environmental microbiology

Environmental Microbiology (2012) **14**(12), 3122–3145 doi:10.1111/j.1462-2920.2012.02893.x

The genome of the ammonia-oxidizing *Candidatus* Nitrososphaera gargensis: insights into metabolic versatility and environmental adaptations

Anja Spang,[1†] Anja Poehlein,[2†] Pierre Offre,[1] Sabine Zumbrägel,[3] Susanne Haider,[4] Nicolas Rychlik,[3] Boris Nowka,[3] Christel Schmeisser,[3] Elena V. Lebedeva,[5] Thomas Rattei,[6] Christoph Böhm,[4] Markus Schmid,[4] Alexander Galushko,[4] Roland Hatzenpichler,[4‡] Thomas Weinmaier,[6] Rolf Daniel,[2] Christa Schleper,[1] Eva Spieck,[3] Wolfgang Streit[3] and Michael Wagner[4*]**

[1]*Department of Genetics in Ecology, University of Vienna, Althanstr. 14, 1090 Vienna, Austria.*
[2]*Laboratorium für Genomanalyse, Universität Göttingen, 37077 Göttingen, Germany.*
[3]*Biozentrum Klein Flottbek, Abteilung für Mikrobiologie und Biotechnologie, Universität Hamburg, Ohnhorststrasse 18, 22609 Hamburg, Germany.*
[4]*Department of Microbial Ecology, University of Vienna, Althanstr. 14, 1090 Vienna, Austria.*
[5]*Winogradsky Institute of Microbiology, Russian Academy of Sciences, Prospect 60-let Oktyabrya 7/2, Moscow 117312, Russia.*
[6]*Department of Computational Systems Biology, University of Vienna, Althanstr., 14, 1090 Vienna, Austria.*

Summary

The cohort of the ammonia-oxidizing archaea (AOA) of the phylum *Thaumarchaeota* is a diverse, widespread and functionally important group of microorganisms in many ecosystems. However, our understanding of their biology is still very rudimentary in part because all available genome sequences of this phylum are from members of the *Nitrosopumilus* cluster. Here we report on the complete genome sequence of *Candidatus* Nitrososphaera gargensis obtained from an enrichment culture, representing a different evolutionary lineage of AOA frequently found in high numbers in many terrestrial environments. With its 2.83 Mb the genome is much larger than that of other AOA. The presence of a high number of (active) IS elements/transposases, genomic islands, gene duplications and a complete CRISPR/Cas defence system testifies to its dynamic evolution consistent with low degree of synteny with other thaumarchaeal genomes. As expected, the repertoire of conserved enzymes proposed to be required for archaeal ammonia oxidation is encoded by *N. gargensis*, but it can also use urea and possibly cyanate as alternative ammonia sources. Furthermore, its carbon metabolism is more flexible at the central pyruvate switch point, encompasses the ability to take up small organic compounds and might even include an oxidative pentose phosphate pathway. Furthermore, we show that thaumarchaeota produce cofactor F420 as well as polyhydroxyalkanoates. Lateral gene transfer from bacteria and euryarchaeota has contributed to the metabolic versatility of *N. gargensis*. This organisms is well adapted to its niche in a heavy metal-containing thermal spring by encoding a multitude of heavy metal resistance genes, chaperones and mannosylglycerate as compatible solute and has the genetic ability to respond to environmental changes by signal transduction via a large number of two-component systems, by chemotaxis and flagella-mediated motility and possibly even by gas vacuole formation. These findings extend our understanding of thaumarchaeal evolution and physiology and offer many testable hypotheses for future experimental research on these nitrifiers.

Received 31 August, 2012; accepted 1 September, 2012. For correspondence. *E-mail wagner@microbial-ecology.net; Tel. (+43) 1 4277 54390; Fax (+43) 1 4277 54389. **E-mail wolfgang.streit@uni-hamburg.de; Tel. (+49) 40 42816 463; Fax (+49) 40 42816 459. †Contributed equally. ‡Current address: Division of Geological and Planetary Sciences, California Institute of Technology.

Introduction

The discovery of ammonia-oxidizing archaea (AOA) after more than 100 years of nitrification research (Könneke *et al.*, 2005; Treusch *et al.*, 2005) has dramatically changed our perception of the diversity (Francis *et al.*, 2005), abundance (Leininger *et al.*, 2006) and niche adaptations (Martens-Habbena *et al.*, 2009; Gubry-Rangin *et al.*, 2011; Verhamme *et al.*, 2011) of nitrifying microbes in the environment. 16S rRNA- and *amoA*-

based surveys of AOA revealed an astonishing diversity of members of this guild with hundreds of predicted species (Schleper *et al*., 2005; Gubry-Rangin *et al*., 2011; Pester *et al*., 2012) belonging to several major clades (Fig. S1A) within the newly described phylum *Thaumarchaeota* (Brochier-Armanet *et al*., 2008; Spang *et al*., 2010). Among them, the *Nitrosopumilus* (also called group I.1a or marine group I) and the *Nitrososphaera* clusters including its recently described sister cluster (Pester *et al*., 2012) (collectively referred to as group I.1b or soil group) seem to be most abundant in aquatic and terrestrial environments respectively. Till today only a few enrichment cultures (Hatzenpichler *et al*., 2008; de la Torre *et al*., 2008; Jung *et al*., 2011; Lehtovirta-Morley *et al*., 2011; French *et al*., 2012; Kim *et al*., 2012; Mosier *et al*., 2012a,b) and two pure cultures (Könneke *et al*., 2005; Tourna *et al*., 2011) of AOA are available for physiological and molecular experiments. So far, only closely related members from the *Nitrosopumilus* cluster (Hallam *et al*., 2006a; Walker *et al*., 2010; Blainey *et al*., 2011; Kim *et al*., 2011; Urakawa *et al*., 2011; Mosier *et al*., 2012a,c) and a few short fosmid sequences from members of the *Nitrososphaera* cluster (Treusch *et al*., 2005; Bartossek *et al*., 2010; 2012) have been investigated genomically (Fig. S1A; Table 1), revealing that AOA exploit a surprisingly different enzymatic repertoire for energy and carbon metabolism compared with their ammonia-oxidizing bacterial counterparts. For example, members of the *Nitrosopumilus* cluster lack homologues for hydroxylamine oxidoreductase or cytochromes c_{552} and c_{554} typically found in ammonia-oxidizing bacteria (AOB), suggesting an alternative substrate conversion pathway and different components for electron transport, likely including copper binding proteins (Walker *et al*., 2010). Furthermore, thaumarchaeota of the *Nitrosopumilus* cluster apparently use a modified version of the 3-hydroxypropionate/4-hydroxybutyrate pathway for autotrophic carbon fixation (Berg *et al*., 2007; 2010a) instead of the Calvin–Benson–Bassham cycle common to all known AOB. However, a comprehensive understanding of the physiology and evolutionary history of *Thaumarchaeota* will require comparative genomic analyses with members of the other evolutionary lines of descent within this phylum.

Here we report on the first complete genome of an AOA from the *Nitrososphaera* cluster. The 2.83 Mb genome sequence of *Candidatus* Nitrososphaera gargensis (Fig. S1B and C; please note that the *Candidatus* designation was omitted in the following text to improve readability) was reconstructed from an enrichment culture obtained from a microbial mat growing in an outflow pond (45°C) of a terrestrial hot spring (Lebedeva *et al*., 2005; Hatzenpichler *et al*., 2008). Thus, *N. gargensis* occupies a fundamentally different niche to those AOA previously investigated on the genomic level as also reflected by

Table 1. General features of the *N. gargensis* genome in comparison with those of other published thaumarchaeal genomes.

Organism	*Nitrososphaera gargensis*	*Nitrosopumilus maritimus* SCM1	*Cenarchaeum symbiosum* A	*Nitrosoarchaeum limnia* SFB1	*Nitrosoarchaeum koreensis* MY1	*Nitrosoarchaeum limnia* BD20	*Nitrosopumilus salaria* BD31
Affiliation	*Nitrososphaera* cluster, group I.1b	*Nitrosopumilus* cluster, group I.1a	*Nitrosopumilus* cluster, group I.1a	*Nitrosopumilus* cluster, group I.1a	*Nitrosopumilus* cluster, group I.1a	*Nitrosopumilus* cluster, group I.1a	*Nitrosopumilus* cluster, group I.1a
Habitat	Microbial mat (outflow pond of Garga spring)	Tropical fish tank, 21–23°C	Sponge symbiont	Low-salinity sediment	Soil rhizosphere of *Caragana sinica*	Low-salinity sediment	Sediment with variable salinity
Temperature range	Moderate thermophilic (46°C optimum)	Mesophilic	Mesophilic	Mesophilic	Mesophilic	Mesophilic	Mesophilic
Origin of genome sequences	Enrichment culture	Pure culture	Natural enrichment (sponge host)	Enrichment culture	Enrichment culture	Enrichment culture	Enrichment culture
Genome size in Mb, number of contigs in brackets	2.83[a] (1)	1.65 (1)	2.05 (1)	1.77 (76)	1.61 (1)	1.86 (343)	1.57 (171)
G+C content	48%	34%	58%	32%	33%	33%	34%
ORPHANS[b]	32%	15%	30%	22%	23%	n.d.[c]	n.d.[c]
16S–23S rRNA genes	1	1	1	1	1	1	1
Separate 5S rRNA	1	1	1	1	1	1	1
Number of tRNAs	40[d]	44	45	45	42	39	40

a. The *N. gargensis* genome contains a large tandem duplication of a genomic fragment that is 16 966 bp in length.
b. Predicted proteins without homology (*E*-value $< 10^{-4}$) to other proteins excluding those predicted for organisms from same phylum.
c. The genome sequence was not available when the ORPHAN analysis was performed.
d. It should be noted that *N. gargensis* possesses some tRNAs with non-canonical introns (for details, see *Supporting information*), which might explain the relatively low number of detected tRNAs.

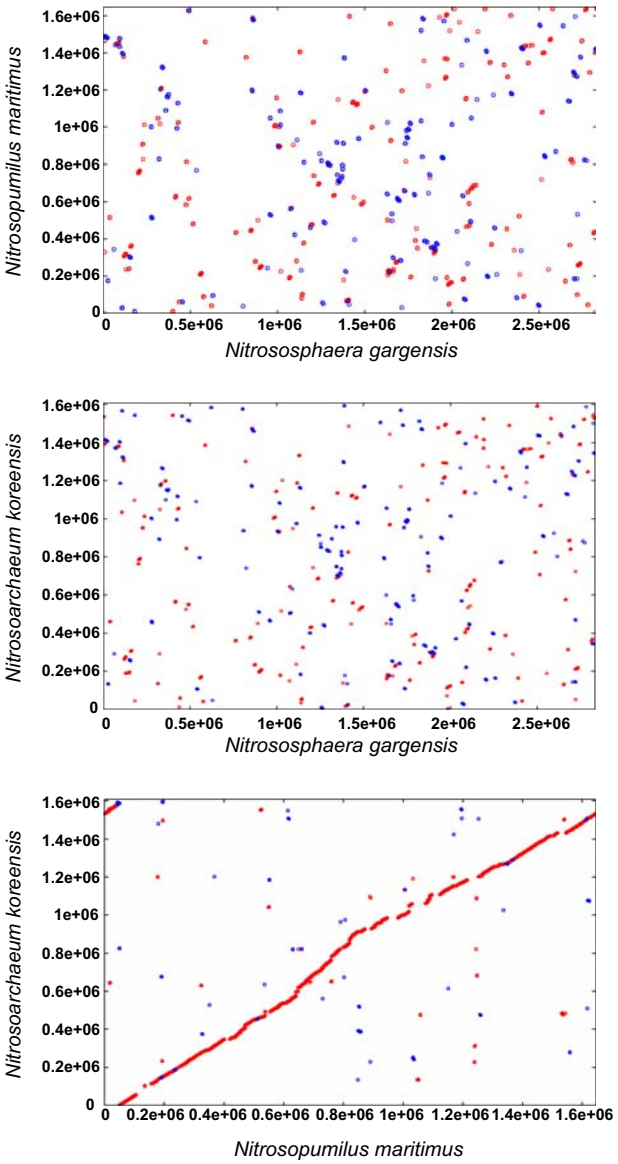

Fig. 1. Dot plot representation of the pairwise alignments of the *N. gargensis*, *N. maritimus* and *N. koreensis* genomes. Alignments were performed on the six-frame amino acid translation of the genome sequences using the Promer program in the MUMmer 3.23 package (Kurtz *et al.*, 2004). In all plots, a dot indicates a match (of at least six AA) between the two genome sequences being compared, with forward matches coloured in red and reverse matches coloured in blue.

its optimal growth temperature of 46°C. Interestingly however, very close relatives of *N. gargensis* have also been detected in various soils where they often thrive at a considerable abundance (Pester *et al.*, 2011; 2012). Consistently, the soil isolate *Nitrososphaera viennensis* (Tourna *et al.*, 2011) has 97.5% 16S rRNA and 95.6% AmoA (corresponding to 86.3% identity on the nucleotide level) sequence identity to *N. gargensis*.

Manual annotation of the *N. gargensis* genome showed on the one hand a gene set for ammonia oxidation and CO_2 fixation similar to previously studied thaumarchaeota, but on the other hand revealed a very dynamically evolving genome with frequent gene duplication and lateral gene transfer events contributing to an unexpected physiological versatility.

Results and discussion

General genomic features and examples of gene family expansions

A closed genome sequence was obtained from a *N. gargensis* enrichment culture, which still contained several contaminating bacteria (Fig. S2). The *N. gargensis* genome shows large differences to the previously published genomes of this phylum which all originated from members of the *Nitrosopumilus* cluster (Hallam *et al.*, 2006a,b; Walker *et al.*, 2010; Blainey *et al.*, 2011; Kim *et al.*, 2011; Mosier *et al.*, 2012a,c). For example, with its 2 833 868 bp and 3566 predicted open reading frames (ORFs) it represents by far the largest thaumarchaeal genome and also differs from those of the free-living members of the *Nitrosopumilus* cluster by its higher overall G+C content (48.4%) and a high number of orphan genes partially explaining the larger genome size (Table 1). Furthermore, more than 2300 predicted proteins of *N. gargensis* do not fall in clusters of orthologous groups (COGs) when compared with the *in silico* proteome of the other thaumarchaeota (Fig. S3A). Likewise, while members of the *Nitrosopumilus* cluster with the exception of *Cenarchaeum symbiosum* exhibit pronounced genome synteny with each other, no larger syntenic regions were observed between the *N. gargensis* genome and those of other thaumarchaeota (Fig. 1). Altogether, these findings are consistent with the phylogenetic placement of *N. gargensis* into a separate thaumarchaeal lineage and suggest that the gene complement of representatives of the *Nitrososphaera* cluster is significantly different from that of *Nitrosopumilus* cluster members. We also explored whether a comparative COG analysis (Fig. S3B) would provide indications for a high number of shared orthologues between *N. gargensis* as moderate thermophile and the recently discovered thermophilic *Caldiarchaeum subterraneum* that was proposed to represent a new archaeal phylum but also shares some features with members of the *Thaumarchaeota* (Nunoura *et al.*, 2011). Interestingly however, *C. subterraneum* shared more orthologues with a member of the *Crenarchaeota* than with either *N. gargensis* or *Nitrosopumilus maritimus* alone or with both of these thaumarchaeota together. Thus, this finding does not provide further support for an affiliation of *C. subterraneum* with the *Thaumarchaeota*, but much more detailed analyses ideally including deep branching members of the *Thau-*

marchaeota will be necessary to clarify this debate (Brochier-Armanet *et al*., 2012).

In the following sections, we present and discuss selected genomic features of *N. gargensis*, including aspects of its informational processing machinery, integrative elements and virus defence system, energy and carbon metabolism, transporters as well as some additional unexpected features and adaptations. The detailed annotation of all discussed genes can be found in Tables S1–S5.

As previously reported, *N. gargensis* encodes a similar set of information processing genes like the other thaumarchaeota (Spang *et al*., 2010; Table S3) supporting the notion of the evolutionary distinctiveness of this phylum. However, in contrast to members of the *Nitrosopumilus* cluster some informational processing genes of *N. gargensis* are present in higher copy number. For example, whereas thaumarchaeota of the *Nitrosopumilus* cluster possess only one or rarely two homologue(s) of the cell division cycle protein cdc6/ORC that is suggested to be involved in selection and binding of replication origin and cell cycle control (e.g. Zhang *et al*., 2009; Akita *et al*., 2010), genes for three homologues of this protein were identified in the *N. gargensis* genome. The presence of several *cdc6* gene copies might indicate additional replication origins as these genes are often, although not always, located adjacent to replication origins (Kelman and Kelman, 2004; Robinson and Bell, 2007). In fact, in several cren- and euryarchaeota, the operation of multiple replication origins has been verified experimentally (Robinson *et al*., 2004; Norais *et al*., 2007). Furthermore, homologues of the archaeal transcription initiation factor TFB are encoded 11 times in the *N. gargensis* genome, all of which group into a monophyletic cluster with the respective proteins from the members of the *Nitrosopumilus* cluster (which possess 5–8 copies; data not shown). TFB gene family expansion has been reported in several archaeal lineages and has recently been experimentally connected to niche adaptation capability in haloarchaea (Turkarslan *et al*., 2011). It is tempting to propose a similar mechanism in thaumarchaeota with *N. gargensis* as the currently most complex representative.

Molecular chaperones (Large *et al*., 2009) represent another group of proteins that have experienced gene duplication events in *N. gargensis*. For instance, the genome encodes six paralogues coding for archaeal thermosomes (group II chaperonin proteins), a more distantly related xenologue coding for an additional putative archaeal thermosome as well as at least eight Hsp20-like chaperones. This is in contrast to members of the *Nitrosopumilus* cluster that contain less archaeal thermosomes (2–3) and Hsp20-like chaperones (2–5). Although several archaea have experienced gene duplications of molecular group II chaperones (Macario *et al*., 2004), the high copy number of these proteins in *N. gargensis* is remarkable and comparable only with the large genomes of few *Methanosarcina* species that in addition contain active group I chaperonins generally found in bacteria (Klunker *et al*., 2003; Macario *et al*., 2004). These findings indicate that *N. gargensis* has a high requirement for assistance in the folding of newly synthesized or denatured proteins aiding in versatile stress response (Macario *et al*., 2004).

Integrative elements, CRISPR defence system

More than 75 transposase genes including pseudogenes with transposase signatures are present in the *N. gargensis* genome and can be assigned to at least three different types of transposases (Table S4) that are associated with insertion sequence families IS605, IS1 and IS4 (Filée *et al*., 2007). Interestingly, the IS element (1332 bp) of *N. gargensis* belonging to group IS4 (subgroup ISH3) herein tentatively referred to as ISNga1 occurs 14 times in the genome. All copies have 100% nucleotide identity to each other, and contain a single ORF (ISH3 transposase) and a terminal inverted repeat of 14 bp as well as flanking repeats (5 bp) indicating recent transposon activity. Although most of the ISNga1 elements occur in poorly conserved genomic regions, several elements appear to have inserted into protein-encoding genes indicating that they significantly contribute to genome evolution of this organism. Noteworthy, ISNga1-12 is located in a 33.5 kb region that might represent a provirus (or derivative thereof) as it is flanked by an integrase, which is encoded next to a lys-tRNA gene (Table S5). It appears possible that ISNga1 has been introduced into the genome of *N. gargensis* along with the integration of this putative genetic element. Furthermore, six other putative integrases were found, some of which are located in genome regions mainly encoding small hypothetical proteins. These regions range from about 9000 bp to 33 500 bp and could represent putative proviruses, integrated plasmids or derivates thereof, as well as other islands of high genomic plasticity. Three of the putative integrase-encoding genes are located in proximity of tRNA genes.

Interestingly, a CRISPR/Cas (clustered regularly interspaced short palindromic repeats/CRISPR-associated genes) viral defence system (Sorek *et al*., 2008; Horvath and Barrangou, 2010; Al-Attar *et al*., 2011) was identified in the *N. gargensis* genome (Table S1). CRISPRfinder (Grissa *et al*., 2007) also detected putative CRISPR loci in the genomes of *N. maritimus* and *Nitrosopumilus salaria* (see also Mosier *et al*., 2012a), but we could not identify associated *cas* genes, which are required for viral defence, in the latter genomes. Furthermore, in contrast to *N. gargensis* the predicted CRISPR sequences

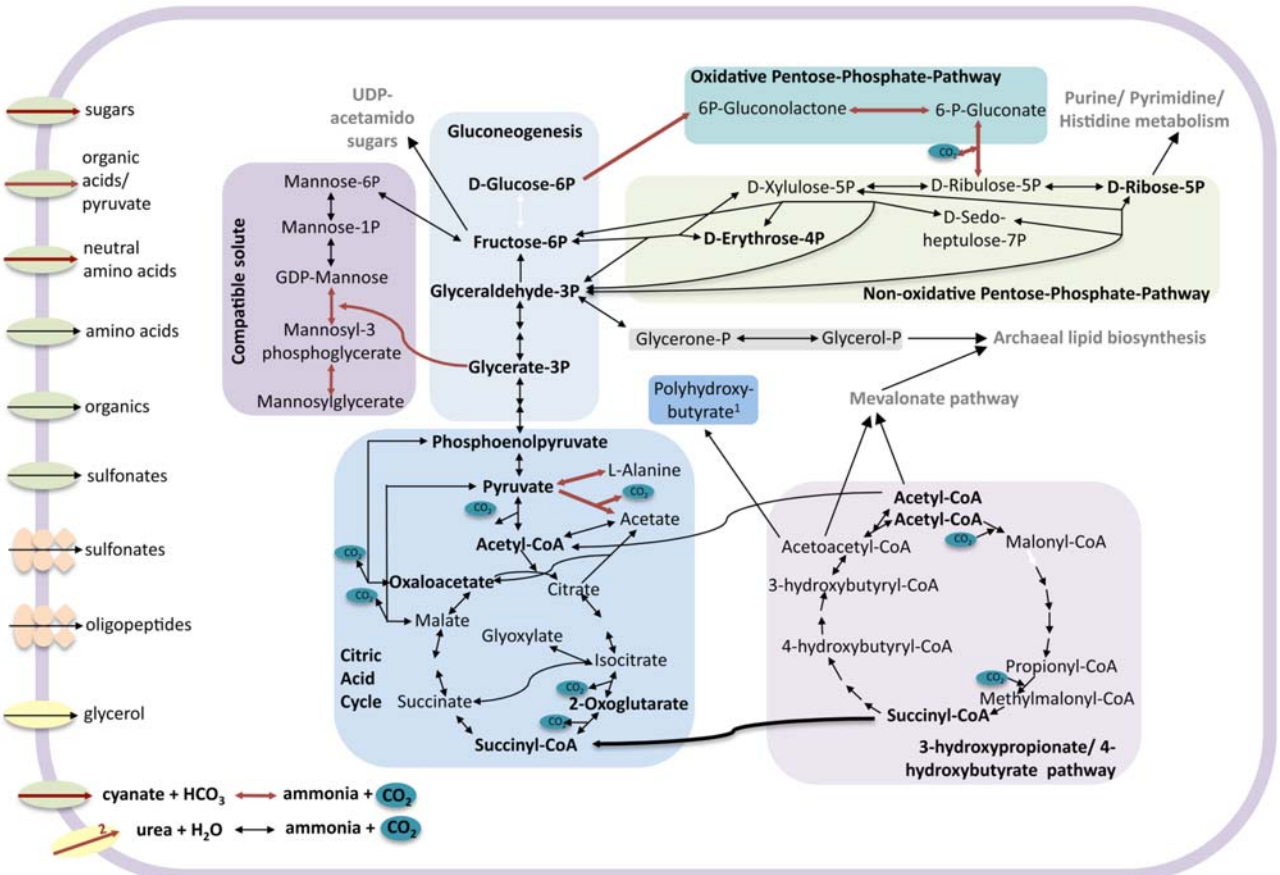

Fig. 2. Predicted carbon metabolism of *N. gargensis* (see also Fig. S4 for further details on respective EC numbers and annotation class as well as Table S1 for gene annotations). Enzymatic reactions for which candidate genes could be identified in the genome of *N. gargensis* and which contain homologues in genomes of thaumarchaeota of the *Nitrosopumilus* cluster are highlighted by solid black arrows. Reactions for which *N. gargensis* encodes candidate enzymes that are absent from thaumarchaeota of the *Nitrosopumilus* cluster (but might be present on fosmids isolated from soil) are shown with red arrows. For reactions presented with white arrows, specific candidate genes have not been identified. [1]Related PHA synthases (type III) of haloarchaeota synthesize polyhydroxybutyrate and/or 3-polyhydroxybutyrate-co-3-hydroxyvalerate. [2]*N. gargensis* encodes two unrelated urea transporters one of which is specific to thaumarchaeota of the *Nitrososphaera* cluster.

comprising the spacer–repeat units of *N. maritimus* and *N. salaria* might rather belong to ORFs encoded in those regions which are conserved between closely related organisms. The CRISPR locus of *N. gargensis* comprises 2763 bp and consists of 36 repeat/spacer sequences. According to a recent reclassification of the diverse CRISPR-Cas systems into three major types (I–III) (Makarova *et al*., 2011), *N. gargensis* harbours a type I system that is proposed to function in virus defence by directly targeting DNA. Interestingly, some spacers show remote similarities to putative provirus regions in the *N. gargensis* genome as well as to the recently published putative provirus of *N. viennensis* (Krupovic *et al*., 2011).

Energy metabolism

Nitrososphaera gargensis can grow chemolithoautotrophically by aerobic oxidation of ammonia to nitrite and concomitant fixation of inorganic carbon as demonstrated by catalysed reporter deposition fluorescence *in situ* hybridization (CARD–FISH)–microautoradiography experiments of an enrichment culture (Hatzenpichler *et al*., 2008). Genomic analyses of *N. gargensis* confirmed its chemolithoautotrophy (Table S1) but also highlighted unexpected variations of this metabolic theme. For example, *N. gargensis* is not dependent on freely available ammonia but can also produce ammonia (and CO_2) from urea and possibly even from cyanate (Fig. 2) as it encodes a urease and a cyanate hydratase, which contains all residues forming the active site of this enzyme in *Escherichia coli* (Walsh *et al*., 2000). Conversion of urea to nitrite by the *N. gargensis* enrichment culture has been verified experimentally (data not shown) and was also shown for its close relative *N. viennensis* (Tourna *et al*., 2011). Interestingly, however, a cyanate hydratase is not present in other thaumarchaeota and it is located in

the genome of *N. gargensis* close to a formate/nitrite transporter family member, which inferred from this colocalization might represent a cyanate transporter (Czyzewski and Wang, 2012) (alternatively cyanate could also be transported by *N. gargensis* by an ABC-type cyanate/nitrite/nitrate transporter). Surprisingly, the *N. gargensis* cyanate hydratase is most closely related to a homologue in the nitrite oxidizer *Nitrospira defluvii* (Lücker *et al*., 2010) which also encodes a closely related transporter in the same operon (Fig. 3). The colocalization of the cyanate uptake and conversion module suggests that the respective genes can easily be transferred together (Fig. 3) and that *N. gargensis* has exchanged these genes with nitrifiers sharing a similar niche.

The exact mechanism of the oxidation of ammonia to nitrite in thaumarchaeota is still unresolved both concerning the enzymes involved as well as the intermediates formed (Klotz and Stein, 2008; Schleper and Nicol, 2010). In AOB, ammonia oxidation follows a two-step mechanism via the intermediate hydroxylamine catalysed by the activity of ammonia monooxygenase (AMO) and hydroxylamine oxidoreductase (HAO) (Prosser, 1989; Arp *et al*., 2007). Like in all published thaumarchaeal genomes (Hallam *et al*., 2006a; Walker *et al*., 2010; Blainey *et al*., 2011; Kim *et al*., 2011; Mosier *et al*., 2012a,c), the *N. gargensis* genome encodes an AMO, which is distantly related to the bacterial version, but a HAO homologue could not be identified. In all AOB, the AMO is encoded by *amoA*, *B* and *C* that occur in one operon of conserved gene order (Arp *et al*., 2007) and these organisms might additionally contain unlinked copies of *amoC* (Norton *et al*., 2002). Homologues of *amoA*, *B* and *C* genes of the *Nitrosopumilus* and *Nitrosocaldus* clusters are arranged in a different gene order (Fig. 4) and contain an additional ORF in proximity to *amoA* that was referred to as ORF38 or *amoX* (Schleper *et al*., 2005; Treusch *et al*., 2005; Bartossek *et al*., 2012). Interestingly, the *amoA*, *B* and *C* genes of *N. gargensis* and *N. viennensis* are not in close proximity or arranged in a gene cluster except for the association of *amoA* and *amoX* (Fig. 4). An additional *amoC* gene copy (97% amino acid identity to the other AmoC) was identified distantly located in the genome (Fig. 4).

Furthermore, it has been postulated that multicopper proteins of thaumarchaeota might be involved in ammonia oxidation (Bartossek *et al*., 2010; Walker *et al*., 2010). *Nitrososphaera gargensis* encodes six multicopper proteins, two of which were classified as nitrite reductases (NirK) generally involved in the reduction of NO_2^- to NO. NirK has not only been found in denitrifying organisms, it also occurs in AOB (Arp *et al*., 2007) and other AOA. Metatranscriptome analyses of thaumarchaeota (Hollibaugh *et al*., 2011; Radax *et al*., 2012; Stewart *et al*., 2012) showed that *nirK* homologues range among the most highly transcribed genes and a potential function of NirK in archaeal ammonia oxidation has recently been proposed in Schleper and Nicol (2010). Besides the *nirK* homologues, no other genes typically involved in denitrification could be identified.

Electron transfer chain

Nitrososphaera gargensis encodes a complete respiratory chain consisting of complexes 1–5, which is used for energy generation and reverse electron transport (Table S1). In consistency with thaumarchaeota of the *Nitrosopumilus* cluster, *N. gargensis* lacks homologues of cytochrome *c* proteins and might instead transfer electrons derived from ammonia oxidation via small blue (type 1) copper proteins to a quinone reductase (Walker *et al*., 2010). Like in *N. maritimus*, in *N. gargensis* one of these proteins contains two cupredoxin and five transmembrane domains and might represent a candidate for such a quinone reductase (Walker *et al*., 2010). Thaumarchaeota (Urakawa *et al*., 2011) including *N. gargensis* have only 11 genes encoding different subunits of type 1 NADH dehydrogenase (two of which are duplicated in *N. gargensis*) but lack close homologues of the peripheral subunits E, F and G that are generally required as canonical electron input module. However, thaumarchaeota contain several FAD/NAD(P) binding oxidoreductases that have similarity to the FAD binding domain of Na^+-translocating NADH–quinone reductase subunit F (NqrF) which serves as electron entry site for sodium-translocating NADH:quinone reductases (type III NADH dehydrogenase) (Türk *et al*., 2004; Verkhovsky and Bogachev, 2010). Although thaumarchaeal FAD/NAD(P) binding oxidoreductases lack the typical 2Fe-2S ferredoxin-type domain of NqrF, these could represent interesting candidate genes for alternative electron input modules acting in concert with separately encoded small ferredoxins present in all genome-sequenced thaumarchaeota. On the other hand, it should be noted that genomes of most archaea and several bacteria do not encode subunits E, F and G of complex I and some of them use alternative non-homologous NADH dehydrogenases (Schäfer *et al*., 1996; 1999; Gomes *et al*., 2001; Melo *et al*., 2004; Pereira *et al*., 2004; Patridge and Ferry, 2006; Moparthi and Hägerhäll, 2011). For example, *N. gargensis* and the other thaumarchaeota might use a membrane-associated single subunit type II NADH dehydrogenase [23% AA identity of *N. gargensis* homologue to *Acidianus ambivalens noxA* (Gomes *et al*., 2001; Bandeiras *et al*., 2002)].

Central carbon metabolism

In accordance with all other published thaumarchaeal genomes, *N. gargensis* encodes all key genes of the

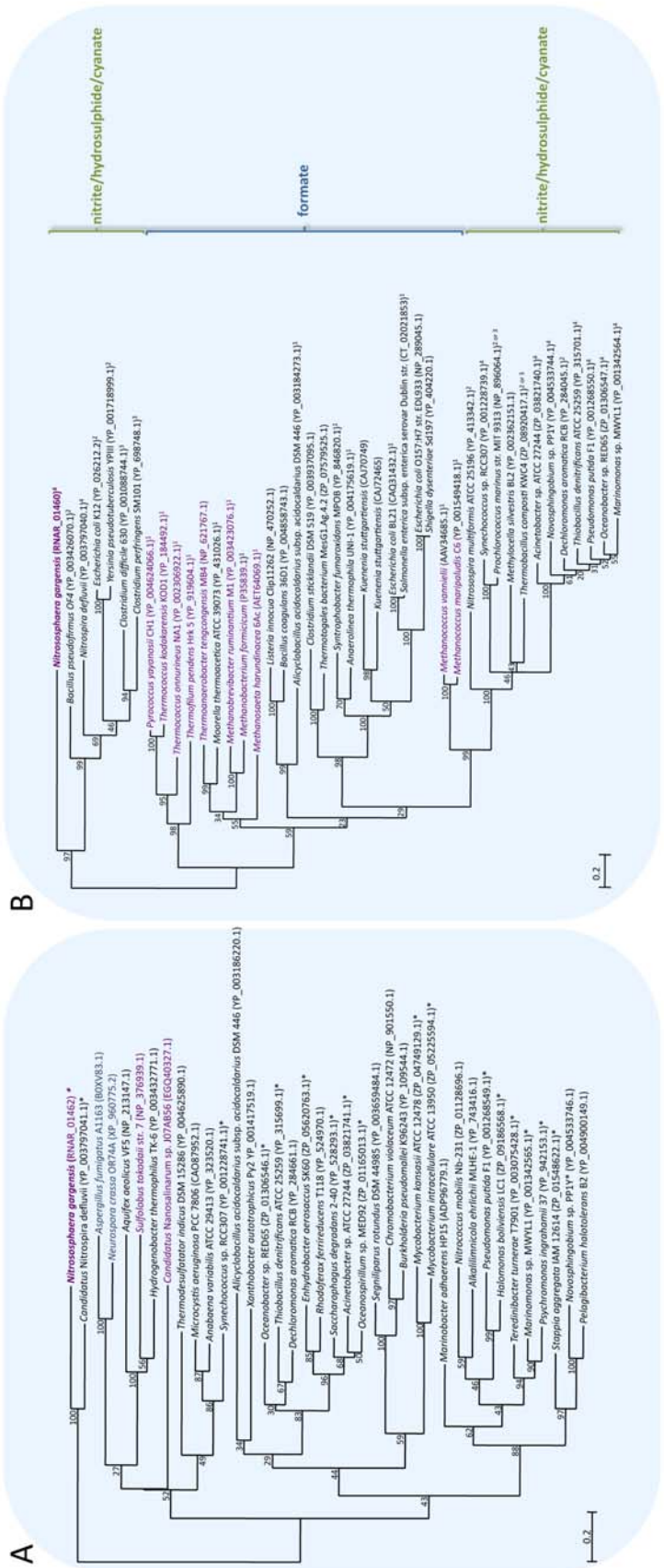

Fig. 3. A. Maximum likelihood phylogenetic analysis of cyanate hydratase based on an alignment of 147 amino acid positions (50% insertion/deletion filter and 100 bootstraps). The scale bar refers to 20% estimated sequence divergence. *Gene for cyanate hydratase in operon with or close to gene encoding putative formate/nitrite family transporter. Archaea are labelled in purple, Eukarotes are labelled in blue and Bacteria are shown in black.
B. Maximum likelihood phylogenetic analysis of formate/nitrite transporter (fnt) family proteins of archaea and bacteria based on a manually corrected alignment of 229 amino acid positions (removal of ambiguously aligned sites, 100 bootstraps). The scale bar refers to 20% estimated sequence divergence. Numbers in superscript indicate putative substrates transported by the respective transporters based on the substrate specificity of the enzymes encoded by flanking genes (Czyzewski and Wang, 2012): [1]formate, [2]nitrite, [3]hydrosulfide sulfite, [4]cyanate. Archaeal sequences are highlighted in purple.

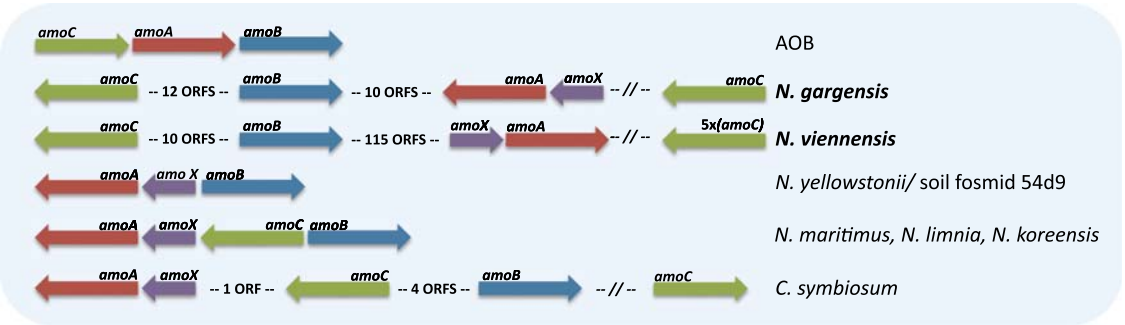

Fig. 4. Ammonia monooxygenase (*amo*) gene order of *N. gargensis* compared with ammonia-oxidizing bacteria (AOB), the close relative *Candidatus* Nitrososphaera viennensis and thaumarchaeota of the *Nitrosopumilus* and *Nitrosocaldus* clusters respectively. (C) The different arrangement of *amoA* and *amoX* genes in the genome of *N. viennensis* compared with *N. gargensis* appears to originate from an inversion of these and more than 100 flanking genes. Gene boxes are not drawn to scale.

3-hydroxypropionate/4-hydroxybutyrate carbon fixation pathway (Berg *et al*., 2007; 2010a,b; Berg, 2011) and probably operates a modified version of it (Hallam *et al*., 2006b; Berg *et al*., 2007; 2010a,b) (Figs 2 and S4, Table S1, and *Supporting information*). Notably, it has recently been shown for *Metallosphaera sedula* that succinyl-CoA (in addition to a minor amount of acetyl-CoA) is the main carbon fixation product of this pathway that can directly enter the tricarboxylic acid (TCA) cycle (Estelmann *et al*., 2011).

Nitrososphaera gargensis harbours candidate enzymes for all steps of an oxidative TCA (Figs 2 and S4, Table S1), but it remains unclear whether it can exploit classical TCA bypass reactions as only one enzyme – isocitrate lyase – catalysing the first step of the glyoxylate shunt could be identified in the genome. Furthermore, the substrate specificity of the 2-oxoacid:ferredoxin oxidoreductases (OFOR) – present in a single copy in all sequenced thaumarchaeota and encoded by a fused alpha and gamma and a separate beta subunit gene – remains to be elucidated. In composition and sequence, this two-subunit enzyme is most similar to pyruvate:ferredoxin oxidoreductase of several other aerobic archaea (Kerscher and Oesterhelt, 1981; Zhang *et al*., 1996). Characterized OFORs of *Sulfolobus* sp. have broad substrate specificity towards diverse oxoacids (Zhang *et al*., 1996; Fukuda and Wakagi, 2002; Park *et al*., 2006). Thus, it appears possible that the thaumarchaeal OFOR substitutes the function of the 2-oxoglutarate dehydrogenase that catalyses the decarboxylation of 2-oxoglutarate to succinyl-CoA in many bacteria, which is lacking in thaumarchaeota including *N. gargensis*. But even an incomplete TCA cycle could be sufficient to support autotrophic growth of *N. gargensis* and yield all intermediates required for biosynthesis as shown for AOB (Wood *et al*., 2004; Hommes *et al*., 2006; Arp *et al*., 2007). Furthermore, succinyl-CoA as putative primary carbon fixation product of thaumarchaeota is ideally suited for an organism without a 2-oxoglutarate:ferredoxin oxidoreductase as it represents this enzyme's reaction product (in the oxidative direction of the TCA cycle).

Nitrososphaera gargensis contains all genes that code for gluconeogenic enzymes involved in the synthesis of fructose-6-phosphate (Figs 2 and S4, Table S1). During autotrophic growth, gluconeogenesis might start from succinyl-CoA or acetyl-CoA. Succinyl-CoA can be oxidized to malate and/or oxaloacetate. Phosphoenolpyruvate (PEP) can directly be synthesized from oxaloacetate by the ATP-dependent PEP carboxykinase of *N. gargensis*, an enzyme present in other thaumarchaeota as well and phylogenetically clustering with bacterial homologues (data not shown), but absent from all other archaea that contain analogous PEP carboxykinases. Theoretically, *N. gargensis* has an alternative route of PEP synthesis using NAD(P)-dependent malic enzyme to produce pyruvate and pyruvate phosphate dikinase for converting it further to PEP (Siebers and Schönheit, 2005), but this route is energetically less favourable. Whereas the latter enzyme is present in all thaumarchaeota, the malic enzyme only occurs in *N. salaria* and *Nitrosoarchaeum limnia* BD20. Gluconeogenesis could also start from acetyl-CoA in *N. gargensis* and the other thaumarchaeota, if they are capable of synthesizing pyruvate from acetyl-CoA with their 2:oxoacid-ferredoxin oxidoreductase operating in the reductive direction. However, this appears unlikely as this reaction does not usually occur during aerobic growth (Berg *et al*., 2010a; Fuchs, 2011). From PEP, all intermediate steps yielding glyceraldehyde-3-phosphate according to the classical reversal of glycolysis are present in *N. gargensis*. The subsequent reaction from glyceraldehyde-3-phosphate via fructose-1.6-bisphosphate to fructose-6-phosphate might be catalysed by a single bifunctional fructose-1.6-bisphosphate (FBP) aldolase/phosphatase as suggested recently (Say and Fuchs, 2010). In contrast to thaumarchaeota of the *Nitrosopumilus* cluster, *N. gargensis* lacks a homologue

of known phosphoglucose isomerases that catalyse the final step of gluconeogenesis from fructose-6-phosphate to glucose-6-phosphate. Considering the high diversity of phosphoglucose isomerases in archaea (e.g. Hansen *et al*., 2001; 2005; Verhees *et al*., 2001; Jeong *et al*., 2003), however, it is likely that another sugar isomerase with a cupin domain substitutes the missing function in *N. gargensis*.

In *N. gargensis* genes coding for enzymes involved in the non-oxidative pentose phosphate pathway are present (Figs 2 and S4, Table S1), which can be used for the generation of pentose sugars as reported for a few archaea (Soderberg, 2005; van de Werken *et al*., 2008; and *Supporting information*). In contrast to most other archaea, homologues of key genes for the alternative ribulose monophosphate pathway (Soderberg and Alver, 2004; Soderberg, 2005) could not be detected in *N. gargensis* and are absent from other thaumarchaeota as well.

While hexose degradation via the Entner Doudoroff pathway or glycolysis could not be confirmed for *N. gargensis* (*Supporting information*), this organism might unexpectedly possess a functional oxidative pentose phosphate pathway (Figs 2 and S4, Table S1) that has not been demonstrated yet to operate in other archaea (van de Werken *et al*., 2008) and appears to be absent from other thaumarchaeota as well. Specifically, *N. gargensis* encodes a putative 6-phosphogluconolactonase (*Supporting information*) and two paralogous genes (one of which is fragmented) coding for putative 6-phosphogluconate dehydrogenase (Fig. S5), a key enzyme of the oxidative pentose phosphate pathway that bridges its oxidative and non-oxidative part. Furthermore, although the second key enzyme of this pathway, NADH-dependent glucose-6-phosphate dehydrogenase, was not found, one of the genes coding for 6-phosphogluconate dehydrogenase is located in proximity to two other genes putatively involved in the pentose phosphate pathway, a ribose-5-phosphate isomerase and a putative F420-dependent oxidoreductase/dehydrogenase. The latter enzyme is closely related to a functionally characterized F420-dependent glucose-6-phosphate dehydrogenase of *Mycobacterium smegmatis* (Purwantini and Daniels, 1996) (*Supporting information*). Whereas this function has been verified in several additional mycobacteria and *Nocardia* sp., F420-dependent glucose-6-phosphate dehydrogenase activity was not detected in other organisms possessing a homologous enzyme (Purwantini *et al*., 1997). The cofactor F420 plays an important role as an electron carrier in methanogenesis (Deppenmeier, 2002), but also occurs in some other organisms (Lin and White, 1986; Selengut and Haft, 2010). Interestingly, reduced F420 is used by mycobacteria for an antioxidant defence system (Hasan *et al*., 2010) and protection from nitrosative damage (Purwantini and Mukhopadhyay, 2009). Consistent with the finding of an F420-dependent enzyme, *N. gargensis* encodes all key genes for the synthesis of this cofactor (we identified these genes also in all other thaumarchaeal genomes) and typical F420 emission spectra were recorded from single *N. gargensis* cells (Fig. 5B). Remarkably, *N. gargensis* produces significant amounts of F420 as its characteristic blue fluorescence can be detected microscopically.

Interestingly, all thaumarchaeota including *N. gargensis* encode a polyhydroxyalkanoate synthase (Figs 2 and S4, Table S1), the key enzyme for the production of polyhydroxyalkanoates (PHAs), which are widely distributed carbon polyester storage compounds of bacteria but seemed to be confined within the archaea to the haloarchaea (Grage *et al*., 2009; Poli *et al*., 2011). The PHA synthase of thaumarchaeota is encoded by two overlapping genes (*phaEC*). The *phaC* subunit shows highest similarity to the respective subunit of haloarchaeal and bacterial type III PHA synthases (Rehm and Steinbüchel, 1999; Lu *et al*., 2008; Han *et al*., 2010; Poli *et al*., 2011). In contrast to other thaumarchaeota, the *N. gargensis* genome harbours three additional *phaC* genes that are not clustered with the *phaEC* operon, with 31–37% AA sequence identity to *phaC1* [a feature that has also been described in *Haloferax mediterranei* (Han *et al*., 2010)]. Synthesis of PHAs in thaumarchaeota likely starts from the carbon fixation pathway intermediate 3-hydroxybutyryl-CoA (Figs 2 and S4). We also experimentally confirmed the production of PHA (putatively polyhydroxybutyrate) by *N. gargensis* and *N. viennensis* using Raman microspectroscopy (Fig. 6) possibly reflecting unbalanced growth conditions (Anderson and Dawes, 1990). Although a close homologue of so far characterized PHA depolymerases is absent from thaumarchaeota, the presence of several hydrolases in *N. gargensis* suggests the possibility of PHA degradation.

Peculiarities of the carbon metabolism of N. gargensis

Although on a first glimpse the central carbon metabolism of *N. gargensis* appears to be similar to thaumarchaeota of the *Nitrosopumilus* cluster, it is noteworthy that *N. gargensis* seems more flexible not only in the uptake (see below) but also in the production and conversion of the central metabolite pyruvate (Figs 2 and S4, Table S1) that strongly enhances growth in its close relative *N. viennensis* (Tourna *et al*., 2011). As mentioned above, *N. gargensis* has a gene with homology to malic enzyme that catalyses the reversible conversion of malate to pyruvate and could thus be involved in gluconeogenesis and anaplerosis. Furthermore, *N. gargensis* but not the other thaumarchaeotes encodes a putative pyruvate oxidase/

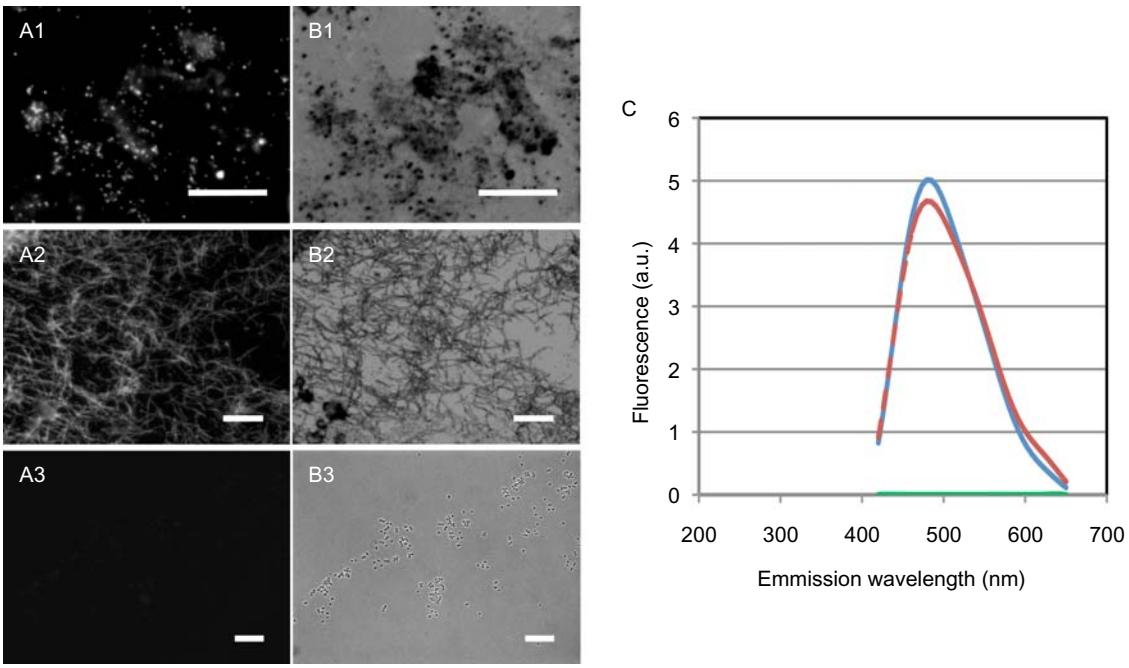

Fig. 5. Fluorescence (A) and phase contrast microscopy (B) of *Nitrososphaera gargensis* enrichment culture (lane 1), *Methanospirillum hungatei* DSM 864 (lane 2) and *Nitrosomonas europaea* Nm50 (lane 3). Bar equals 10 μm. The methanogen and *N. gargensis* displayed the typical F420 fluorescence while no fluorescence was visible for *N. europaea* cells. (C) The presence of F420 was confirmed by recording Lambda scans from individual cells of *M. hungatei* (red), *N. gargensis* (blue) and *N. europaea* (green) using a confocal laser scanning microscope with a 405 nm diode for excitation. The displayed spectra were smoothed by Fourier transformation. For each organism 10 highly similar single-cell emission spectra were recorded and a representative spectrum is shown.

dehydrogenase, which produces acetate or acetyl-phosphate from pyruvate. In addition, and again in contrast to other studied thaumarchaeotes, *N. gargensis* encodes an alanine dehydrogenase that catalyses the reversible deamination of alanine to pyruvate (Figs 2, S4 and S6, and *Supporting information*). As *N. gargensis* possesses several amino acid and an oligo/dipeptide transporter, alanine might thus represent an important source for pyruvate for this organism. Alternatively, alanine formation from pyruvate that can also be catalysed by this enzyme (see *Supporting information*) might not only be important for protein biosynthesis but could also have a role in ammonia storage and ammonia toxicity alleviation.

A remarkable feature of *N. gargensis* is the presence of two paralogous genes encoding the key enzyme mannosyl-3-phosphoglycerate synthase that is involved in the two-step synthesis of the compatible solute mannosylglycerate of a few, mostly (hyper)thermophilic archaea and bacteria from GDP mannose (Empadinhas *et al*., 2001, Empadinhas and da Costa, 2006; 2008; 2011) (Figs 2, S4 and S7). These microbes use mannosylglycerate for cellular osmotic adjustment and thermal protection (Neves *et al*., 2005), which is particularly interesting as *N. gargensis* is a moderate thermophile (Hatzenpichler *et al*., 2008). This feature is apparently more common in the *Nitrososphaera* and *Nitrosocaldus* clusters (but absent in the *Nitrosopumilus* cluster) as mannosyl-3-phosphoglycerate synthase genes are also present on environmentally retrieved fosmids carrying genome fragments from members of these clades (acc. BAE95227, CAD42692, CAF28675) (Quaiser *et al*., 2002). However, the synthesis of this compatible solute requires a second enzyme, mannosyl-3-phosphoglycerate phosphatase, which is not encoded by *N. gargensis*. In this respect it is worth mentioning that one of the two mannosyl-3-phosphoglycerate synthase genes of *N. gargensis* and the one encoded by soil fosmid 29i4 are located next to an uncharacterized phosphatase belonging to the superfamily of haloacid dehalogenases that encompasses canonical mannosyl-3-phosphoglycerate phosphatases (Burroughs *et al*., 2006). Thus, it seems possible that *N. gargensis* and some other thaumarchaeota are capable of producing mannosylglycerate.

Transport

The genome of *N. gargensis* encodes at least 116 putative full-length transporter proteins (Table S2), which are the structural elements of approximately 90 transport systems (some of them consisting of several proteins)

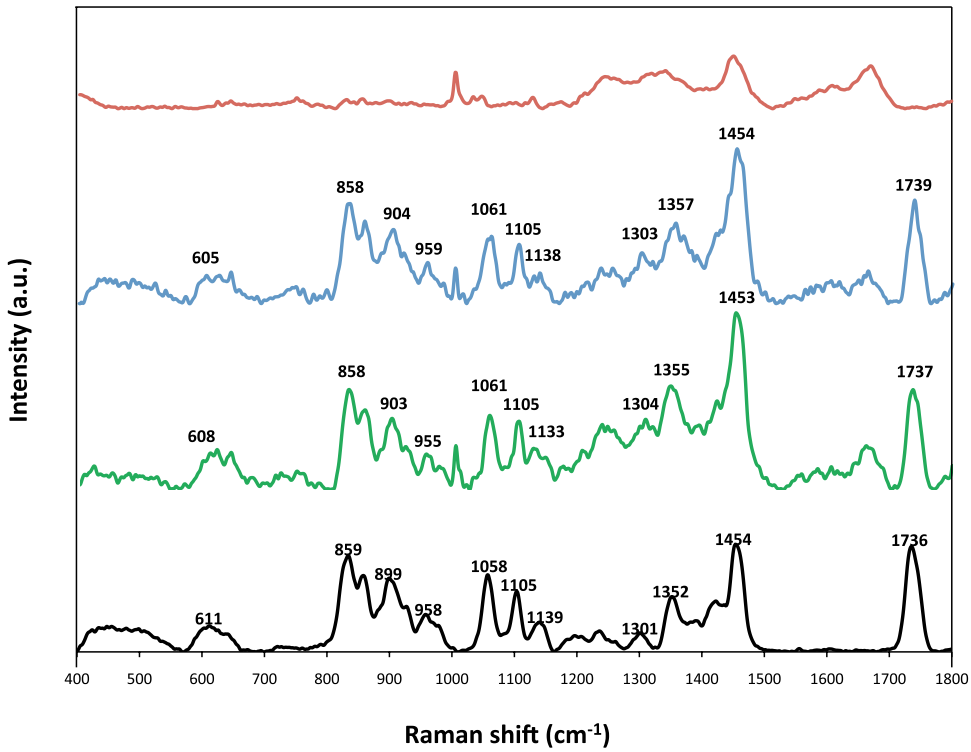

Fig. 6. Raman spectroscopy reveals PHA storage in thaumarchaeotes. Spectra of *N. gargensis* cells without (red) and with (blue) typical peaks of polyhydroxyalkanoates that include the signature peak at 1734 cm^{-1} (Gelder *et al.*, 2008) slightly shifted to 1739 cm^{-1} in *N. gargensis*. Both spectra were normalized by their phenylalanine peak. In addition, a spectrum from a *N. viennensis* cell also containing the typical PHA peaks is shown (green). For comparison a PHA reference spectrum from *Crenothrix polyspora* is displayed in black. The signature peak shift reflects varying abundances of amorphous versus crystalline PHA in different cells (Izumi and Temperini, 2010). Consequently, the analysed *N. gargensis* and *N. viennensis* contained mostly amorphous PHA consistent with the absence of the crystalline PHA signature peak at around 434 cm^{-1} (Murakami *et al.*, 2007). Only a fraction of the *N. gargensis* cells in the enrichment culture showed indications for storage of polyhydroxyalkanoates.

and represent 41 transporter families (Busch and Saier, 2002; Saier *et al.*, 2006; 2009). Furthermore, all crucial components of the general secretion (Sec) and twin-arginine translocation (Tat) pathways have been identified (*Supporting information*). Normalized to its genome size the number of transporter proteins is around average for an archaeon, with about 40 transporter per Mb of genome (Paulsen *et al.*, 2000; Albers *et al.*, 2007), and compares well with other autotrophic members of this domain (which on average do not differ from heterotrophic archaea regarding transporter-coding gene density; Ren *et al.*, 2004; 2007) although electrochemical potential-driven transporters are underrepresented (*Supporting information*). Among the *N. gargensis* transport proteins, those belonging to the ATP binding cassette (ABC) superfamily are the most represented ($n = 38$) and are the constituent of 12 transport systems. The complement of transport systems in *N. gargensis* is roughly reminiscent of other thaumarchaeota but 41 of the *N. gargensis* transport proteins do not have a homologue in the other genomes.

Nitrososphaera gargensis possesses transporters for ammonia, urea and cyanate (discussed above), which we predict as substrates for its energy metabolism. In detail, *N. gargensis* contains three transporters of the ammonia channel transporter (Amt) family. Interestingly, these thaumarchaeal genes have been shown to be among the most highly transcribed ones in three independent metatranscriptome studies (Hollibaugh *et al.*, 2011; Radax *et al.*, 2012; Stewart *et al.*, 2012). These transporters may be ammonia or CO_2 gas channels or serve for ammonium or even methylammonium transport (Zheng *et al.*, 2004; Andrade and Einsle, 2007; Fong *et al.*, 2007; Loqué *et al.*, 2007; Musa-Aziz *et al.*, 2009) and most likely are important for nitrogen assimilation and possibly for sensing/regulation in *N. gargensis* (Lorenz and Heitman, 1998; Yakunin and Hallenbeck, 2002). *Nitrososphaera gargensis* possesses in addition the genes for two urea transporters. Those genes are located close to the urease gene cluster. One of these urea transporters is homologous to the predicted urea transporter of *C. symbiosum* (Hallam *et al.*, 2006a) and is an electrochemical

potential-driven transporter belonging to the solute:sodium symporter (SSS) family. The second urea transporter is particularly interesting as it belongs to the urea transporter (UT) family that has not been documented so far in archaea (Levin *et al*., 2009) and is only present in the close relative *N. viennensis* (Tourna *et al*., 2011). Phylogenetic analysis suggests that *N. gargensis* acquired this transporter from a bacterial donor (Fig. S8). It is interesting to note that similar transporters are present in AOB as well as in *Desulfovibrio vulgaris*, whose UT transporter has been structurally and functionally characterized recently (Levin *et al*., 2009) and was found to be a channel. The functional difference between the two urea transporters of *N. gargensis* is, however, currently unclear.

Another remarkable feature of the transporter complement is the presence of 26 transport systems that might be involved in the uptake of organic molecules [among which the major facilitator superfamily (MFS) is the most represented] and which complements the higher flexibility of *N. gargensis* in the central carbon metabolism. Among these transporters 15 do not have a homologue in other thaumarchaeota suggesting broader scavenging capabilities of *N. gargensis*. For example, four putative amino acid uptake systems are present and two of these are not found in published thaumarchaeal genomes. In addition, *N. gargensis* also possesses, like the other thaumarchaeota, an oligo/dipeptide transport system of the ABC superfamily as well as several aminopeptidases (methionyl-, leucyl- and a putative alanyl-aminopeptidase, Ngar_c01308, Ngar_c34750, Ngar_c35070), which could be used to release amino acids from the imported peptides. The uptake of amino acids or peptides may help reducing the energy cost of protein synthesis in this autotrophic organism, but could also serve for acquisition of substrates for anaplerotic reactions.

In addition, *N. gargensis* has two closely related transport systems, which belong to the bile acid/arsenite/riboflavin transporter (BART) superfamily, which are absent from the other thaumarchaeota. Phylogenetic analysis places these two transporters within the ACR3 family where they form a previously unrecognized cluster with other archaeal and bacterial sequences (Fig. S9). Two members of the ACR3 family were shown to be arsenite exporters, and uncharacterized family members might be involved in the transport of organic molecules while the bile acid sodium symporter (BASS) family shows broad specificity for organic substrates (Mansour *et al*., 2007). It is particularly noteworthy that a BASS transporter found in plants is a sodium-dependent pyruvate transporter (Furumoto *et al*., 2011) and the presence of uncharacterized BART transporters in *Nitrososphaera* species might thus be linked to the pyruvate dependence of *N. viennensis* (Tourna *et al*., 2011).

In addition, *N. gargensis* has a putative sugar transporter (MFS superfamily), which is absent from other thaumarchaeal genomes. Furthermore, it might take up glycerol and sulfonates (but no indications for degradative enzymes were found) via similar transporters previously described in *N. maritimus* (Walker *et al*., 2010). Since *N. gargensis* possesses several enzymes putatively involved in glycerol-P conversions (Fig. 2), it is tempting to speculate that glycerol enhances growth of this organism.

Motility, chemotaxis and two-component systems

Similar to many archaea and the thaumarchaeon *N. limnia*, *N. gargensis* is equipped with numerous motility-associated genes (Table S1). First of all, it encodes all genes required to assemble an archaeal flagellar apparatus that is composed of the flagellar filament, the motor, and its switch and a flagellum could be observed by electron microscopy (Fig. 7A). The biosynthesis and assembly of archaeal flagella as well as the structural protein(s) and some accessory proteins share homology with bacterial type IV pili (Thomas *et al*., 2001; Bardy *et al*., 2004; Szabó *et al*., 2007a), which are involved in surface adhesion, cell–cell contact, autoaggregation, twitching motility and DNA uptake (Craig and Li, 2008; Albers and Pohlschröder, 2009; Ellen *et al*., 2010; Pohlschroder *et al*., 2011). The *fla* gene cluster of *N. gargensis* contains six genes including one gene for structural flagellin subunit FlaB/FlaA as well as the flagellar accessory genes *flaG*, *flaF*, *flaH*, *flaJ* and *flaI* (Fig. 7B). In addition, three copies of the membrane peptidase FlaK, which is involved in the preprocessing of flagellins, could be identified in the genome. Overall, the *fla* genes of *N. gargensis* show similarities to the type II cluster (*fla2*) of *Sulfolobales*, which is characterized by the presence of a single *flaB*/*flaA* gene (Faguy *et al*., 1996; Albers *et al*., 2003; Szabó *et al*., 2007b) and genes for the accessory subunits FlaC, FlaD/E as well as by a conserved gene order of *flaG* and *flaH* (Albers *et al*., 2003; Desmond *et al*., 2007). Although the flagella gene cluster of *N. limnia* is closely related to that in *N. gargensis* the former one differs by the presence of four *flaB* copies (Fig. 7B).

Besides the *fla* genes, *N. gargensis* possesses at least one pili gene cluster that shows remote similarity to the two characterized pili operons (*ups* and *bas* pili operons) of *Sulfolobus solfataricus* P2 (Zolghadr *et al*., 2007; 2011; Fröls *et al*., 2008). The cluster of *N. gargensis* encompasses a putative membrane bound type IV pili protein (*basF*-like) and a type IV pilus assembly ATPase (*basE*-like) that belong to the same protein families like FlaJ and

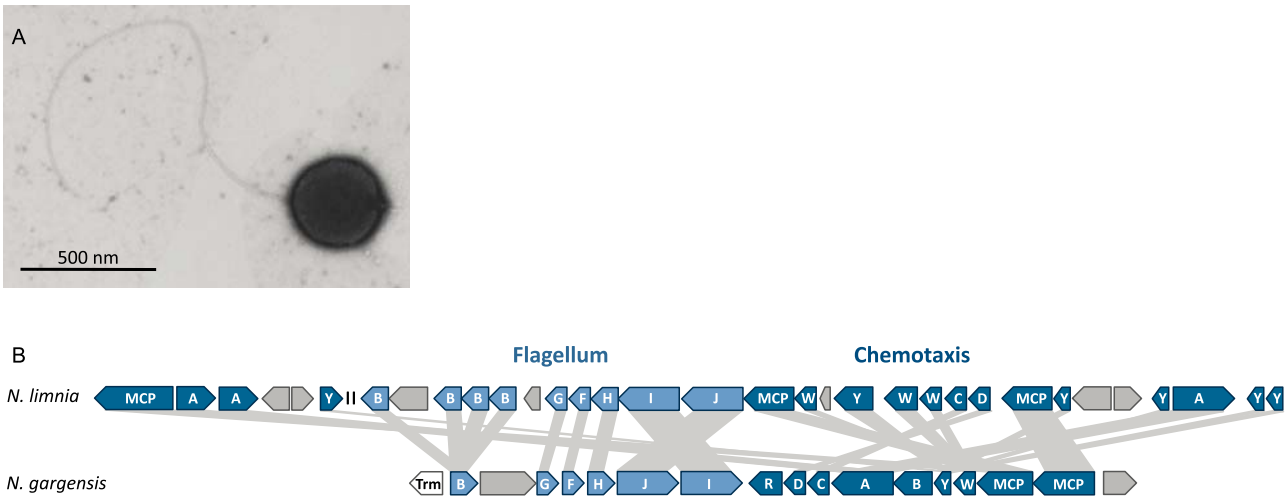

Fig. 7. A. EM picture of *N. gargensis* showing a flagellum.
B. Organization of flagellar and chemotaxis genes in thaumarchaeota. Flagellum-associated genes (*flaB*, *flaG*, *flaF*, *flaH*, *flaJ*, *flaI*) are shown in light blue, chemotaxis genes are shown in dark blue (MCP, methyl-accepting chemotaxis protein; *cheA*, histidine kinase; *cheY*, response regulator; *cheW*, adapter protein; *cheC*, *cheD*, *cheR*, *cheB*, adaptation proteins) and the gene for the transcriptional regulator TrmB is shown in white. Genes coding for unknown proteins are shown in grey. A II sign indicates separated genome regions.

FlaI respectively (Ghosh and Albers, 2011). Furthermore, an uncharacterized protein and two putative pilin proteins with similarity to the BasA and UpsA proteins of *S. solfataricus* P2 could be identified. Archaeal pili have been suggested to play a role in DNA acquisition (Ajon *et al.*, 2011), the uptake of extracellular substrates (Zolghadr *et al.*, 2011) and/or cell aggregation (Pohlschroder *et al.*, 2011).

Interestingly, in contrast to crenarchaeota but similar to euryarchaeota and *N. limnia* (Desmond *et al.*, 2007; Blainey *et al.*, 2011), *N. gargensis* harbours all genes for chemotaxis (*cheR*, *cheD*, *cheC*, *cheA*, *cheB*, *cheY*, *cheW*) that are encoded in a single gene cluster next to the *fla* genes (Fig. 7B). In addition, 25 histidine kinase-like proteins (nine of which are membrane-associated) and 23 putative signal transduction response regulators were identified. Many of these histidine kinase and response regulator-encoding genes are located in close proximity to each other suggesting that they are involved in two-component signal transduction pathways allowing *N. gargensis* to respond to various environmental stimuli. Whereas *N. limnia* has a comparable number of genes putatively involved in signalling cascades including the genes associated with chemotaxis ($n = 52$), other thaumarchaeota of the *Nitrosopumilus* cluster encode significantly less signal transduction proteins [*N. maritimus* ($n = 24$), *Nitrosoarchaeum koreensis* ($n = 26$) and *C. symbiosum* ($n = 6$)]. From a phylogenetic perspective it is interesting to note that crenarchaeota (Ashby, 2006) and korarchaeota do not encode two-component systems which is consistent with the notion that AOA are members of a separate phylum (Brochier-Armanet *et al.*, 2008; Spang *et al.*, 2010).

Gas vesicles

In contrast to all other thaumarchaeota, *N. gargensis* contains one gene cluster encoding the gas vesicle proteins GvpA, J, K, L, F and G (in this order) and a distantly located homologue of GvpD. In addition, the locus includes a small hypothetical protein that might represent a *N. gargensis*-specific gas vesicle protein and a protein with a CobQ/CobB/MinD/ParA-like nucleotide binding domain directly upstream of the gene cluster (Table S1). Although *N. gargensis* does not contain homologues of GvpM and GvpO that belong to the eight proteins required for gas vesicle formation in haloarchaea (also present in *Bacillus megaterium*) (Li and Cannon, 1998; Offner *et al.*, 2000), it might still be able to synthesize gas vesicle proteins in analogy to gas vesicle-carrying cyanobacteria that also lack *gvpM* and *gvpO* and have smaller *gvp* operons (Pfeifer, 2006). So far, gas vesicles have not been observed in the enrichment culture (by screening many TEM images) and the ability of *N. gargensis* to produce gas vesicles has thus to be verified. In this respect it is, however, interesting to mention that the gas vesicle genes are flanked by genes encoding two signal transduction histidine kinases (one of which is membrane bound) and one associated response regulator which might suggest that gas vesicle production of *N. gargensis* is regulated by a two-component system and is dependent on specific environmental conditions.

Heavy metal resistance

Nitrososphaera gargensis has been enriched from a runoff basin of the Garga spring, a slightly alkaline (pH 7.9) hot spring located in the Baikal rift zone, Russia (Lebedeva *et al*., 2005). The concentration of several heavy metals in this spring is between 2- and 2000-fold higher than in pristine freshwater or marine environments (Table S6). Consistently, *N. gargensis* is well adapted to living in a habitat containing heavy metals as it encodes multiple resistance mechanisms including at least 21 putative metal ion efflux proteins belonging to 10 different transporter families. Most noteworthy, eight of these putative metal efflux proteins, representing five transporter families, do not have a homologue in any of the other sequenced thaumarchaeota.

The high concentration of nickel in the Garga spring correlates well with the presence of six putative nickel transporter genes in the *N. gargensis* genome (*Supporting information*). All these transporters are absent in *C. symbiosum* and only one of them is present in *N. maritimus*, *N. limnia* and *N. koreensis*. Some of these *N. gargensis* transporters (CDF-type) might also play a role in tolerance to the relatively high levels of zinc and cobalt in the Garga spring. *Nitrososphaera gargensis* possesses in addition homologues of the CorA and CorC proteins which function together as a Co efflux system and may confer Co^{2+} resistance (Gibson *et al*., 1991). Another permease in the *N. gargensis* genome which belongs to the Ca^{2+}:cation antiporter family and which is absent from thaumarchaeota of the *Nitrosopumilus* cluster might also be a heavy metal efflux pump and confer resistance to Zn^{2+} and/or Co^{2+} ions (Shigaki *et al*., 2005).

The genome of *N. gargensis* encodes two NADPH-dependent FMN reductases and at least four nitroreductases that could have a chromate reductase activity (Schröder *et al*., 2003; Ramírez-Díaz *et al*., 2008) (see also iron uptake in *Supporting information*) and are thought to be determinants of chromate resistance in bacteria (Ramírez-Díaz *et al*., 2008). This might therefore explain the presence of *N. gargensis* in an environment containing a high level of chromium (Table S6).

In accordance with the high concentration of copper in the Garga spring (Table S6) *N. gargensis* possesses a copper-translocating P-type ATPase (CopA homologue), which may be involved in copper tolerance (Chan *et al*., 2010) (*Supporting information*). Another putative copper transport system (CopC–CopD) is present in *N. gargensis* and all other thaumarchaeota and might provide additional resistance to copper (Silver and Ji, 1994; Lee *et al*., 2002).

Lithium and cesium, which are toxic to both micro- and macroorganisms (Inaba *et al*., 1994; Avery, 1995; Mechold *et al*., 2006; Zhang *et al*., 2006), are also much more concentrated in the Garga spring than in other aquatic systems (Table S6). Cesium is the most toxic of the alkali metals and notably for an organism living in an alkaline habitat its toxicity increases with pH (Perkins and Gadd, 1995). Resistance to Li^+ in *E. coli* is partly mediated by Na^+/H^+ antiporters (Inaba *et al*., 1994; Liu *et al*., 2005) and it is reasonable to assume that resistance to Cs^+ might also be conferred by monovalent cation efflux systems such as K^+/H^+ antiporters (Avery, 1995). Therefore, it is interesting to note that *N. gargensis* possesses at least seven monovalent cation/H^+ antiporters and seems well adapted to living in the presence of high lithium and cesium concentrations. Notably, all these antiporters belong to the monovalent cation:proton antiporter-2 (CPA2) family that is less abundant (0–5 copies) in all other archaeal genomes (Ren *et al*., 2004; 2007).

Conclusions and outlook

Manual annotation of the *N. gargensis* genome confirmed that the gene set for ammonia oxidation and CO_2 fixation is conserved in different thaumarchaeal lineages, but also revealed many features not yet found in other thaumarchaeota. Like several other thermophiles, *N. gargensis* likely produces mannosylglycerate as compatible solute. Furthermore, it is well equipped for sensing changing environmental conditions by its array of two component systems and has the genetic potential to relocate itself by active swimming and possibly by using gas vacuoles. It also encodes an impressive collection of resistance mechanisms that will confer a selective advantage for living in the spring runoff waters that contain high concentrations of various heavy metals. Unfortunately, no information on cyanate concentrations in the spring water is currently available. Cyanate can be formed as metabolic by-product, but can also originate from anthropogenic activities (Koshiishi *et al*., 1997; Kraus and Kraus, 1998) and from the spontaneous photooxidation of cyanide (Nowakowska *et al*., 2006), which is produced largely by plants (Peiser *et al*., 1984) and could thus represent an interesting substrate in several environments. We will explore in an experimental follow-up study whether cyanate can indeed be used by *N. gargensis* as alternative energy source. In addition, *N. gargensis* has more flexibility regarding its central carbon metabolism than other thaumarchaeota including a putative oxidative pentose phosphate pathway and the capability to use a variety of organic compounds including alanine, glycerol, pyruvate and even cyanate. If also present in other members of this clade, this metabolic flexibility, which is also consistent with the recently reported strong growth stimulating effect of pyruvate for *N. viennensis* (Tourna *et al*., 2011), might be an essential factor for the

competitive success of the *Nitrososphaera* cluster in many terrestrial environments. It will be exciting to explore further whether members of this lineage can grow mixo- or possibly even heterotrophically as recently suggested for a close relative of *N. gargensis* thriving in an industrial waste water treatment plant (Mußmann *et al*., 2011).

Experimental procedures

Biomass production for genome sequencing and experimental analyses

The ammonia-oxidizing *N. gargensis* enrichment culture containing approximately 90% of this organism (compared with DAPI-stained cells) according to CARD–FISH using probe RHGA702 (Hatzenpichler *et al*., 2008) was grown in modified mineral medium (Hatzenpichler *et al*., 2008; Pitcher *et al*., 2010) with 0.5 mM NH_4Cl. A total of 4.5 l of medium was prepared in 5 l bottles and inoculated with 50 ml of an active preculture. Incubation was carried out in the dark at 46°C for 3 weeks with moderate stirring (150 r.p.m.). The pH was adjusted to 7.8 and kept constant by daily titration using 15% (w/v) $NaHCO_3$. The consumption of ammonia was regularly measured by test sticks (Merck KGaA, Darmstadt, Germany) followed by readjusting the substrate concentration by adding small amount of a 5 M sterile stock solution. In total, 3–4 mM NH_4Cl was oxidized before cells were harvested by centrifugation (14 000 *g*) and washed in 0.9% NaCl. Several 5 l bottles were collected and the pellet was stored at −20°C until further treatment. In the enrichments used in this study, a single archaeal OTU [99% 16S rRNA sequence homology to *N. gargensis* studied by Hatzenpichler *et al*. (2008)] was present according to 16S rRNA analysis.

Electron microscopy

For observation by electron microscopy, cells from a 5 l culture were collected at 15 000 *g*, washed in 0.9% NaCl and fixed with 2.5% (v/v) glutaraldehyde and 1% (w/v) osmium tetroxide as described by Spieck and Lipski (2011). After dehydration embedding was performed in Spurr. The dried samples were sectioned with a diamond knife, collected on copper grids and double-stained with 2% uranylacetate in distilled water and 2% lead citrate for 10 min respectively. Examination was carried out with a transmission electron microscope (Zeiss model Leo 906E with a CCD camera model 794). For whole-cell preparations a few millilitres of the culture was centrifuged (15 000 *g*), the harvested cells dropped on a copper grid, dried and stained with uranyl acetate in distilled water (2%).

Raman spectroscopy

For Raman microspectroscopy, enrichments of *N. gargensis* and *N. viennensis* were dried onto aluminum slides, washed with ddH_2O and placed dry under a confocal LabRAM HR800 Raman microspectrometer (Horiba, Germany) equipped with a 50 mW 532.17 nm laser. Before the samples were analysed, the system was aligned using a silica reference with a distinct Raman peak at 520 cm^{-1}. Cells for Raman analysis were chosen by morphology (for *N. gargensis*) in the live-view mode of the Labspec software, ver. 5.25.15 (Horiba) using an Olympus 100 × MPlan N objective (NA 0.9, Olympus, Japan). The exposure time for all cells was set to 40 s with the laser intensity dimmed to 50% by an intensity filter to prevent photo damage. The confocal pinhole was set to 300 μm, leading to an axial resolution of about 3 μm. Raman spectra were acquired between 400 and 1800 cm^{-1}. Raman spectra were baseline-corrected, normalized and exported to a file format readable by Excel (Microsoft).

Detection of F420

Formaldehyde-fixed cells of *Methanospirillum hungatei* (DSM 864), *Nitrosomonas europaea* Nm50 and the *N. gargensis* enrichment culture were applied to a microscopic slide and let dry. The slides were briefly dipped into deionized water to wash away residual salts and air-dried again. For image acquisition the slides were placed directly or embedded in 50% glycerol in PBS (30 mM sodium phosphate pH 5, 390 mM NaCl) under a LMD 7000 (Leica, Germany) equipped with a BGR for LMD filter cube, a 63 × air objective for dry samples or a 100 × oil immersion objective for embedded samples. Of the BGR for LMD filter cube only the BP 420/30 excitation was used. All images of the different samples were acquired with the same settings with the LAS software package (version 3.6, Leica) delivered with the instrument. Lambda scans were acquired with a Leica SP5X confocal laser scanning microscope using the 405 nm diode for excitation.

Sequencing, gap closure, annotation and bioinformatic tools

DNA for 454 sequencing was isolated using the peqGOLD Bacterial DNA Kits (peqlab Biotechnology GmbH, Erlangen, Germany). The 454 shotgun sequencing was carried out according to the manufacturer's protocols (Roche Applied Science, Mannheim, Germany) and yielded approximately 200 Mb sequence information. The raw 454 reads were assembled using the 454 *de novo* Newbler Assembly software, which yielded 6705 contigs with a length up to 370 784 nt (4.8 Mb non-redundant sequence information). For scaffolding a metagenomic fosmid library was constructed according to the manufacturer's protocols in pCC1FOS with *E. coli* EPI300 as heterologous host (Epicentre Biotechnologies, Madison, USA) using DNA isolated by classical methods with chemical and enzymatic cell lysis followed by phenol chloroform extraction and a final dialysis (Sambrook and Russell, 2001). In total, 2500 clones were picked and stored at −70°C. End sequencing of 500 randomly chosen fosmid clones by Sanger sequencing allowed to physically map 88 contigs into one large scaffold with a size of 2.8 Mb. PCR-based techniques (Schmeisser *et al*., 2009) were used to close the remaining gaps using genomic DNA as template. All manual editing steps were performed using the GAP4 software package v4.5 and v4.6 (Schiex *et al*., 2003). Coding sequences (CDS) and ORFs were predicted with YACOP (Tech and Merkl, 2003) using the ORF finders Glimmer, Critica and Z-curve. All CDS were manually curated

and verified by comparison with the publicly available databases SwissProt (Magrane and UniProt Consortium, 2011), InterproScan via EMBL (Cochrane *et al*., 2009), GenBank (Benson *et al*., 2011), COG (Tatusov *et al*., 1997) and Prosite (Sigrist *et al*., 2010) using the annotation software ERGO (Overbeek *et al*., 2003) as well as the comparative genome analysis and manual functional annotation platform Microscope (Vallenet *et al*., 2006; 2009). CRISPR loci were identified using CRISPRFinder (Grissa *et al*., 2007). All genes discussed in the paper have been manually annotated using a slightly modified version of the annotation parameters described in Lücker and colleagues (2010). These included the annotation of coding sequences (CDS) that had an amino acid identity $\geq 35\%$ and sequence coverage $> 75\%$ to proteins in the SwissProt/Trembl database (Magrane and UniProt Consortium, 2011) as homologous to proteins with known function (class 2). Protein sequences with an amino acid identity between 25% and 35% have been annotated as putative homologue to a protein with a specific function or as member of a protein family (class 3). CDS with lower sequence identities to proteins in the mentioned databases and/or with protein coverage below 75% were annotated as hypothetical proteins (eventually with conserved domains if detected) (class 4). Furthermore, additional information from protein signatures of the InterPro database (Hunter *et al*., 2009), from the COGs database (Tatusov *et al*., 1997) as well as enzyme profile data provided by PRIAM (Claudel-Renard *et al*., 2003) and HAMAP (Lima *et al*., 2009) were taken into account to refine the annotation of proteins of all annotation classes presented above. Transporter gene annotations were carried out by additionally taking into account the information provided in the transporter classification database (Saier *et al*., 2006; 2009). Detailed information on the annotated genes ordered by functional category can be found in Tables S1–S5.

The completed 2.83 Mb genome has been submitted to GenBank under Accession Number CP002408.

Phylogenetic analyses

Predicted amino acid sequences of selected genes were automatically aligned with reference sequences using the multiple sequence alignment tool Clustal2W (Larkin *et al*., 2007) provided by the European Bioinformatics Institute (EMBL-EBI) (Goujon *et al*., 2010) and manually corrected with the sequence editor BioEdit (Hall, 1999). If not mentioned otherwise, phylogenetic trees were constructed using the maximum likelihood software PhyML (Guindon *et al*., 2010) (default parameters and 100 bootstraps) and visualized with the tree-drawing program NJPlot (Perrière and Gouy, 1996).

Genome synteny comparisons

Pairwise alignments for dot plot representations were performed on the six-frame amino acid translation of the genome sequences of *Nitrososphaera gargensis*, *N. maritimus* and *N. koreensis* using the Promer program in the MUMmer 3.23 package (Kurtz *et al*., 2004). For all analyses default parameters were applied; that is, exact matches longer than six amino acids were identified, adjacent exact matches were joined if separated by a gap no longer than 30 amino acids, the resulting clusters were further processed if the total length of their matches was longer than 20 amino acids and then aligned using a BLOSUM62 amino acid substitution matrix.

Calculation of COGs for Venn diagrams

The predicted proteome sequences of selected genomes (*M. sedula*, *Halobacterium salinarum*, *C. subterraneum*, *N. maritimus*, *C. symbiosum*, *N. koreensis* and *N. limnia*) except for *N. gargensis* were retrieved from the July 2012 release of the NCBI RefSeq database (Pruitt *et al*., 2012). Bidirectional best sequence similarities were obtained from the SIMAP database (Rattei *et al*., 2010) using two cut-off values: maximal *E*-value = $1e^{-10}$; minimal coverage of both proteins by local alignment = 70%. Bidirectional best hits (BBH) between all genomes were clustered into COGs using a SIMAP-based implementation of the *COGtriangles* algorithm (Kristensen *et al*., 2010). Due to the small number of involved genomes, also isolated BBH not part of any triangle were treated as COGs.

Acknowledgements

We would like to thank Rolf Thauer, Sonja V. Albers, Sebastian Lücker and Holger Daims for helpful discussions as well as A.M. Plyusnin for making available the heavy metal data from Garga spring. Furthermore, we are grateful to the LABGeM bioinformatics team of Genoscope for providing access to the MicroScope platform. M. Wagner and C. Böhm were supported by the ERC Advanced Grant NITRICARE (Grant No.: 294343). S. Haider was supported by a grant from the FWF (P20775) to M. W. A. Spang was supported by a DOCfForte fellowship from the Austrian Academy of Sciences, P. Offre was funded by a research grant (09-EuroEEFG-FP-034) from the European Science Foundation and C. Schleper was supported by the FWF Grant P23000. E. Spieck was supported by a research grant of the DFG (DFG SP 667/7-1).

References

Ajon, M., Fröls, S., van Wolferen, M., Stoecker, K., Teichmann, D., Driessen, A.J.M., *et al*. (2011) UV-inducible DNA exchange in hyperthermophilic archaea mediated by type IV pili. *Mol Microbiol* **82:** 807–817.

Akita, M., Adachi, A., Takemura, K., Yamagami, T., Matsunaga, F., and Ishino, Y. (2010) Cdc6/Orc1 from *Pyrococcus furiosus* may act as the origin recognition protein and Mcm helicase recruiter. *Genes Cells* **15:** 537–552.

Al-Attar, S., Westra, E.R., van der Oost, J., and Brouns, S.J.J. (2011) Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. *Biol Chem* **392:** 277–289.

Albers, S.V., and Pohlschröder, M. (2009) Diversity of archaeal type IV pilin-like structures. *Extremophiles* **13:** 403–410.

Albers, S.V., Szabó, Z., and Driessen, A.J.M. (2003) Archaeal homolog of bacterial type IV prepilin signal peptidases with broad substrate specificity. *J Bacteriol* **185:** 3918–3925.

Albers, S.V., Konings, W.N., and Driessen, A.J. (2007) Solute transport. In *Archaea: Molecular and Cellular Biology*. Cavicchioli, R. (ed.). Washington, DC, USA: American Society for Microbiology, pp. 354–368.

Anderson, A.J., and Dawes, E.A. (1990) Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. *Microbiol Rev* **54:** 450–472.

Andrade, S.L.A., and Einsle, O. (2007) The Amt/Mep/Rh family of ammonium transport proteins. *Mol Membr Biol* **24:** 357–365.

Arp, D.J., Chain, P.S.G., and Klotz, M.G. (2007) The impact of genome analyses on our understanding of ammonia-oxidizing bacteria. *Annu Rev Microbiol* **61:** 503–528.

Ashby, M.K. (2006) Distribution, structure and diversity of 'bacterial' genes encoding two-component proteins in the Euryarchaeota. *Archaea* **2:** 11–30.

Avery, S.V. (1995) Caesium accumulation by microorganisms: uptake mechanisms, cation competition, compartmentalization and toxicity. *J Ind Microbiol* **14:** 76–84.

Bandeiras, T.M., Salgueiro, C., Kletzin, A., Gomes, C.M., and Teixeira, M. (2002) *Acidianus ambivalens* type-II NADH dehydrogenase: genetic characterisation and identification of the flavin moiety as FMN. *FEBS Lett* **531:** 273–277.

Bardy, S.L., Ng, S.Y.M., and Jarrell, K.F. (2004) Recent advances in the structure and assembly of the archaeal flagellum. *J Mol Microbiol Biotechnol* **7:** 41–51.

Bartossek, R., Nicol, G.W., Lanzen, A., Klenk, H.-P., and Schleper, C. (2010) Homologues of nitrite reductases in ammonia-oxidizing archaea: diversity and genomic context. *Environ Microbiol* **12:** 1075–1088.

Bartossek, R., Spang, A., Weidler, G., Lanzen, A., and Schleper, C. (2012) Metagenomic analysis of ammonia-oxidizing archaea affiliated with the soil group. *Front Microbiol* **3:** 208.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2011) GenBank. *Nucleic Acids Res* **39:** D32–D37.

Berg, I.A. (2011) Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways. *Appl Environ Microbiol* **77:** 1925–1936.

Berg, I.A., Kockelkorn, D., Buckel, W., and Fuchs, G. (2007) A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science* **318:** 1782–1786.

Berg, I.A., Kockelkorn, D., Ramos-Vera, W.H., Say, R.F., Zarzycki, J., Hügler, M., *et al*. (2010a) Autotrophic carbon fixation in archaea. *Nat Rev Microbiol* **8:** 447–460.

Berg, I.A., Ramos-Vera, W.H., Petri, A., Huber, H., and Fuchs, G. (2010b) Study of the distribution of autotrophic CO_2 fixation cycles in Crenarchaeota. *Microbiology* **156:** 256–269.

Blainey, P.C., Mosier, A.C., Potanina, A., Francis, C.A., and Quake, S.R. (2011) Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLoS ONE* **6:** e16626.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. (2008) Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* **6:** 245–252.

Brochier-Armanet, C., Gribaldo, S., and Forterre, P. (2012) Spotlight on the Thaumarchaeota. *ISME J* **6:** 227–230.

Burroughs, A.M., Allen, K.N., Dunaway-Mariano, D., and Aravind, L. (2006) Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. *J Mol Biol* **361:** 1003–1034.

Busch, W., and Saier, M.H., Jr (2002) The transporter classification (TC) system, 2002. *Crit Rev Biochem Mol Biol* **37:** 287–337.

Chan, H., Babayan, V., Blyumin, E., Gandhi, C., Hak, K., Harake, D., *et al*. (2010) The p-type ATPase superfamily. *J Mol Microbiol Biotechnol* **19:** 5–104.

Claudel-Renard, C., Chevalet, C., Faraut, T., and Kahn, D. (2003) Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Res* **31:** 6633–6639.

Cochrane, G., Akhtar, R., Bonfield, J., Bower, L., Demiralp, F., Faruque, N., *et al*. (2009) Petabyte-scale innovations at the European Nucleotide Archive. *Nucleic Acids Res* **37:** D19–D25.

Craig, L., and Li, J. (2008) Type IV pili: paradoxes in form and function. *Curr Opin Struct Biol* **18:** 267–277.

Czyzewski, B.K., and Wang, D.-N. (2012) Identification and characterization of a bacterial hydrosulphide ion channel. *Nature* **483:** 494–497.

Deppenmeier, U. (2002) Redox-driven proton translocation in methanogenic Archaea. *Cell Mol Life Sci* **59:** 1513–1533.

Desmond, E., Brochier-Armanet, C., and Gribaldo, S. (2007) Phylogenomics of the archaeal flagellum: rare horizontal gene transfer in a unique motility structure. *BMC Evol Biol* **7:** 106.

Ellen, A.F., Zolghadr, B., Driessen, A.M.J., and Albers, S.-V. (2010) Shaping the archaeal cell envelope. *Archaea* **2010:** 608243.

Empadinhas, N., and da Costa, M.S. (2006) Diversity and biosynthesis of compatible solutes in hyper/thermophiles. *Int Microbiol* **9:** 199–206.

Empadinhas, N., and da Costa, M.S. (2008) Osmoadaptation mechanisms in prokaryotes: distribution of compatible solutes. *Int Microbiol* **11:** 151–161.

Empadinhas, N., and da Costa, M.S. (2011) Diversity, biological roles and biosynthetic pathways for sugar-glycerate containing compatible solutes in bacteria and archaea. *Environ Microbiol* **13:** 2056–2077.

Empadinhas, N., Marugg, J.D., Borges, N., Santos, H., and Costa, M.S. (2001) Pathway for the synthesis of mannosylglycerate in the hyperthermophilic archaeon *Pyrococcus horikoshii*. Biochemical and genetic characterization of key enzymes. *J Biol Chem* **276:** 43580–43588.

Estelmann, S., Hügler, M., Eisenreich, W., Werner, K., Berg, I.A., Ramos-Vera, W.H., *et al*. (2011) Labeling and enzyme studies of the central carbon metabolism in *Metallosphaera sedula*. *J Bacteriol* **193:** 1191–1200.

Faguy, D.M., Bayley, D.P., Kostyukova, A.S., Thomas, N.A., and Jarrell, K.F. (1996) Isolation and characterization of flagella and flagellin proteins from the Thermoacidophilic archaea *Thermoplasma volcanium* and *Sulfolobus shibatae*. *J Bacteriol* **178:** 902–905.

Filée, J., Siguier, P., and Chandler, M. (2007) Insertion sequence diversity in archaea. *Microbiol Mol Biol Rev* **71:** 121–157.

Fong, R.N., Kim, K.-S., Yoshihara, C., Inwood, W.B., and Kustu, S. (2007) The W148L substitution in the *Escherichia*

coli ammonium channel AmtB increases flux and indicates that the substrate is an ion. *Proc Natl Acad Sci USA* **104:** 18706–18711.

Francis, C.A., Roberts, K.J., Beman, J.M., Santoro, A.E., and Oakley, B.B. (2005) Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proc Natl Acad Sci USA* **102:** 14683–14688.

French, E., Kozlowski, J.A., Mukherjee, M., Bullerjahn, G., and Bollmann, A. (2012) Ecophysiological characterization of ammonia-oxidizing archaea and bacteria from freshwater. *Appl Environ Microbiol* **78:** 5773–5780.

Fröls, S., Ajon, M., Wagner, M., Teichmann, D., Zolghadr, B., Folea, M., *et al*. (2008) UV-inducible cellular aggregation of the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by pili formation. *Mol Microbiol* **70:** 938–952.

Fuchs, G. (2011) Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? *Annu Rev Microbiol* **65:** 631–658.

Fukuda, E., and Wakagi, T. (2002) Substrate recognition by 2-oxoacid:ferredoxin oxidoreductase from *Sulfolobus* sp. strain 7. *Biochim Biophys Acta* **1597:** 74–80.

Furumoto, T., Yamaguchi, T., Ohshima-Ichie, Y., Nakamura, M., Tsuchida-Iwata, Y., Shimamura, M., *et al*. (2011) A plastidial sodium-dependent pyruvate transporter. *Nature* **476:** 472–475.

Gelder, J.D., Willemse-Erix, D., Scholtes, M.J., Sanchez, J.I., Maquelin, K., Vandenabeele, P., *et al*. (2008) Monitoring poly(3-hydroxybutyrate) production in *Cupriavidus necator* DSM 428 (H16) with raman spectroscopy. *Anal Chem* **80:** 2155–2160.

Ghosh, A., and Albers, S.-V. (2011) Assembly and function of the archaeal flagellum. *Biochem Soc Trans* **39:** 64–69.

Gibson, M.M., Bagga, D.A., Miller, C.G., and Maguire, M.E. (1991) Magnesium transport in *Salmonella typhimurium*: the influence of new mutations conferring Co^{2+} resistance on the CorA Mg^{2+} transport system. *Mol Microbiol* **5:** 2753–2762.

Gomes, C.M., Bandeiras, T.M., and Teixeira, M. (2001) A new type-II NADH dehydrogenase from the Archaeon *Acidianus ambivalens*: characterization and *in vitro* reconstitution of the respiratory chain. *J Bioenerg Biomembr* **33:** 1–8.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* **38:** W695–W699.

Grage, K., Jahns, A.C., Parlane, N., Palanisamy, R., Rasiah, I.A., Atwood, J.A., and Rehm, B.H.A. (2009) Bacterial polyhydroxyalkanoate granules: biogenesis, structure, and potential use as nano-/micro-beads in biotechnological and biomedical applications. *Biomacromolecules* **10:** 660–669.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* **35:** W52–W57.

Gubry-Rangin, C., Hai, B., Quince, C., Engel, M., Thomson, B.C., James, P., *et al*. (2011) Niche specialization of terrestrial archaeal ammonia oxidizers. *Proc Natl Acad Sci USA* **108:** 21206–21211.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59:** 307–321.

Hall, T.A. (1999) BioEdit: a user-friendly tbiological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41:** 95–98.

Hallam, S.J., Konstantinidis, K.T., Putnam, N., Schleper, C., Watanabe, Y., Sugahara, J., *et al*. (2006a) Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc Natl Acad Sci USA* **103:** 18296–18301.

Hallam, S.J., Mincer, T.J., Schleper, C., Preston, C.M., Roberts, K., Richardson, P.M., and DeLong, E.F. (2006b) Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *PLoS Biol* **4:** e95.

Han, J., Hou, J., Liu, H., Cai, S., Feng, B., Zhou, J., and Xiang, H. (2010) Wide distribution among halophilic archaea of a novel polyhydroxyalkanoate synthase subtype with homology to bacterial type III synthases. *Appl Environ Microbiol* **76:** 7811–7819.

Hansen, T., Oehlmann, M., and Schönheit, P. (2001) Novel type of glucose-6-phosphate isomerase in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Bacteriol* **183:** 3428–3435.

Hansen, T., Schlichting, B., Felgendreher, M., and Schönheit, P. (2005) Cupin-type phosphoglucose isomerases (Cupin-PGIs) constitute a novel metal-dependent PGI family representing a convergent line of PGI evolution. *J Bacteriol* **187:** 1621–1631.

Hasan, M.R., Rahman, M., Jaques, S., Purwantini, E., and Daniels, L. (2010) Glucose 6-phosphate accumulation in mycobacteria: implications for a novel F420-dependent anti-oxidant defense system. *J Biol Chem* **285:** 19135–19144.

Hatzenpichler, R., Lebedeva, E.V., Spieck, E., Stoecker, K., Richter, A., Daims, H., and Wagner, M. (2008) A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. *Proc Natl Acad Sci USA* **105:** 2134–2139.

Hollibaugh, J.T., Gifford, S., Sharma, S., Bano, N., and Moran, M.A. (2011) Metatranscriptomic analysis of ammonia-oxidizing organisms in an estuarine bacterioplankton assemblage. *ISME J* **5:** 866–878.

Hommes, N.G., Kurth, E.G., Sayavedra-Soto, L.A., and Arp, D.J. (2006) Disruption of *sucA*, which encodes the E1 subunit of alpha-ketoglutarate dehydrogenase, affects the survival of *Nitrosomonas europaea* in stationary phase. *J Bacteriol* **188:** 343–347.

Horvath, P., and Barrangou, R. (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327:** 167–170.

Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., *et al*. (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res* **37:** D211–D215.

Inaba, K., Kuroda, T., Shimamoto, T., Kayahara, T., Tsuda, M., and Tsuchiya, T. (1994) Lithium toxicity and $Na^+(Li^+)/H^+$ antiporter in *Escherichia coli*. *Biol Pharm Bull* **17:** 395–398.

Izumi, C.M.S., and Temperini, M.L.A. (2010) FT-Raman investigation of biodegradable polymers: poly(3-hydroxybutyrate) and poly(3-hydroxybutyrate-co-3-hydroxyvalerate). *Vib Spectrosc* **54:** 127–132.

Jeong, J.-J., Fushinobu, S., Ito, S., Jeon, B.-S., Shoun, H., and Wakagi, T. (2003) Characterization of the cupin-type phosphoglucose isomerase from the hyperthermophilic archaeon *Thermococcus litoralis*. *FEBS Lett* **535:** 200–204.

Jung, M.-Y., Park, S.-J., Min, D., Kim, J.-S., Rijpstra, W.I.C., Sinninghe Damsté, J.S., *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.

Kelman, L.M., and Kelman, Z. (2004) Multiple origins of replication in archaea. *Trends Microbiol* **12:** 399–401.

Kerscher, L., and Oesterhelt, D. (1981) Purification and properties of two 2-oxoacid:ferredoxin oxidoreductases from *Halobacterium halobium*. *Eur J Biochem* **116:** 587–594.

Kim, B.K., Jung, M.-Y., Yu, D.S., Park, S.-J., Oh, T.K., Rhee, S.-K., and Kim, J.F. (2011) Genome sequence of an ammonia-oxidizing soil archaeon, '*Candidatus* Nitrosoarchaeum koreensis' MY1. *J Bacteriol* **193:** 5539–5540.

Kim, J.-G., Jung, M.-Y., Park, S.-J., Rijpstra, W.I.C., Sinninghe Damsté, J.S., Madsen, E.L., *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.

Klotz, M.G., and Stein, L.Y. (2008) Nitrifier genomics and evolution of the nitrogen cycle. *FEMS Microbiol Lett* **278:** 146–156.

Klunker, D., Haas, B., Hirtreiter, A., Figueiredo, L., Naylor, D.J., Pfeifer, G., *et al*. (2003) Coexistence of group I and group II chaperonins in the archaeon *Methanosarcina mazei*. *J Biol Chem* **278:** 33256–33267.

Könneke, M., Bernhard, A.E., Torre, J.R., Walker, C.B., Waterbury, J.B., and Stahl, D.A. (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437:** 543–546.

Koshiishi, I., Mamura, Y., and Imanari, T. (1997) Cyanate causes depletion of ascorbate in organisms. *Biochim Biophys Acta* **1336:** 566–574.

Kraus, L.M., and Kraus, A.P., Jr (1998) The search for the uremic toxin: the case for carbamoylation of amino acids and proteins. *Wien Klin Wochenschr* **110:** 521–530.

Kristensen, D.M., Kannan, L., Coleman, M.K., Wolf, Y.I., Sorokin, A., Koonin, E.V., and Mushegian, A. (2010) A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. *Bioinformatics* **26:** 1481–1487.

Krupovic, M., Spang, A., Gribaldo, S., Forterre, P., and Schleper, C. (2011) A thaumarchaeal provirus testifies for an ancient association of tailed viruses with archaea. *Biochem Soc Trans* **39:** 82–88.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004) Versatile and open software for comparing large genomes. *Genome Biol* **5:** R12.

Large, A.T., Goldberg, M.D., and Lund, P.A. (2009) Chaperones and protein folding in the archaea. *Biochem Soc Trans* **37:** 46–51.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., *et al*. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947–2948.

Lebedeva, E.V., Alawi, M., Fiencke, C., Namsaraev, B., Bock, E., and Spieck, E. (2005) Moderately thermophilic nitrifying bacteria from a hot spring of the Baikal rift zone. *FEMS Microbiol Ecol* **54:** 297–306.

Lee, S.M., Grass, G., Rensing, C., Barrett, S.R., Yates, C.J.D., Stoyanov, J.V., and Brown, N.L. (2002) The Pco proteins are involved in periplasmic copper handling in *Escherichia coli*. *Biochem Biophys Res Commun* **295:** 616–620.

Lehtovirta-Morley, L.E., Stoecker, K., Vilcinskas, A., Prosser, J.I., and Nicol, G.W. (2011) Cultivation of an obligate acidophilic ammonia oxidizer from a nitrifying acid soil. *Proc Natl Acad Sci USA* **108:** 15892–15897.

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G.W., *et al*. (2006) Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442:** 806–809.

Levin, E.J., Quick, M., and Zhou, M. (2009) Crystal structure of a bacterial homologue of the kidney urea transporter. *Nature* **462:** 757–761.

Li, N., and Cannon, M.C. (1998) Gas vesicle genes identified in *Bacillus megaterium* and functional expression in *Escherichia coli*. *J Bacteriol* **180:** 2450–2458.

Lima, T., Auchincloss, A.H., Coudert, E., Keller, G., Michoud, K., Rivoire, C., *et al*. (2009) HAMAP: a database of completely sequenced microbial proteome sets and manually curated microbial protein families in UniProtKB/Swiss-Prot. *Nucleic Acids Res* **37:** D471–D478.

Lin, X.L., and White, R.H. (1986) Occurrence of coenzyme F420 and its gamma-monoglutamyl derivative in non-methanogenic archaebacteria. *J Bacteriol* **168:** 444–448.

Liu, J., Xue, Y., Wang, Q., Wei, Y., Swartz, T.H., Hicks, D.B., *et al*. (2005) The activity profile of the NhaD-type $Na^+(Li^+)/H^+$ antiporter from the soda Lake Haloalkaliphile *Alkalimonas amylolytica* is adaptive for the extreme environment. *J Bacteriol* **187:** 7589–7595.

Loqué, D., Lalonde, S., Looger, L.L., von Wirén, N., and Frommer, W.B. (2007) A cytosolic trans-activation domain essential for ammonium uptake. *Nature* **446:** 195–198.

Lorenz, M.C., and Heitman, J. (1998) The MEP2 ammonium permease regulates pseudohyphal differentiation in *Saccharomyces cerevisiae*. *EMBO J* **17:** 1236–1247.

Lu, Q., Han, J., Zhou, L., Zhou, J., and Xiang, H. (2008) Genetic and biochemical characterization of the poly(3-hydroxybutyrate-co-3-hydroxyvalerate) synthase in *Haloferax mediterranei*. *J Bacteriol* **190:** 4173–4180.

Lücker, S., Wagner, M., Maixner, F., Pelletier, E., Koch, H., *et al*. (2010) A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proc Natl Acad Sci USA* **107:** 13479–13484.

Macario, A.J.L., Malz, M., and de Macario, E. (2004) Evolution of assisted protein folding: the distribution of the main chaperoning systems within the phylogenetic domain archaea. *Front Biosci* **9:** 1318–1332.

Magrane, M., and UniProt Consortium (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database* bar009.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., *et al*. (2011) Evolution and

classification of the CRISPR-Cas systems. *Nat Rev Microbiol* **9:** 467–477.

Mansour, N.M., Sawhney, M., Tamang, D.G., Vogl, C., and Saier, M.H., Jr (2007) The bile/arsenite/riboflavin transporter (BART) superfamily. *FEBS J* **274:** 612–629.

Martens-Habbena, W., Berube, P.M., Urakawa, H., De La Torre, J.R., and Stahl, D.A. (2009) Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature* **461:** 976–979.

Mechold, U., Ogryzko, V., Ngo, S., and Danchin, A. (2006) Oligoribonuclease is a common downstream target of lithium-induced pAp accumulation in *Escherichia coli* and human cells. *Nucleic Acids Res* **34:** 2364–2373.

Melo, A.M.P., Bandeiras, T.M., and Teixeira, M. (2004) New insights into type II NAD(P)H:quinone oxidoreductases. *Microbiol Mol Biol Rev* **68:** 603–616.

Moparthi, V.K., and Hägerhäll, C. (2011) The evolution of respiratory chain complex I from a smaller last common ancestor consisting of 11 protein subunits. *J Mol Evol* **72:** 484–497.

Mosier, A.C., Allen, E.E., Kim, M., Ferriera, S., and Francis, C.A. (2012a) Genome sequence of '*Candidatus* Nitrosopumilus salaria' BD31, an ammonia-oxidizing archaeon from the San Francisco Bay Estuary. *J Bacteriol* **194:** 2121–2122.

Mosier, A.C., Lund, M.B., and Francis, C.A. (2012b) Ecophysiology of an ammonia-oxidizing archaeon adapted to low-salinity habitats. *Microb Ecol*. doi:10.007/s00248-012-0075-1.

Mosier, A.C., Allen, E.E., Kim, M., Ferriera, S., and Francis, C.A. (2012c) Genome sequence of '*Candidatus* Nitrosoarchaeum limnia' BG20, a low-salinity ammonia-oxidizing archaeon from the San Francisco Bay Estuary. *J Bacteriol* **194:** 2119–2120.

Murakami, R., Sato, H., Dybal, J., Iwata, T., and Ozaki, Y. (2007) Formation and stability of β-structure in biodegradable ultra-high-molecular-weight poly(3-hydroxybutyrate) by infrared, Raman, and quantum chemical calculation studies. *Polymer* **48:** 2672–2680.

Musa-Aziz, R., Chen, L.-M., Pelletier, M.F., and Boron, W.F. (2009) Relative CO_2/NH_3 selectivities of AQP1, AQP4, AQP5, AmtB, and RhAG. *Proc Natl Acad Sci USA* **106:** 5406–5411.

Mußmann, M., Brito, I., Pitcher, A., Sinninghe Damsté, J.S., Hatzenpichler, R., Richter, A., *et al*. (2011) Thaumarchaeotes abundant in refinery nitrifying sludges express *amoA* but are not obligate autotrophic ammonia oxidizers. *Proc Natl Acad Sci USA* **108:** 16771–16776.

Neves, C., da Costa, M.S., and Santos, H. (2005) Compatible solutes of the hyperthermophile *Palaeococcus ferrophilus*: osmoadaptation and thermoadaptation in the order thermococcales. *Appl Environ Microbiol* **71:** 8091–8098.

Norais, C., Hawkins, M., Hartman, A.L., Eisen, J.A., Myllykallio, H., and Allers, T. (2007) Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS Genet* **3:** e77.

Norton, J.M., Alzerreca, J.J., Suwa, Y., and Klotz, M.G. (2002) Diversity of ammonia monooxygenase operon in autotrophic ammonia-oxidizing bacteria. *Arch Microbiol* **177:** 139–149.

Nowakowska, M., Sterzel, M., and Szczubiałka, K. (2006) Photosensitized oxidation of cyanide in aqueous solutions of photoactive modified hydroxyethylcellulose. *J Polym Environ* **14:** 59–64.

Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., *et al*. (2011) Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res* **39:** 3204–3223.

Offner, S., Hofacker, A., Wanner, G., and Pfeifer, F. (2000) Eight of fourteen *gvp* genes are sufficient for formation of gas vesicles in halophilic archaea. *J Bacteriol* **182:** 4328–4336.

Overbeek, R., Larsen, N., Walunas, T., D'Souza, M., Pusch, G., Selkov, E., Jr, *et al*. (2003) The ERGO genome analysis and discovery system. *Nucleic Acids Res* **31:** 164–171.

Park, Y.-J., Yoo, C.-B., Choi, S.Y., and Lee, H.-B. (2006) Purifications and characterizations of a ferredoxin and its related 2-oxoacid:ferredoxin oxidoreductase from the hyperthermophilic archaeon, *Sulfolobus solfataricus* P1. *J Biochem Mol Biol* **39:** 46–54.

Patridge, E.V., and Ferry, J.G. (2006) WrbA from *Escherichia coli* and *Archaeoglobus fulgidus* is an NAD(P)H:quinone oxidoreductase. *J Bacteriol* **188:** 3498–3506.

Paulsen, I.T., Nguyen, L., Sliwinski, M.K., Rabus, R., and Saier, M.H., Jr (2000) Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. *J Mol Biol* **301:** 75–100.

Peiser, G.D., Wang, T.-T., Hoffman, N.E., Yang, S.F., Liu, H., and Walsh, C.T. (1984) Formation of cyanide from carbon 1 of 1-aminocyclopropane-1-carboxylic acid during its conversion to ethylene. *Proc Natl Acad Sci USA* **81:** 3059–3063.

Pereira, M.M., Bandeiras, T.M., Fernandes, A.S., Lemos, R.S., Melo, A.M., and Teixeira, M. (2004) Respiratory chains from aerobic thermophilic prokaryotes. *J Bioenerg Biomembr* **36:** 93–105.

Perkins, J., and Gadd, G.M. (1995) The influence of pH and external K^+ concentration on caesium toxicity and accumulation in *Escherichia coli* and *Bacillus subtilis*. *J Ind Microbiol* **14:** 218–225.

Perrière, G., and Gouy, M. (1996) WWW-query: an on-line retrieval system for biological sequence banks. *Biochimie* **78:** 364–369.

Pester, M., Schleper, C., and Wagner, M. (2011) The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Curr Opin Microbiol* **14:** 300–306.

Pester, M., Rattei, T., Flechl, S., Gröngröft, A., Richter, A., Overmann, J., *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.

Pfeifer, F. (2006) Gas vesicles of archaea and bacteria. In *Complex Intracellular Structures in Prokaryotes*. Shively, J.M. (ed.). Heidelberg, Germany: Springer-Verlag, pp. 115–140.

Pitcher, A., Rychlik, N., Hopmans, E.C., Spieck, E., Rijpstra, W.I.C., Ossebaar, J., *et al*. (2010) Crenarchaeol dominates the membrane lipids of *Candidatus* Nitrososphaera gargensis, a thermophilic group I.1b Archaeon. *ISME J* **4:** 542–552.

Pohlschroder, M., Ghosh, A., Tripepi, M., and Albers, S.-V. (2011) Archaeal type IV pilus-like structures-evolutionarily conserved prokaryotic surface organelles. *Curr Opin Microbiol* **14:** 357–363.

Poli, A., Di Donato, P., Abbamondi, G.R., and Nicolaus, B. (2011) Synthesis, production, and biotechnological applications of exopolysaccharides and polyhydroxyalkanoates by archaea. *Archaea* **2011:** 693253.

Prosser, J.I. (1989) Autotrophic nitrification in bacteria. *Adv Microb Physiol* **30:** 125–181.

Pruitt, K.D., Tatusova, T., Brown, G.R., and Maglott, D.R. (2012) NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res* **40:** D130–D135.

Purwantini, E., and Daniels, L. (1996) Purification of a novel coenzyme F420-dependent glucose-6-phosphate dehydrogenase from *Mycobacterium smegmatis*. *J Bacteriol* **178:** 2861–2866.

Purwantini, E., and Mukhopadhyay, B. (2009) Conversion of NO_2 to NO by reduced coenzyme F420 protects mycobacteria from nitrosative damage. *Proc Natl Acad Sci USA* **106:** 6333–6338.

Purwantini, E., Gillis, T.P., and Daniels, L. (1997) Presence of F420-dependent glucose-6-phosphate dehydrogenase in *Mycobacterium* and *Nocardia* species, but absence from *Streptomyces* and *Corynebacterium* species and methanogenic Archaea. *FEMS Microbiol Lett* **146:** 129–134.

Quaiser, A., Ochsenreiter, T., Klenk, H., Kletzin, A., Treusch, A.H., Meurer, G., *et al*. (2002) First insight into the genome of an uncultivated crenarchaeote from soil. *Environ Microbiol* **4:** 603–611.

Radax, R., Rattei, T., Lanzen, A., Bayer, C., Rapp, H.T., Urich, T., and Schleper, C. (2012) Metatranscriptomics of the marine sponge *Geodia barretti*: tackling phylogeny and function of its microbial community. *Environ Microbiol* **14:** 1308–1324.

Ramírez-Díaz, M.I., Díaz-Pérez, C., Vargas, E., Riveros-Rosas, H., Campos-García, J., and Cervantes, C. (2008) Mechanisms of bacterial resistance to chromium compounds. *Biometals* **21:** 321–332.

Rattei, T., Tischler, P., Götz, S., Jehl, M.-A., Hoser, J., Arnold, R., *et al*. (2010) SIMAP – a comprehensive database of pre-calculated protein sequence similarities, domains, annotations and clusters. *Nucleic Acids Res* **38:** D223–D226.

Rehm, B.H., and Steinbüchel, A. (1999) Biochemical and genetic analysis of PHA synthases and other proteins required for PHA synthesis. *Int J Biol Macromol* **25:** 3–19.

Ren, Q., Kang, K.H., and Paulsen, I.T. (2004) TransportDB: a relational database of cellular membrane transport systems. *Nucleic Acids Res* **32:** D284–D288.

Ren, Q., Chen, K., and Paulsen, I.T. (2007) TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res* **35:** D274–D279.

Robinson, N.P., and Bell, S.D. (2007) Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proc Natl Acad Sci USA* **104:** 5806–5811.

Robinson, N.P., Dionne, I., Lundgren, M., Marsh, V.L., Bernander, R., and Bell, S.D. (2004) Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell* **116:** 25–38.

Saier, M.H., Jr, Tran, C.V., and Barabote, R.D. (2006) TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res* **34:** D181–D186.

Saier, M.H., Jr, Yen, M.R., Noto, K., Tamang, D.G., and Elkan, C. (2009) The Transporter Classification Database: recent advances. *Nucleic Acids Res* **37:** D274–D278.

Sambrook, J., and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn. New York, NY, USA: Cold Spring Harbor Laboratory Press.

Say, R.F., and Fuchs, G. (2010) Fructose 1,6-bisphosphate aldolase/phosphatase may be an ancestral gluconeogenic enzyme. *Nature* **464:** 1077–1081.

Schäfer, G., Purschke, W., and Schmidt, C.L. (1996) On the origin of respiration: electron transport proteins from archaea to man. *FEMS Microbiol Rev* **18:** 173–188.

Schäfer, G., Engelhard, M., and Müller, V. (1999) Bioenergetics of the archaea. *Microbiol Mol Biol Rev* **63:** 570–620.

Schiex, T., Gouzy, J., Moisan, A., and de Oliveira, Y. (2003) FrameD: a flexible program for quality check and gene prediction in prokaryotic genomes and noisy matured eukaryotic sequences. *Nucleic Acids Res* **31:** 3738–3741.

Schleper, C., and Nicol, G.W. (2010) Ammonia-oxidising archaea – physiology, ecology and evolution. *Adv Microb Physiol* **57:** 1–41.

Schleper, C., Jurgens, G., and Jonuscheit, M. (2005) Genomic studies of uncultivated archaea. *Nat Rev Microbiol* **3:** 479–488.

Schmeisser, C., Liesegang, H., Krysciak, D., Bakkou, N., Le Quéré, A., Wollherr, A., *et al*. (2009) *Rhizobium* sp. strain NGR234 possesses a remarkable number of secretion systems. *Appl Environ Microbiol* **75:** 4035–4045.

Schröder, I., Johnson, E., and de Vries, S. (2003) Microbial ferric iron reductases. *FEMS Microbiol Rev* **27:** 427–447.

Selengut, J.D., and Haft, D.H. (2010) Unexpected abundance of coenzyme F(420)-dependent enzymes in *Mycobacterium tuberculosis* and other actinobacteria. *J Bacteriol* **192:** 5788–5798.

Shigaki, T., Barkla, B.J., Miranda-Vergara, M.C., Zhao, J., Pantoja, O., and Hirschi, K.D. (2005) Identification of a crucial histidine involved in metal transport activity in the *Arabidopsis* cation/H^+ exchanger CAX1. *J Biol Chem* **280:** 30136–30142.

Siebers, B., and Schönheit, P. (2005) Unusual pathways and enzymes of central carbohydrate metabolism in Archaea. *Curr Opin Microbiol* **8:** 695–705.

Sigrist, C.J.A., Cerutti, L., de Castro, E., Langendijk-Genevaux, P.S., Bulliard, V., Bairoch, A., and Hulo, N. (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res* **38:** D161–D166.

Silver, S., and Ji, G. (1994) Newer systems for bacterial resistances to toxic heavy metals. *Environ Health Perspect* **102:** 107–113.

Soderberg, T. (2005) Biosynthesis of ribose-5-phosphate and erythrose-4-phosphate in archaea: a phylogenetic analysis of archaeal genomes. *Archaea* **1:** 347–352.

Soderberg, T., and Alver, R.C. (2004) Transaldolase of *Methanocaldococcus jannaschii*. *Archaea* **1:** 255–262.

Sorek, R., Kunin, V., and Hugenholtz, P. (2008) CRISPR – a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* **6:** 181–186.

Spang, A., Hatzenpichler, R., Brochier-Armanet, C., Rattei, T., Tischler, P., Spieck, E., *et al*. (2010) Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. *Trends Microbiol* **18:** 331–340.

Spieck, E., and Lipski, A. (2011) Cultivation, growth physiology and chemotaxonomy of nitrite-oxidizing bacteria. In *Methods of Enzymology, Vol. 486, Part A, Microbial Nitrification and Related Processes*. Klotz, M.G. (ed.). Oxford, UK: Academic Press/Elsevier, pp. 109–130.

Stewart, F.J., Ulloa, O., and DeLong, E.F. (2012) Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environ Microbiol* **14:** 23–40.

Szabó, Z., Stahl, A.O., Albers, S.V., Kissinger, J.C., Driessen, A.J.M., and Pohlschröder, M. (2007a) Identification of diverse archaeal proteins with class III signal peptides cleaved by distinct archaeal prepilin peptidases. *J Bacteriol* **189:** 772–778.

Szabó, Z., Sani, M., Groeneveld, M., Zolghadr, B., Schelert, J., Albers, S.V., *et al*. (2007b) Flagellar motility and structure in the hyperthermoacidophilic archaeon *Sulfolobus solfataricus*. *J Bacteriol* **189:** 4305–4309.

Tatusov, R.L., Koonin, E.V., and Lipman, D.J. (1997) A genomic perspective on protein families. *Science* **278:** 631–637.

Tech, M., and Merkl, R. (2003) YACOP: enhanced gene prediction obtained by a combination of existing methods. *In Silico Biol* **3:** 441–451.

Thomas, N.A., Bardy, S.L., and Jarrell, K.F. (2001) The archaeal flagellum: a different kind of prokaryotic motility structure. *FEMS Microbiol Rev* **25:** 147–174.

de la Torre, J.R., Walker, C.B., Ingalls, A.E., Könneke, M., and Stahl, D.A. (2008) Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. *Environ Microbiol* **10:** 810–818.

Tourna, M., Stieglmeier, M., Spang, A., Könneke, M., Schintlmeister, A., Urich, T., *et al*. (2011) *Nitrososphaera viennensis*, an ammonia oxidizing archaeon from soil. *Proc Natl Acad Sci USA* **108:** 8420–8425.

Treusch, A.H., Leininger, S., Kletzin, A., Schuster, S.C., Klenk, H., and Schleper, C. (2005) Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. *Environ Microbiol* **7:** 1985–1995.

Türk, K., Puhar, A., Neese, F., Bill, E., Fritz, G., and Steuber, J. (2004) NADH oxidation by the Na^+-translocating NADH-:quinone oxidoreductase from *Vibrio cholerae*: functional role of the NqrF subunit. *J Biol Chem* **279:** 21349–21355.

Turkarslan, S., Reiss, D.J., Gibbins, G., Su, W.L., Pan, M., Bare, J.C., *et al*. (2011) Niche adaptation by expansion and reprogramming of general transcription factors. *Mol Syst Biol* **7:** 554.

Urakawa, H., Martens-Habbena, W., and Stahl, D.A. (2011) Physiology and genomics of ammonia-oxidizing archaea. In *Nitrification*. Ward, B.B., Arp, D.J., and Klotz, M.G. (eds). Washington, DC, USA: American Society for Microbiology, pp. 117–156.

Vallenet, D., Labarre, L., Rouy, Z., Barbe, V., Bocs, S., Cruveiller, S., *et al*. (2006) MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res* **34:** 53–65.

Vallenet, D., Engelen, S., Mornico, D., Cruveiller, S., Fleury, L., Lajus, A., *et al*. (2009) MicroScope: a platform for microbial genome annotation and comparative genomics. *Database* bap021.

Verhamme, D.T., Prosser, J.I., and Nicol, G.W. (2011) Ammonia concentration determines differential growth of ammonia-oxidising archaea and bacteria in soil microcosms. *ISME J* **5:** 1067–1071.

Verhees, C.H., Huynen, M.A., Ward, D.E., Schiltz, E., de Vos, W.M., and van der Oost, J. (2001) The phosphoglucose isomerase from the hyperthermophilic archaeon *Pyrococcus furiosus* is a unique glycolytic enzyme that belongs to the cupin superfamily. *J Biol Chem* **276:** 40926–40932.

Verkhovsky, M.I., and Bogachev, A.V. (2010) Sodium-translocating NADH:quinone oxidoreductase as a redox-driven ion pump. *Biochim Biophys Acta* **1797:** 738–746.

Walker, C.B., Torre, J.R., Klotz, M.G., Urakawa, H., Pinel, N., Arp, D.J., *et al*. (2010) *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc Natl Acad Sci USA* **107:** 8818–8823.

Walsh, M.A., Otwinowski, Z., Perrakis, A., Anderson, P.M., and Joachimiak, A. (2000) Structure of cyanase reveals that a novel dimeric and decameric arrangement of subunits is required for formation of the enzyme active site. *Structure* **8:** 505–514.

van de Werken, H.J.G., Brouns, S.J.J., and van der Oost, J. (2008) Pentose metabolism in archaea. In *Archaea: New Models for Prokaryotic Biology*. Blum, P. (ed.). Norfolk, UK: Caister Academic Press, pp. 71–94.

Wood, A.P., Aurikko, J.P., and Kelly, D.P. (2004) A challenge for 21st century molecular biology and biochemistry: what are the causes of obligate autotrophy and methanotrophy? *FEMS Microbiol Rev* **28:** 335–352.

Yakunin, A.F., and Hallenbeck, P.C. (2002) AmtB is necessary for NH_4^+-induced nitrogenase switch-off and ADP-ribosylation in *Rhodobacter capsulatus*. *J Bacteriol* **184:** 4081–4088.

Zhang, J.-Y., Zou, J., Bao, Q., Chen, W.-L., Wang, L., Yang, H., and Zhang, C.-C. (2006) A lithium-sensitive and sodium-tolerant 3′-phosphoadenosine-5′-phosphatase encoded by *halA* from the cyanobacterium *Arthrospira platensis* is closely related to its counterparts from yeasts and plants. *Appl Environ Microbiol* **72:** 245–251.

Zhang, L., Zhang, L., Liu, Y., Yang, S., Gao, C., Gong, H., *et al*. (2009) Archaeal eukaryote-like Orc1/Cdc6 initiators physically interact with DNA polymerase B1 and regulate its functions. *Proc Natl Acad Sci USA* **106:** 7792–7797.

Zhang, Q., Iwasaki, T., Wakagi, T., and Oshima, T. (1996) 2-Oxoacid:ferredoxin oxidoreductase from the thermoacidophilic archaeon, *Sulfolobus* sp. strain 7. *J Biochem* **120:** 587–599.

Zheng, L., Kostrewa, D., Bernèche, S., Winkler, F.K., and Li, X.-D. (2004) The mechanism of ammonia transport based on the crystal structure of AmtB of *Escherichia coli*. *Proc Natl Acad Sci USA* **101:** 17090–17095.

Zolghadr, B., Weber, S., Szabó, Z., Driessen, A.J.M., and Albers, S.-V. (2007) Identification of a system required for the functional surface localization of sugar binding proteins with class III signal peptides in *Sulfolobus solfataricus*. *Mol Microbiol* **64:** 795–806.

Zolghadr, B., Klingl, A., Rachel, R., Driessen, A.J.M., and Albers, S.-V. (2011) The bindosome is a structural component of the *Sulfolobus solfataricus* cell envelope. *Extremophiles* **15:** 235–244.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Additional interesting features of the *N. gargensis* genome.

Fig. S1. A. Phylogeny of thaumarchaeota including *N. gargensis*. Majority consensus trees based on the 16S rRNA gene (1067 nucleic acid positions conserved in > 50% of all archaea) as inferred by maximum likelihood, distance and maximum parsimony methods. The bars represent 10% Jukes–Cantor corrected sequence divergence. Dots indicate bootstrap support above 80% as inferred by maximum likelihood. Abbreviations: SAGMCG-1, South African gold mine crenarchaeotic group 1; HWCG III, hot water crenarchaeotic group III. Names of described thaumarchaeota within the clades are listed. *Genome sequence not yet available/published. Tree modified from Pester and colleagues (2011). B and C. Electron micrographs of *N. gargensis* showing its morphology and ultrastructure. (B) Single cells are arranged to small aggregates and reveal a regular structure of the surface layer. Cells are stained with uranylacetate. (C) Ultrathin section showing the thick cell wall and electron dense structures in the cytoplasm.

Fig. S2. Phylogenetic maximum likelihood tree of bacterial 16S rRNA sequences from the enrichment culture Ga 9-2. Bootstrap values are given in % from 1000 replicate resamplings; only bootstrap values higher than 80% are shown. 16S rRNA gene clones from the enrichment are labelled in red and bold. If the respective 16S rRNA gene sequence was also found in the 454 metagenome data set this is also indicated. The scale represents base substitutions per nucleotide. Specific oligonucleotide probes were designed for the *Thermaerobacter*-related (probe GaBI830, GGT CAA ACC CAC CCA CAC) and *Betaproteobacteria*-affiliated (RHG1130, AGT GCC CAC CTC TCG CGT) 16S rRNA sequences. Both probes were evaluated via Clone FISH (Schramm *et al*., 2002) and used at 35% formamide in the hybridization buffer. The probes were applied together with the general bacterial probe mix EUB338mix (Daims *et al*., 1999) using established protocols (Daims *et al*., 2005). The two newly designed probes stained all bacterial cells detected by the EUB338mix (data not shown) suggesting that the other contaminants occurred at an abundance below the detection limit of FISH or were not amenable to the applied FISH protocol, e.g. due to low ribosomal numbers.

Fig. S3. Venn diagrams showing the numbers of COGs shared between the predicted proteomes of *N. gargensis*, *N. maritimus*, *N. limnia*, *N. koreensis* and *C. symbiosum* (A) as well as between the predicted proteomes of *N. gargensis*, *N. maritimus*, the euryarcheon *Halobacterium salinarum*, the crenarchaeon *Metallosphaera sedula* and *Caldiarchaeum subterraneum* (B) (see *Experimental procedures* for details of COG determination). Proteins that were not grouped into COGs are represented as specific proteins for each organism. Numbers in brackets behind species names indicate total number of predicted proteins.

Fig. S4. Predicted carbon metabolism of *N. gargensis*. Enzymatic reactions for which candidate genes could be identified in the genome are highlighted by solid arrows. Depending on homology conservation of the predicted proteins to described proteins in other organisms, the arrows and EC numbers are marked in either black [> 35% AA similarity to characterized proteins (class 2)] or grey [putative proteins with respective function (class 3)]. White dashed arrows and EC numbers indicate reactions for which no putative candidate genes were found in the *N. gargensis* genome, based on homology searches to known enzymes involved in the respective steps. Proteins labelled with a red asterisk are absent from other published thaumarchaeal genomes. *Related PHA synthases (type III) of haloarchaeota synthesize polyhydroxybutyrate and/or 3-polyhydroxybutyrate-co-3-hydroxyvalerate. Annotations of all corresponding genes are listed in Table S1.

Fig. S5. Maximum likelihood phylogenetic analysis of one of the predicted 6-phosphogluconate dehydrogenase enzymes of *N. gargensis* compared with related enzymes of archaea and bacteria based on an alignment of 293 amino acid positions (50% insertion/deletion filter and 100 bootstraps). The fragmented paralogous 6-phosphogluconate dehydrogenase of *N. gargensis* forms a monophyletic cluster together with the former enzyme (not shown). The scale bar refers to 20% estimated sequence divergence. *Protein has been characterized and shown to participate in oxidative pentose phosphate pathway (Chistoserdova *et al*., 2000). **Function of this protein in *Bacillus subtilis* is obscure as this organism contains a canonical 6-phosphogluconate dehydrogenase (Zamboni *et al*., 2004).

Fig. S6. Amino acid biosynthetic pathways of *Nitrososphaera gargensis*. Only selected catabolic pathways are shown. tRNA genes, aminopeptidases and some selected transporters are displayed, too. *Nitrososphaera gargensis*-specific enzymes refer to enzymes not found in the genomes of other thaumarchaeotes publically available until October 2010. For more information, see respective section in *Supporting information*.

Fig. S7. Maximum likelihood phylogenetic analysis of almost all (closely related proteins were excluded) mannosyl-3-phosphoglycerate synthases based on an alignment of 384 amino acid positions (removal of ambiguously aligned sites, 100 bootstraps). The scale bar refers to 20% estimated sequence divergence. Archaea are labelled in purple, eukaryotes are labelled in dark green and bacteria are shown in black.

Fig. S8. Phylogeny of the urea transporter from the UT family of *N. gargensis* using selected reference sequences. Maximum likelihood phylogenetic analysis based on 266 aligned amino acid positions (alignment based on Clustal2w with removal of ambiguously aligned regions in this poorly

conserved protein, 50% insertion/deletion filter and 100 bootstraps). The scale bar refers to 20% estimated sequence divergence. Archaea are labelled in purple. Eukaryotes are shown in green.

Fig. S9. Phylogeny of the BART superfamily including the two paralogues of *N. gargensis* and selected reference sequences. Maximum likelihood analysis based on 289 amino acid positions (alignment based on Clustal2w with removal of ambiguously aligned regions in this poorly conserved protein, 50% insertion/deletion filter and 100 bootstraps). Only bootstrap support > 30% is indicated. The scale bar refers to 50% estimated sequence divergence. Archaea are labelled in purple. Eukaryotes are shown in green. The tree is rooted with the UBS1 family (unknown BART superfamily-1 family), the third member of the BART superfamily. ACR3, arsenite resistance-3; BASS, bile acid sodium symporter. Transported substrates: [1]arsenite, [2]arsenite/antimonite efflux, [3]taurocholate/Na^+ symport, [4]bile acid/Na^+ symport, [5]estrone-3-sulfate/dehydropiandrosterone sulfate/Na^+ symport, [6]dehydropiandrosterone sulfate/estrone-3-sulfate/prenenolone sulfate/sulfoconjugated taurolithocholate/Na^+ symport, [7]glucosinolate, [8]pyruvate/Na^+ symport.

Table S1. Genes encoding proteins of diverse metabolic pathways.

Table S2. Genes encoding transporter proteins.

Table S3. Selected genes encoding informational processing proteins.

Table S4. IS elements and transposase genes.

Table S5. Integrases and putative associated genomic islands.

Table S6. Concentration of selected chemical elements in the Garga spring. For comparison concentrations of these elements in pristine marine (Turekian, 1968) and freshwater samples (see references in the table) are listed.