tsuchida-Iwata Y, et al. (2011) A plastidial sodium-dependent pyruvate transporter. Nature 476: 472–475.
74. Sola-Landa A, Moura RS, Martin JF (2003) The two-component PhoR-PhoP system controls both primary metabolism and secondary metabolite biosynthesis in *Streptomyces lividans*. Proc Natl Acad Sci U S A 100: 6133–6138.
75. Wende A, Furtwangler K, Oesterheld D (2009) Phosphate-dependent behavior of the archaeon *Halobacterium salinarum* Strain R1. J Bacteriol 191: 3852–3860.
76. Leblanc SKD, Oates CW, Raivio TL (2011) Characterization of the induction and cellular role of the BacSR two-component envelope stress response of *Escherichia coli*. J Bacteriol 193: 3367–3375.
77. Giner-Lamia J, López-Maury L, Reyes JC, Florencio EJ (2012) The CopRS two-component system is responsible for resistance to copper in the cyanobacterium *Synechocystis* sp. PCC 6803. Plant Physiol 159: 1806–1818.
78. Fried L, Behr S, Jung K (2013) Identification of a target gene and activating stimulus for the YpdA/YpdB histidine kinase/response regulator system in *Escherichia coli*. J Bacteriol 195: 807–815.
79. Tortosa P, Logsdon L, Kraigher B, Itoh Y, Mandic-Mulec I, et al. (2001) Specificity and genetic polymorphism of the *Bacillus* competence quorum-sensing system. J Bacteriol 183: 451–460.

80. Korkhin Y, Unligil UM, Littlefield O, Nelson PJ, Stuart DI, et al. (2009) Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. *PLOS Biol* 7: 0001–0010.
81. de Koning, Blombach F, Wu H, Brouns SJJ, Oost JVD (2009) Role of multiprotein bridging factor 1 in archaea: bridging the domains? *Biochem Soc Trans* 37: 52–57.
82. Brochier-Armanet C, Forterre P, Gribaldo S (2011) Phylogeny and evolution of the Archaea: one hundred genomes later. *Curr Opin Microbiol* 14: 274–281.
83. Ragan MA, Logsdon JM, Sensen CW, Charlebois RL, Doolittle WF (1996) An archaeobacterial homolog of pelota, a meiotic cell division protein in eukaryotes. *FEMS Microbiol Lett* 144: 151–155.
84. Kobayashi K, Kikuno I, Kuroha K, Saito K, Ito K, et al. (2010) Structural basis for mRNA surveillance by archaeal Pelota and GTP-bound EF1 α complex. *Proc Natl Acad Sci USA* 107: 17575–9.
85. de Koning B, Blombach F, Brouns SJ, van der Oost J (2009) Fidelity in archaeal information processing. *Archaea (Vancouver, BC)* 2010: doi:10.1155/2010/960298
86. White MF (2003) Archaeal DNA repair: paradigms and puzzles. *Biochem Soc Trans* 31: 690–693.
87. Snyder JC, Bateson MM, Lavin M, Young MJ (2010) Use of cellular CRISPR (clusters of regularly interspaced short palindromic repeats) spacer-based microarrays for detection of viruses in environmental samples. *Appl Environ Microbiol* 76: 7251–7258.
88. Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327: 167–170.
89. Marraffini L, Sontheimer E (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* 11: 181–90.
90. Lehtovirta-Morley L, Stoecker K, Vilcinskas A, Prosser J, Nicol G (2011) Cultivation of an obligate acidophilic ammonia oxidizer from a nitrifying acid soil. *Proc Natl Acad Sci USA* 108: 15892–7.
91. Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, et al. (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 437: 543–546.
92. Filée J, Siguier P, Chandler M (2007) Insertion sequence diversity in archaea. *Microbiology and Mol Biol Rev* 71: 121–157.
93. Hatzepichler R, Lebedeva EV, Spieck E, Stoecker K, Richter A, et al. (2008) A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. *Proc Natl Acad Sci U S A* 105: 2134–2139.
94. Caporaso GJ, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7: 335–336.
95. Francis CA, Roberts KJ, Beman JM, Santoro AE, Oakley BB (2005) Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proc Natl Acad Sci U S A* 102: 14683–14688.
96. Rothauwe JH, Witzel KP, Liesack W (1997) The ammonia monoxygenase structural gene *amoA* as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing populations. *Appl Environ Microbiol* 63: 4704–12.
97. Peng Y, Leung HC, Yiu S-M, Chin FY (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28: 1420–1428.
98. Chevreur B, Wetter T, Suhai Sn (1999) Genome sequence assembly using trace signals and additional sequence information. In: German conference on Bioinformatics 45: 56.
99. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32: W273–279.
100. Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14: 1394–1403.
101. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, et al. (2008) The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9: 75.
102. Markowitz VM, Mavromatis K, Ivanova NN, Chen I, Chu K, et al. (2009) IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 25: 2271–2278.
103. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, et al. (2013) GenBank. *Nucleic Acids Res* 41(D): 36–42.
104. Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, et al. (2013) TIGRFAMs and genome properties in 2013. *Nucleic Acids Res* 41(D): 387–395.
105. Tatusov RL, Fedorova ND, Jackson JD, Jackobs AR, Kiryutin B, et al. (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4: 41.
106. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Deryshire MK, et al. (2011) "CDD: a conserved domain database for the functional annotation of proteins". *Nucleic Acids Res* 39(D): 225–229.
107. Grissa I, Vergnaud G, Poucel C (2007) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35: W52–W57.
108. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792–1797.
109. Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56: 564–577.
110. Criscuolo A (2011) MorePhyML: improving the phylogenetic tree space exploration with PhyML 3. *Mol Phylogenet Evol* 61: 944–948.
111. Delcher AL, Salzberg SL, Phillippy AM (2003) Using MUMmer to identify similar regions in large sequence sets. *Curr Protoc Bioinformatics Chapter* 10: Unit 10.13.
112. Richter M, Rosselló-Móra R (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* 106: 19126–19131.
113. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26: 2460–2461.
114. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, et al. (2004) Versatile and open software for comparing large genomes. *Genome Biology* 5: R12.