

RESEARCH ARTICLE

Open Access



Genome-resolved metagenomics analysis provides insights into the ecological role of Thaumarchaeota in the Amazon River and its plume

Otávio H. B. Pinto¹, Thais F. Silva¹, Carla S. Vizzotto^{1,2}, Renata H. Santana³, Fabyano A. C. Lopes⁴, Bruno S. Silva⁵, Fabiano L. Thompson⁵ and Ricardo H. Kruger^{1*} 

Abstract

Background: Thaumarchaeota are abundant in the Amazon River, where they are the only ammonia-oxidizing archaea. Despite the importance of Thaumarchaeota, little is known about their physiology, mainly because few isolates are available for study. Therefore, information about Thaumarchaeota was obtained primarily from genomic studies. The aim of this study was to investigate the ecological roles of Thaumarchaeota in the Amazon River and the Amazon River plume.

Results: The archaeal community of the shallow in Amazon River and its plume is dominated by Thaumarchaeota lineages from group 1.1a, which are mainly affiliated to *Candidatus Nitrosotenuis uzonensis*, members of order Nitrosopumilales, *Candidatus Nitrosoarchaeum*, and *Candidatus Nitrosopelagicus* sp. While Thaumarchaeota sequences have decreased their relative abundance in the plume, *Candidatus Nitrosopelagicus* has increased. One genome was recovered from metagenomic data of the Amazon River (ThauR71 [1.05 Mpb]), and two from metagenomic data of the Amazon River plume (ThauP25 [0.94 Mpb] and ThauP41 [1.26 Mpb]). Phylogenetic analysis placed all three Amazon genome bins in Thaumarchaeota Group 1.1a. The annotation revealed that most genes are assigned to the COG subcategory coenzyme transport and metabolism. All three genomes contain genes involved in the hydroxypropionate/hydroxybutyrate cycle, glycolysis, tricarboxylic acid cycle, oxidative phosphorylation. However, ammonia-monooxygenase genes were detected only in ThauP41 and ThauR71. Glycoside hydrolases and auxiliary activities genes were detected only in ThauP25.

Conclusions: Our data indicate that Amazon River is a source of Thaumarchaeota, where these organisms are important for primary production, vitamin production, and nitrification.

Keywords: Thaumarchaeota, Amazon River, Amazon River plume, Metagenome-assembled genome

Background

Thaumarchaeota was proposed as an archaeal phylum in 2008 [1]. These organisms were previously known as “mesophilic Crenarchaeota” based on phylogenetic analyses of the large-subunit and small-subunit rRNA gene, which weakly suggested that they form a sister group with Crenarchaeota [1–3]. However, analysis of ribosomal

proteins and comparisons of specific proteins showed that Thaumarchaeota represents a phylum more closely related to Euryarchaeota than to Crenarchaeota [1, 4–6]. In addition, genomic analysis showed that Thaumarchaeota possesses genomic features that are not present in either Euryarchaeota or Crenarchaeota [1, 7].

Thaumarchaeota has several lineages that are not yet well defined; the most commonly used nomenclature for these lineages is Groups 1.1a, 1.1a-associated, 1.1b, ThAOA, 1.1c, and 1.3 [8–11]. Groups 1.1a, 1.1a-associated, ThAOA, and 1.1b, which comprise the class

* Correspondence: kruger@unb.br

¹Department of Enzymology, Institute of Biological Sciences, University of Brasília, Brasília 70910-900, Brazil

Full list of author information is available at the end of the article



Nitrososphaeria, are the only ammonia-oxidizing archaea (AOA) [9, 12, 13]. The gene encoding ammonia monooxygenase subunit A (*amoA*) is used as a marker to study the diversity and abundance of AOA in various habitats [13–25]. However, the lack of cultured representatives of AOA has impeded our study of their physiology. Currently, just ten members of group 1.1a [26–28], three of 1.1b [28, 29], and two from the order Nitrosocaldales [28] could be cultured as pure isolates. Other members have been studied using enrichment cultures [9, 13]; however, Groups 1.1c and 1.3 have no cultured representatives [11, 30]. Thus, most of our information about Thaumarchaeota originates from genomic studies.

Metagenomics studies indicate that Thaumarchaeota is one of the most abundant archaea on the planet [2, 31]. These ubiquitous organisms [9, 14, 32] are present in the euphotic zone and suboxic, eutrophic, oligotrophic, thermophilic, alkaline, and extremely acidic environments [11, 30, 33–38]. They play a major role in nitrification [16, 21, 39, 40] and are crucial in the open ocean, where ammonia concentrations are very low, and ammonia-oxidizing bacteria show low activity [41]. Thaumarchaeota are typically chemolithoautotrophic and fix CO₂ using a variant of the hydroxypropionate/hydroxybutyrate (HP/HB) cycle, which is the most energy-efficient aerobic CO₂ fixation pathway [30, 42, 43]. However, despite the importance of Thaumarchaeota, their role in several habitats remains underexplored.

Previous studies have confirmed the presence of Thaumarchaeota in the Amazon River [44, 45], which is considered the world's largest riverine system, accounting for 20% of the global freshwater discharge into the Atlantic Ocean [46, 47]. The Amazon River and its plume carry a large amount of suspended and dissolved terrestrial materials, such as organic matter and nutrients [48, 49]. The plume can travel many hundreds of kilometers and cover ~ 2 million km², thus contributing to nutrient cycling at a global scale [50–53]. It therefore affects the entire ecosystem, decreasing luminosity and salinity and increasing biological productivity [54]. The high microbial activity in the Amazon River may contribute to its global-scale environmental effects. Therefore, in this study we performed a detailed phylogenetic and functional analysis of three distinct thaumarchaeal genomes obtained from random shotgun sequencing of DNA from the Amazon River and its plume to infer the ecological roles and possible impacts of Thaumarchaeota in these habitats.

Results

In total, 26,207,858 Paired-end sequencing (PE) reads and 31,379 contigs were recovered from the co-assembly data of Amazon River, plus 38,598,192 PE reads and 138,265 contigs from its plume. Metagenomic analysis showed that Thaumarchaeota was the most abundant archaeal phylum in both the Amazon River and Amazon

River plume samples Fig. 1, although the relative abundance of Thaumarchaeota was higher in the river than in its plume. In the Amazon River samples Fig. 1a, Thaumarchaeota comprised approximately 10.85% of the entire microbial while Bacteria 32.91%. The archaeal community is dominated by Thaumarchaeota 97.2%, followed by Euryarchaeota and Crenarchaeota, which comprise less than 1% of the microbial community. The Thaumarchaeota sequences were most phylogenetic related to *Candidatus Nitrosotenuis uzonensis* (Thaumarchaeota archaeon N4; GenBank ID NZ_CBTY000000000.1) (82.1%), followed by other members of order Nitrosopumilales (5.5%) and *Candidatus Nitrosoarchaeum* (1.6%).

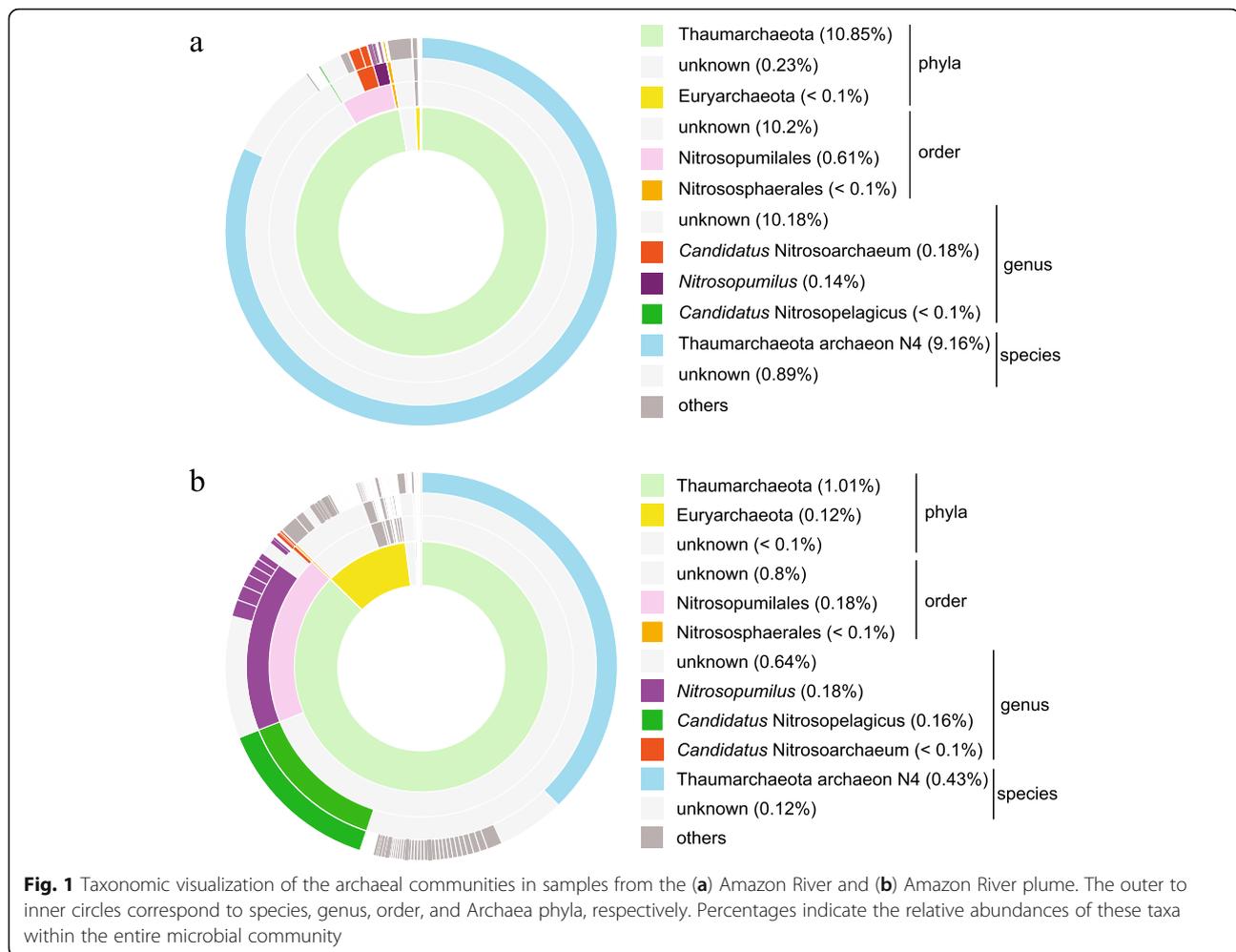
In the plume samples Fig. 1b, Thaumarchaeota comprised approximately 1.01% of the microbial community while Bacteria 80.64%, but was still the most representative Archaea in these samples (87.8%), followed by Euryarchaeota (10.4%). The Archaeal phyla Bathyarchaeota, Crenarchaeota, and Woesearchaeota each comprised less than 1% of the microbial community. Thaumarchaeota sequences in the plume were more phylogenetic related to *Candidatus Nitrosotenuis uzonensis* (37.4%), *Candidatus Nitrosopelagicus* sp. (13.9%), and other members of order Nitrosopumilales (18.3%) Fig. 1.

General genomic analyses

One near complete genome was recovered from the co-assembly data of the Amazon River (ThauR71 [1.05 Mbp]), and two from co-assembly data of the Amazon River plume (ThauP25 [0.94 Mbp] and ThauP41 [1.26 Mbp]). Phylogenetic analysis placed all three genomes in Thaumarchaeota Group 1.1a (Nitrosopumilales) Fig. 2. ThauP41 and ThauR71 were placed in the same clade as *Candidatus Nitrosotenuis cloacae* SAT1 and *Candidatus Nitrosotenuis uzonensis*, which correspond to the most abundant taxon in these areas. ThauP25 was placed in the same clade as *Candidatus Nitrosopelagicus*, which is more abundant in the Amazon River plume (0.16%) than in the Amazon River (< 0.1%).

The highest ANI value was calculated between ThauP41 and ThauR71 (98.86%), the other values were below to 95%. The ANI of ThauP41 and ThauR71 among *Candidatus Nitrosotenuis cloacae* SAT1 was 76%, whereas between ThauP25 and *Candidatus Nitrosopelagicus brevis* was 81%. The ANI of *Candidatus Nitrosotenuis uzonensis* among the three Thaumarchaeota genomes was below to 73%. All ANI were measured in both directions, but the results never varied by more than 0.01%.

Of the 38 single-copy archaeal genes identified used to measure completeness, the ThauR71 genome contained 37 (97% completeness), the ThauP25 genome contained 35 (92% completeness), and the ThauP41 genome contained 38 (complete), suggesting that each binned



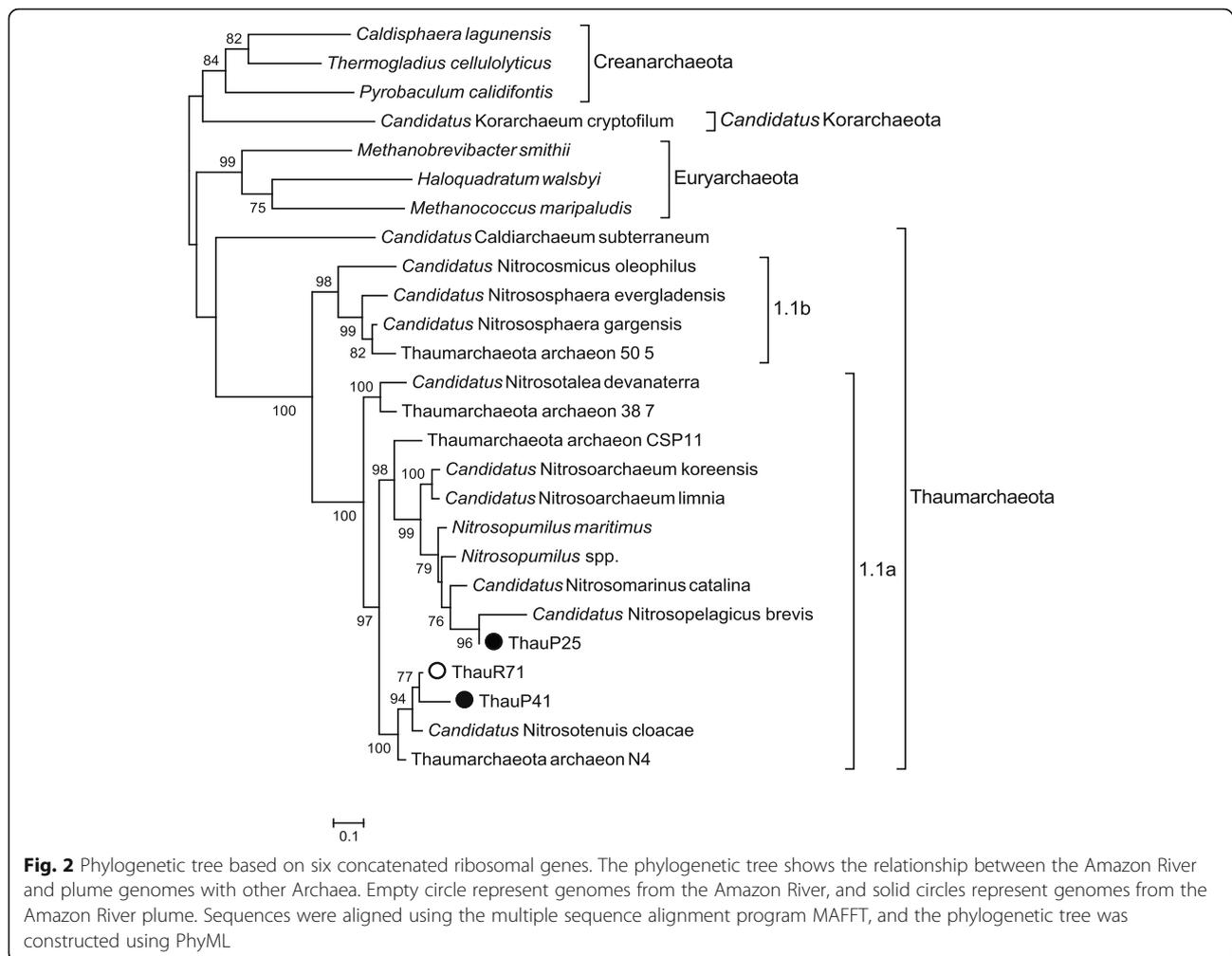
genome represented a substantial fraction of a single draft genome. The general features of these three Thaumarchaeota genomes were compared with those of the most closely related genomes: *Candidatus Nitrosotenuis cloacae* SAT1, *Candidatus Nitrosotenuis uzonensis* (Thaumarchaeota archaeon N4), and *Candidatus Nitrosopelagicus brevis* CN25 (Table 1).

The GC content of ThauP41 and ThauR71 (38%), which are both related to *Nitrosotenuis*, was greater than that of ThauP25 (33%), which is more closely related to *Candidatus Nitrosopelagicus brevis*. The GC content of *Candidatus Nitrosotenuis uzonensis* N4 was higher than the genomes obtained in this study (42.25%).

The estimated number of CDSs was 1454 for ThauR71, 1406 for ThauP25, and 1721 for ThauP41, with functions predicted for approximately 63% of these genes (40–45% predicted using the COG database); approximately 33% did not have predicted functions. The percentage of CDSs predicted to be signal peptides was 1.2–2.9% for the three genomes obtained

in this study, which is similar to that of *Candidatus Nitrosotenuis uzonensis*, *Candidatus Nitrosotenuis cloacae*, and *Candidatus Nitrosopelagicus brevis* (2.6–4.6%). Similarly, the percentage of CDSs predicted to be transmembrane proteins for the genomes obtained in this study (15.9–18.0%) is similar to that of *Candidatus Nitrosotenuis uzonensis*, *Candidatus Nitrosotenuis cloacae*, and *Candidatus Nitrosopelagicus brevis* (19.0–22.0%).

The distribution of CDSs with predicted function was similar for the Amazon River and plume genomes, which shared 186 CDSs Fig. 3. However, there was a greater number of shared CDSs between ThauP41 and ThauR71 (175) than between ThauP41 and ThauP25 (62), which were from the same habitat. The number of unique CDSs in ThauP25 (189) was higher than that of ThauP41 (136) and ThauR71 (79). *Candidatus Nitrosotenuis uzonensis* N4 shared 39 CDSs with ThauP25, 38 CDSs with ThauP41, and 33 CDSs with ThauP71. A total of 137 CDSs were shared by all four Thaumarchaeota genomes analyzed.



To analyze the three Amazon River and plume genomes, we used the COG database [55] to assign CDSs to the following main functional categories: metabolism, cellular processes and signaling, information storage and processing, and poorly characterized. Within the metabolism category, the highest percentage of genes was assigned to the subcategory coenzyme transport and metabolism (11–12%), which consists of genes involved mainly in the synthesis of cobalamin, biotin, pantothenate, menaquinone, phyloquinone, siroheme, coenzyme F430, coenzyme A, cytochrome c oxidase, coenzyme A, and bacteriochlorophyll c and d. Coenzyme transport and metabolism genes were most abundant in ThauP25 (12.52%), followed by ThauR71 (11.75%). The abundance of genes in other metabolism subcategories was 8–9% for amino acid transport and metabolism, 5–7% for energy production and conversion, 5–6% for nucleotide transport and metabolism, 5% for inorganic ion transport and metabolism, 4–5% for carbohydrate transport and metabolism, 2–3% for lipid transport and metabolism, and 1% for secondary metabolites biosynthesis, transport, and catabolism.

In the cellular processes and signaling category, the highest percentage of genes was assigned to the subcategory posttranslational modification, protein turnover, chaperones (5–6%), followed by signal transduction mechanisms (1–2%), wall/membrane/envelope biogenesis (1–2%), cell cycle control, cell division, chromosome partitioning (1%), defense mechanisms (1%), intracellular trafficking, secretion, and vesicular transport (0.8–1%), and cell motility (0.3–0.8%).

For the information storage and processing category, the highest percentage of genes was assigned to the subcategory translation, ribosomal structure, and biogenesis (15–17%), followed by transcription (4–5%), and replication, recombination, and repair (3%). Genes assigned to the subcategories of chromatin structure and dynamics (<1%) and RNA processing and modification (<1%) were not present in all three genomes. Some genes were considered poorly characterized or had only general function predicted (6–7%), or function was unknown (3%) (Table 2). The percentages described above for each category represent the range of these genes in the three genomes obtained in this study.

Table 1 Comparison of general genome features of three Amazon Thaumarchaeota genomes and their phylogenetically closest members

	ThauP25	ThauP41	ThauR71	N4	SAT1	CN25
Genome size, bp	934,797	1,256,699	1,047,532	1,636,125	1,620,156	1,232,128
DNA coding region, bp	878,769	1,159,316	970,632	1,500,929	1,499,274	1,164,105
G + C content, mol%	33.15	37.82	37.99	42.25	41.00	33.16
Total RNA genes, n	47	50	30	46	48	47
tRNA genes, n	44	43	27	41	43	42
rRNA genes, n	0	3	1	3	3	3
Other RNA genes, n	3	4	2	2	2	2
Total number of genes, n	1453	1771	1454	1955	1924	1516
Total CDSs, n (%)	1406 (96.77)	1721 (97.18)	1424 (97.94)	1909 (97.65)	1876 (97.51)	1469 (96.90)
With predicted function, n (%)	920 (63.32)	1119 (63.18)	944 (64.92)	1233 (63.07)	1212 (62.99)	985 (64.97)
Predicted with COG database, n (%)	583 (40.12)	791 (44.66)	660 (45.39)	1006 (51.46)	998 (51.87)	850 (56.07)
Predicted with TIGRFAMs database, n (%)	333 (22.92)	402 (22.70)	343 (23.59)	456 (23.32)	463 (24.06)	413 (27.24)
Encoding signal peptides, n (%)	18 (1.24)	51 (2.88)	36 (2.48)	81 (4.14)	89 (4.63)	40 (2.64)
Encoding transmembrane proteins, n (%)	231 (15.90)	319 (18.01)	255 (17.54)	428 (21.89)	424 (22.04)	289 (19.06)
Without predicted function, n (%)	486 (33.45)	602 (33.99)	480 (33.01)	676 (34.58)	644 (34.51)	484 (31.93)
In internal clusters, n (%)	381 (26.22)	457 (25.80)	234 (16.09)	149 (7.62)	137 (7.12)	66 (4.35)
Metadata						
Isolation source, habitat	Seawater	Seawater	Freshwater	Geothermal hot spring	Wastewater	Seawater
	Atlantic Ocean	Atlantic Ocean	Brazil	Russian	China	Pacific Ocean
Thaumarchaeota group	1.1a	1.1a	1.1a	1.1a	1.1a	1.1a
Species	<i>Nitrosopelagicus</i> sp.	<i>Nitrosotenuis</i> sp.	<i>Nitrosotenuis</i> sp.	<i>Nitrosotenuis uzonensis</i>	<i>Nitrosotenuis cloacae</i>	<i>Nitrosopelagicus brevis</i>

Abbreviations: CDSs coding sequences, COG clusters of orthologous groups, rRNA ribosomal RNA, tRNA transfer RNA

Ecophysiology

All three genomes obtained in this study contained genes involved in energy conversion by nitrogen metabolism, oxidative phosphorylation, sulfur metabolism. Genes for carbon fixation were also found in these genomes, they had genes representing a near-complete

hydroxypropionate/hydroxybutyrate (HP/HB) cycle and the presence of the gene encoding acetyl/propionyl-CoA carboxylase, which is responsible for HCO₃ assimilation. However, genes encoding malonyl-CoA reductase and succinate-semialdehyde dehydrogenase were not detected in any of the genomes. The three genomes also

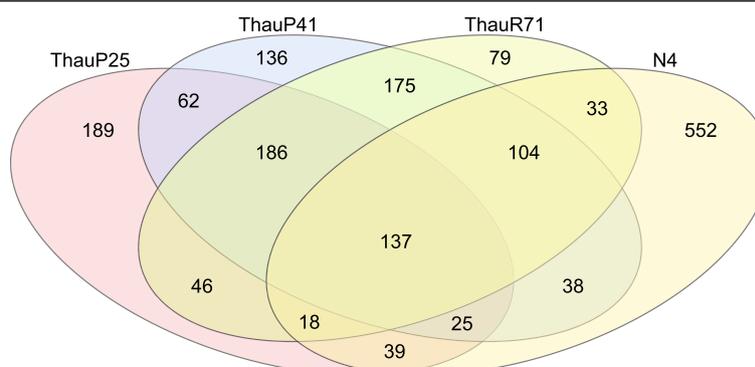


Fig. 3 Venn diagram of CDSs that are unique or shared between ThauP25, ThauP41, ThauR71, and *Candidatus Nitrosotenuis uzonensis* N4 (N4)

Table 2 Comparison of gene content by COG functional categories

COG		Proteins among:											
		ThauP25		ThauP41		ThauR71		N4		SAT1		CN25	
Code	Description	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
[C]	Energy production and conversion	51	7.98	52	6.08	41	5.73	67	6.08	69	6.29	60	6.49
[G]	Carbohydrate transport and metabolism	32	5.01	38	4.44	34	4.76	43	3.9	46	4.19	43	4.65
[E]	Amino acid transport and metabolism	60	9.39	72	8.42	65	9.09	102	9.26	106	9.66	90	9.74
[F]	Nucleotide transport and metabolism	37	5.79	56	6.55	46	6.43	60	5.44	59	5.38	53	5.74
[H]	Coenzyme transport and metabolism	80	12.52	97	11.35	84	11.75	101	9.17	102	9.3	100	10.82
[I]	Lipid transport and metabolism	20	3.13	30	3.51	20	2.8	30	2.72	33	3.01	31	3.35
[P]	Inorganic ion transport and metabolism	33	5.16	49	5.73	41	5.73	59	5.35	60	5.47	38	4.11
[Q]	Secondary metabolites biosynthesis, transport and catabolism	6	0.94	13	1.52	13	1.82	14	1.27	16	1.46	13	1.41
[D]	Cell cycle control, cell division, chromosome partitioning	10	1.56	12	1.4	7	0.98	11	1	11	1	8	0.87
[M]	Cell wall/membrane/envelope biogenesis	12	1.88	18	2.11	11	1.54	30	2.72	36	3.28	35	3.79
[N]	Cell motility	5	0.78	3	0.35	2	0.28	20	1.81	20	1.82	5	0.54
[O]	Posttranslational modification, protein turnover, chaperones	40	6.26	50	5.85	49	6.85	56	5.08	64	5.83	46	4.98
[T]	Signal transduction mechanisms	11	1.72	23	2.69	20	2.8	48	4.36	44	4.01	14	1.52
[U]	Intracellular trafficking, secretion, and vesicular transport	9	1.41	7	0.82	6	0.84	12	1.09	10	0.91	9	0.97
[V]	Defense mechanisms	9	1.41	12	1.4	8	1.12	17	1.54	17	1.55	9	0.97
[A]	RNA processing and modification	0	0	1	0.12	1	0.14	1	0.09	1	0.09	0	0
[B]	Chromatin structure and dynamics	1	0.16	2	0.23	2	0.28	2	0.18	2	0.18	2	0.22
[J]	Translation, ribosomal structure and biogenesis	102	15.96	152	17.78	121	16.92	169	15.34	172	15.68	165	17.86
[K]	Transcription	33	5.16	41	4.8	38	5.31	61	5.54	61	5.56	44	4.76
[L]	Replication, recombination and repair	24	3.76	31	3.63	28	3.92	43	3.9	43	3.92	41	4.44
[R]	General function prediction only	41	6.42	67	7.84	55	7.69	106	9.62	88	8.02	79	8.55
[S]	Function unknown	23	3.6	29	3.39	23	3.22	47	4.26	36	3.28	36	3.9
(-)	Not in COG	870	59.88	980	55.34	794	54.61	949	48.54	926	48.13	666	43.93

contained genes for carbon fixation through the Calvin cycle, but the ribulose 1,5-biphosphate carboxylase were not detected for these genomes.

For nitrogen metabolism, genes encoding ammonia monooxygenase were detected in ThauP41 (*amoA*, *amoB*, and *amoC*) and ThauR71 (*amoB* and *amoC*) but not in ThauP25. All three Amazon River and plume genomes contained *nirK*, which is involved in the production of nitric oxide and nitrous oxide, as well as genes involved in glycolysis, TCA cycle, and oxidative phosphorylation.

Analysis of carbohydrate-active enzymes showed that GH genes were detected just in ThauP25, which have 5 genes from the GH1 family (Additional file 2: Table S1). Genes assigned to the enzyme classes AA (15 genes from AA1 family) were only detected in ThauP25. The families GT1 and GT2 were detected in all three genomes, whereas the ThauP25 has the higher abundance of GT2 (123 vs. mean of 57.5). The family GT66 were just detected for ThauP41 and ThauR71, whereas GT7 and GT90 were just detected for ThauR71. The enzyme classes CE was not detected in any of these genomes.

The genomes ThauP41 and ThauR71 encode a complete assimilatory sulfate reduction pathway, whereas in ThauP25 the *sir* gene was not detected. In addition, genes specific for the anaerobic synthesis of cobalamin were detected in ThauP25, ThauP41 (*gbix*, *cbiG*, and *cbiD*), and ThauR71 (*cbiG* and *cbiD*). Genes of the second cluster of biosynthesis of cobalamin were also detected in ThauR71 (*cobY*, *cobS*, *cobD*, and *cobQ*), the likewise occurs for ThauP25 and ThauP41, though, *cobQ* was not detected.

All three Thaumarchaeota genomes contained the complete set of genes for transport of zinc (*znuA*, *znuB*, *znuC*). Moreover, it was detected genes for transport of phosphate (*pstS*, *pstA*, *pstC*, *pstB*) and lipoprotein (*lolC*, *lolE*, *lolD*) in ThauP41.

Discussion

Because of the lack of representative Thaumarchaeota genomes, our metagenomics analysis classified many of the sequences obtained in our study as unknown Fig. 1. Most of the sequences of the Thaumarchaeota from the Amazon River and its plume were assigned to members of group

1.1a (Nitrosopumilales) [13] or Marine Group I [2, 19, 56]), suggesting that this is the most abundant archaeal group in these habitats. These results are consistent with previous studies that report that 73% of Thaumarchaeota from estuarine-coastal AOA belong to Nitrosopumilales, and 37% are associated specifically with the clade of *Nitrosopumilus* spp. [13]. The relative abundance of Thaumarchaeota was higher in the Amazon River than in its plume, which might be attributed to the difference in their physical and chemical characteristics. Salinity is lower in the Amazon River (0.01–0.02 g/kg) than in its plume (12.17–36.34 g/kg), as is pH (6.58–6.99 vs. 7.67–8.1) and surface dissolved inorganic carbon (113–502 vs. 818–2030 $\mu\text{mol}\cdot\text{C}/\text{kg}$). In contrast, partial pressure of carbon dioxide is higher in the Amazon River (1039–5488 μatm) than its plume (234–569 μatm), as is ammonium (0.72–3.12 vs. 0.0–0.014 μM), nitrate + nitrite (5.64–17.9 vs. 0.0–8.52 μM), and dissolved organic carbon (244–381 vs. 61–191 μM) [57]. The higher pressure of carbon dioxide in the Amazon River is probably associated with heterotrophic carbon processing of allochthonous organic matter derived from plant material [58, 59]. Ammonium, nitrate+nitrite and dissolved carbon concentrations may also be associated with the allochthonous organic matter input from the surrounding Amazon forest [60]. Thus, greater availability of nitrogen ammonium, nitrate, and nitrite suggests the predominance of Thaumarchaeota Group 1.1a, since they are nitrifying Archaea. Soil metagenomics studies have also found that Group 1.1a is associated with higher nitrate concentrations and lower pH ($5 \leq \text{pH} < 7$) [11, 61]. Furthermore, the clades NP- η (*Candidatus Nitrosotenuis uzonensis*) and NP- γ (*Nitrosopumilus* spp.) appear to be associated with low-salinity and freshwater environments [13, 62, 63], which would account for the considerable decrease in Thaumarchaeota in Amazon River plume, since the most abundant taxa in our analysis belong to these clades. Nevertheless, the change from estuarine to marine environment would explain the increase of clade NP- ϵ (*Candidatus Nitrosopelagicus*), which suggests niche specialization to plume habitat that is correlated with saturation of dissolved oxygen, surface dissolved inorganic carbon, pH, and salinity. Other studies have corroborated the association of this clade with seawater [13]. In addition, clades NP- ϵ (*Candidatus Nitrosopelagicus*) and NP- γ (*Nitrosopumilus*) are associated with “shallow” water (Water Column A) [13, 64–69]. These findings may account for the abundance of Group 1.1a and the low representation of Groups 1.1a-associated, 1.1b, ThAOA, 1.1c, and 1.3. Thaumarchaeota represent an important fraction of the microbial community in the Amazon River. Although more abundant in the Amazon River, their presence in the plume extending to the Atlantic Ocean.

The genomes recovered in this study (ThauP25, ThauP41, and ThauR71) were assigned to the most abundant taxa of these environments Fig. 1. However, it

happens probably because of lack of the known species in the database, since ANI analysis suggests that these genomes represent different Thaumarchaeota species from those of literature. Studies have shown that ANI values equal to approximately 95% correspond to the cut-off values of 70% similarity used on DNA-DNA hybridization as a standard for the classification of new species [70–73]. Therefore, ThauP41, and ThauP25 belong to known genera according to the presented phylogeny.

Our findings suggest that these organisms are prototroph for cobalamin (vitamin B12), biotin (vitamin B7), pantothenate (vitamin B5), menaquinone (vitamin K2), and phyloquinone (vitamin K1) and not require uptake, which has previously been shown by culture of AOA without vitamin supplements [27, 74]. Other studies have reported that Thaumarchaeota plays an important role in the global production of cobalamin [75], thus Thaumarchaeota may be acting and the production of vitamins to this environment.

These three recovered genomes contain genes for the HP/HB cycle, although genes encoding malonyl-CoA reductase and succinate-semialdehyde dehydrogenase were not identified. However, many genes encoding enzymes in Thaumarchaeota do not have homologs in databases and cannot be identified using bioinformatics approach, including a malonyl-CoA reductase in *Nitrosopumilus maritimus* [43, 69, 76]. *N. maritimus* has an alternative enzyme for this step, malonic semialdehyde reductase, which was detected by homology in ThauP25 and ThauP41. AOA are thought to be major contributors to carbon fixation in oceans [77], indicating these organisms contribute to primary production in Atlantic Ocean. On the other hand, the high nutrient and CO_2 levels in the Amazon River [78] may increase primary production, leading to eutrophication, higher consumption of dissolved oxygen. The presence of genes encoding glycosyl hydrolases detected in ThauP25 and proteins involved in glycolysis and the TCA cycle suggests that these organisms mixotrophic metabolism, which supports the hypothesis that Thaumarchaeota are not only chemoautotrophic [30]. Previous studies have shown that addition of organic acids can improve the growth of *N. maritimus* [27], furthermore, they demonstrate that the metabolism of some Thaumarchaeota keeps activated when nitrification inhibitor was added [79].

Two of the genomes (ThauP41 and ThauR71) suggest participation in nitrification; however, the most likely reason that AOA genes were not detected in ThauP25 is that this genome is not complete. The gene *nirK* was detected in all three genomes, indicating participation in the production of the greenhouse gas nitrous oxide, which was demonstrated by heterologous expression of *nirK* from *Nitrososphaera viennensis* [80]. The presence of genes encoding coenzyme F420 are widespread in

Thaumarchaeota genomes [9, 81]. This coenzyme has been detected in high amount in a proteomic study [82], suggesting an important role in its metabolism. The most probable hypothesis for this coenzyme is a protective role against reactive nitrogen species like observed in mycobacteria [83].

Conclusions

Thaumarchaeota is the most abundant archaeal phylum in the Amazon River and Amazon River plume. Both areas are dominated by taxa closely related to *Candidatus Nitrosotenuis uzonensis*. The relative abundance of these Thaumarchaeota is associated with neutral to acidic pH and higher concentrations of nitrate, nitrite, and ammonia. The three genomes obtained in this study provide evidence of a typical Thaumarchaeota metabolism, with energy production by nitrification and carbon assimilation through the HP/HB cycle, suggesting that they are autotrophs. The presence of genes related to the TCA cycle and glycolysis suggest mixotrophic metabolism. In addition, these genomes contain genes involved in the production of nitrous oxide, and vitamins. Similarities between the genomes of Thaumarchaeota from the Amazon River and its plume suggest that all three organisms are present in both habitats, with the Amazon River representing a potential source of Thaumarchaeota in this habitat, where these organisms are of paramount importance.

Methods

Sampling and DNA extraction

We collected and analyzed 12 water samples from the Amazon River and its plume as described by Silva et al. 2017 [57]. Microbial DNA from samples preserved in SET buffer (20% sucrose, 50 mM EDTA, 0.5 mM Tris HCl) was extracted using the NucleoSpin Tissue kit (Macherey-Nagel, Dueren, Germany) with the following modifications for DNA extraction from Sterivex™ filters [84]: after incubation with lysozyme (final concentration 0.1%) at 37 °C for 60 min, 20% SDS (final concentration 1%) with proteinase K was added to the samples at 37 °C for 30 min, followed by washing with chloroform:isoamyl alcohol (24:1) solution. The extracted DNA was analyzed by agarose gel electrophoresis (1% agarose in 1× Tris-borate-EDTA) and then quantified spectrophotometrically with a NanoDrop ND-1000 (Thermo Scientific, DE, USA) and enzymatically with a Qubit High Sensitivity Fluorometer (Life Technologies). The DNA was amplified with the GenomiPhi V3 Kit (GE Healthcare) according to the manufacturer's instructions but with the amplification time increased to 16 h or 18 h [85].

Illumina library construction and sequencing

DNA libraries were prepared using the Nextera XT Sample Preparation Kit (Illumina). Library size distribution was determined using the Agilent 2100 Bioanalyzer and High Sensitivity DNA Kit (Agilent) and quantified using the Applied Biosystems 7500 Real Time PCR System and KAPA Library Quantification Kit (Kapa Biosystems). The PhiX sequencing control v3 (Illumina) was added at 1%, and paired-end sequencing (2 × 250 bp) was performed on a MiSeq System (Illumina).

Metagenomic analysis, binning, and annotation

Reads were trimmed and quality checked using Sickle version 1.33 with default parameters. The reads of the five samples of Amazon River (Tapajós, Óbidos, North Macapá, South Macapá, and Belém) were combined and then assembled into scaffolds using IDBA-ud [86]. The same was proceeded for the seven samples (St1, St3, St4, St6, St10, St11, and St15) of the Amazon River plume. All metagenome data from each sample were uploaded to ggKbase.

The microbial community classification was based on phylogenetic makeup among the genes that compose a specific contig. When more than 50% of the genes that make up a contig belong to an organism in the NCBI database, taxonomic attribution is given. The taxonomy of sequences from the Amazon River and its plume was visualized using the “View project taxonomy” tool. All this information is available on the website (<https://ggkbase-help.berkeley.edu/>).

Genomes were binned using ggKbase binning tools (ggkbase.berkeley.edu) and manually curated to remove scaffolds with anomalous GC content, coverage, or taxonomic classification [87]. Genome completeness was estimated based on 38 archaeal single-copy genes, as described previously [88, 89]. Each assembled genome was uploaded to the Integrated Microbial Genomes (IMG) system at the US Department of Energy's Joint Genome Institute for annotation [90, 91].

Protein-coding genes of ThauP25, ThauP41, ThauR71, and N4 were assigned to COG & KOG functional category [55] by comparison to COG PSSMs using RPS-BLAST with E-value cutoff of 1e-2, and minimum alignment length of 70% of the consensus sequence length with the top hit [91].

Metabolic reconstruction and pathway mapping of each assembled genome was completed using the KEGG Automatic Annotation Server [92].

Catalytic and carbohydrate-binding modules of enzymes that degrade, modify, or create glycosidic bonds were identified by searching against the Carbohydrate-Active enZymes database (<http://www.cazy.org/>) [93] using Diamond version 0.9.22 [94] with E-value cutoff of <1e-102. These enzymes and associated modules include glycoside hydrolases (GHs), enzymes with auxiliary activities (AAs),

carbohydrate esterases (CEs), glycosyl transferases (GTs), and carbohydrate-binding modules (CBMs).

We construct a Venn diagram using the *gplots* package in R software [95] to identify unique and shared genes, utilizing protein-coding genes with predicted function of *Candidatus Nitrosopelagicus brevis* bin25 (ThauP25; IMG Genome ID 2781125713), *Candidatus Nitrosotenuis* sp. bin41 (ThauP41; IMG Genome ID 2781125712), *Candidatus Nitrosotenuis* sp. bin71 (ThauR71; IMG Genome ID 2778260922), and *Candidatus Nitrosotenuis uzonensis* N4 (N4; IMG Genome ID 2619619229). We compared the genomes acquired in this study (ThauP25, ThauP41, and ThauR71) with the genome of *Candidatus Nitrosotenuis uzonensis* N4, isolated from geothermal environments [96].

Phylogenetic analysis

To construct the phylogenetic tree, sequences of the concatenated ribosomal genes S11, S7, S2, L21, L11, and L3 were aligned with MAFFT version 7.388 [97], using default parameters, and then manually edited to delete the extra end of the long sequences. The phylogenetic tree was constructed in PhyML version 20,131,022 [98] with the BLOSUM62 matrix and 1000 bootstrap replications. The Average Nucleotide Identity (ANI) was calculated using a Pairwise ANI tools available on IMG [90, 99]. The GenBank accession numbers of these sequences are available in Additional file 1: Dataset S1.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12866-020-1698-x>.

Additional file 1: Dataset S1. provides GenBank accession numbers and ribosomal gene sequences for the genomes used to construct the phylogenetic tree.

Additional file 2: Table S1. lists the protein families of the Carbohydrate-Active enZYmes database detected in the genomes of ThauP25, ThauP41, and ThauR71.

Abbreviations

ANI: Average Nucleotide Identity; AOA: Ammonia-oxidizing archaea; CDS: Coding sequence; CE: Carbohydrate esterase; COG: Clusters of orthologous groups; DNA: Deoxyribonucleic acid; GH: Glycoside hydrolase; GT: Glycosyl transferase; HP/HB: Hydroxypropionate/Hydroxybutyrate; PE: Paired-end sequencing; RNA: Ribonucleic acid; rRNA: Ribosomal RNA; TCA: Tricarboxylic acid cycle/rNATransfer RNA

Acknowledgments

The authors thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), and Fundação de Apoio à Pesquisa do Distrito Federal (FAP-DF) for financial support. Pinto, Silva and Lopes acknowledge fellowships from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). We would like to thank Dr. Jillian F. Banfield and Brian C. Thomas from the UC Berkeley for the use of the ggkbase pipeline.

Authors' contributions

RK, TS, CV, and RS designed the experiments. Data analyses were performed by OP and FL, data collection by FT and BS. The manuscript was drafted by OP

and revised by all authors. All authors have read and approved the manuscript.

Funding

This study was financially supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), which also provided financial support for the first author (OP), Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), and Fundação de Apoio à Pesquisa do Distrito Federal (FAP-DF). The funding bodies had no role in the design of the study, collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

Genomes described in this article are available on the Integrated Microbial Genomes system (Accession numbers: ThauP25; IMG Genome ID 2781125713, ThauP41; IMG Genome ID 2781125712, ThauR71; IMG Genome ID 2778260922). Accession numbers of sequences used to construct the phylogenetic tree are available in the Supplemental Material Dataset S1.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Enzymology, Institute of Biological Sciences, University of Brasília, Brasília 70910-900, Brazil. ²Department of Civil and Environmental Engineering, University of Brasília, Brasília 70910-900, Brazil. ³Federal Institute of Brasília, Brasília 70830-450, Brazil. ⁴Laboratory of Microbiology, Federal University of Tocantins, Palmas 77500-000, Brazil. ⁵Department of Genetics, Institute of Biology, Federal University of Rio de Janeiro, Rio de Janeiro 21941-901, Brazil.

Received: 30 May 2019 Accepted: 9 January 2020

Published online: 15 January 2020

References

- Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P. Mesophilic crenarchaeota: Proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol*. 2008;6:245–52.
- Delong EF. Archaea in coastal marine environments. *Proc Natl Acad Sci USA*. 1992;89:5685–9.
- Schleper C, Jurgens G, Jonuscheit M. Genomic studies of uncultivated archaea. *Nat Rev Microbiol*. 2005;3:479–88.
- Matte-Tailliez O, Brochier C, Forterre P, Philippe H. Archaeal phylogeny based on ribosomal proteins. *Mol Biol Evol*. 2002;19:631–9.
- Spang A, Hatzepichler R, Brochier-Armanet C, Rattei T, Tischler P, Spieck E, et al. Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. *Trends Microbiol*. 2010;18:331–40.
- Petitjean C, Deschamps P, López-García P, Moreira D. Rooting the domain archaea by phylogenomic analysis supports the foundation of the new kingdom Proteoarchaeota. *Genome Biol Evol*. 2014;7:191–204.
- Brochier-Armanet C, Gribaldo S, Forterre P. Spotlight on the Thaumarchaeota. *ISME J*. 2012;6:227–30.
- Hallam SJ, Konstantinidis KT, Putnam N, Schleper C, Watanabe Y, Sugahara J, et al. Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc Natl Acad Sci U S A*. 2006;103:18296–301.
- Stieglmeier M, RJE A, Schleper C. The Prokaryotes; 2014.
- Lin X, Handley KM, Gilbert JA, Kostka JE. Metabolic potential of fatty acid oxidation and anaerobic respiration by abundant members of Thaumarchaeota and Thermoplasmata in deep anoxic peat. *ISME J*. 2015;9:2740–4.
- Oton EV, Quince C, Nicol GW, Prosser JI, Gubry-Rangin C. Phylogenetic congruence and ecological coherence in terrestrial Thaumarchaeota. *ISME J*. 2016;10:85–96.
- Kerou M, Alves RJE, Schleper C, Stieglmeier, Klingl, Alves, Rittmann, Melcher, Leisch and Schleper 2014a, 2747; 2016.

13. Alves RJE, Minh BQ, Ulrich T, Von Haeseler A, Schleper C. Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on amoA genes. *Nat Commun*. 2018;9:1–17.
14. Francis CA, Roberts KJ, Beman JM, Santoro AE, Oakley BB. Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proc Natl Acad Sci*. 2005;102:14683–8.
15. Treusch AH, Leininger S, Kietzin A, Schuster SC, Klenk HP, Schleper C. Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. *Environ Microbiol*. 2005;7:1985–95.
16. Leininger S, Ulrich T, Schloter M, Schwark L, Qi J, Nicol GW, et al. Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature*. 2006;442:806–9.
17. Mincer TJ, Church MJ, Taylor LT, Preston C, Karl DM, DeLong EF. Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. *Environ Microbiol*. 2007;9:1162–75.
18. Adair KL, Schwartz E. Evidence that ammonia-oxidizing archaea are more abundant than ammonia-oxidizing bacteria in semiarid soils of northern Arizona, USA. *Microb Ecol*. 2008;56:420–6.
19. Agogue H, Brink M, Dinasquet J, Herndl GJ. Major gradients in putatively nitrifying and non-nitrifying Archaea in the deep North Atlantic. *Nature*. 2008;456:788–92.
20. Reigstad LJ, Richter A, Daims H, Ulrich T, Schwark L, Schleper C. Nitrification in terrestrial hot springs of Iceland and Kamchatka. *FEMS Microbiol Ecol*. 2008;64:167–74.
21. Shen JP, Zhang LM, Zhu YG, Zhang JB, He JZ. Abundance and composition of ammonia-oxidizing bacteria and ammonia-oxidizing archaea communities of an alkaline sandy loam. *Environ Microbiol*. 2008;10:1601–11.
22. Dang H, Luan XW, Chen R, Zhang X, Guo L, Klotz MG. Diversity, abundance and distribution of amoA-encoding archaea in deep-sea methane seep sediments of the Okhotsk Sea. *FEMS Microbiol Ecol*. 2010;72:370–85.
23. Gubry-Rangin C, Nicol GW, Prosser JI. Archaea rather than bacteria control nitrification in two agricultural acidic soils. *FEMS Microbiol Ecol*. 2010;74:566–74.
24. Mussmann M, Brito I, Pitcher A, Sinnighe Damste JS, Hatzepichler R, Richter A, et al. Thaumarchaeotes abundant in refinery nitrifying sludges express amoA but are not obligate autotrophic ammonia oxidizers. *Proc Natl Acad Sci*. 2011;108:16771–6.
25. Zheng M, Fu HZ, Ho YS. Research trends and hotspots related to ammonia oxidation based on bibliometric analysis. *Environ Sci Pollut Res*. 2017;24:20409–21.
26. Könneke M, Bernhard AE, De La Torre JR, Walker CB, Waterbury JB, Stahl DA. Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature*. 2005;437:543–6.
27. Qin W, Heal KR, Ramdasi R, Kobelt JN, Martens-Habbena W, Bertagnolli AD, et al. *Nitrosopumilus maritimus* gen. nov., sp. nov., *Nitrosopumilus cobalaminigenes* sp. nov., *Nitrosopumilus oxyclinae* sp. nov., and *Nitrosopumilus urephilus* sp. nov., four marine ammoniaoxidizing archaea of the phylum thaumarchaeo. *Int J Syst Evol Microbiol*. 2017;67:5067–79.
28. Stein LY. Insights into the physiology of ammonia-oxidizing microorganisms. *Curr Opin Chem Biol*. 2019;49:9–15.
29. Tournai M, Stieglmeier M, Spang A, Könneke M, Schintlmeister A, Ulrich T, et al. *Nitrososphaera viennensis* sp. nov., an ammonia oxidizing archaeon from soil. *Proc Natl Acad Sci*. 2011;108:8420 LP – 8425.
30. Pester M, Schleper C, Wagner M. The Thaumarchaeota: An emerging view of their phylogeny and ecophysiology. *Curr Opin Chem Biol*. 2011;14:300–6.
31. Karner MB, Delong EF, Karl DM. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature*. 2001;409:507–10.
32. Ochsenreiter T, Selezni D, Quaiser A, Bonch-Osmolovskaya L, Schleper C. Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. *Environ Microbiol*. 2003;5:787–97.
33. Tolar BB, King GM, Hollibaugh JT. An analysis of thaumarchaeota populations from the northern Gulf of Mexico. *Front Microbiol*. 2013;4:72.
34. Erguder TH, Boon N, Wittebolle L, Marzorati M, Verstraete W. Environmental factors shaping the ecological niches of ammonia-oxidizing archaea. *FEMS Microbiol Rev*. 2009;33:855–69.
35. Lehtovirta-Morley LE, Stoecker K, Vilcinskas A, Prosser JI, Nicol GW. Cultivation of an obligate acidophilic ammonia oxidizer from a nitrifying acid soil. *Proc Natl Acad Sci*. 2011;108:15892–7.
36. Hu A, Yang Z, Yu CP, Jiao N. Dynamics of Autotrophic Marine Planktonic Thaumarchaeota in the East China Sea. *PLoS ONE*. 2013;8:20–3.
37. Vissers EW, Blaga CI, Bodelier PLE, Muyzer G, Schleper C, Sinnighe Damsté JS, et al. Seasonal and vertical distribution of putative ammonia-oxidizing thaumarchaeotal communities in an oligotrophic lake. *FEMS Microbiol Ecol*. 2013;83:515–26.
38. De La Torre JR, Walker CB, Ingalls AE, Könneke M, Stahl DA. Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. *Environ Microbiol*. 2008;10:810–8.
39. Wuchter C, Abbas B, Coolen MJL, Herfort L, van Bleijswijk J, Timmers P, et al. Archaeal nitrification in the ocean. *Proc Natl Acad Sci*. 2006;103:12317–22.
40. Stahl DA, de la Torre JR. Physiology and Diversity of Ammonia-Oxidizing Archaea. *Annu Rev Microbiol*. 2012;66:83–101.
41. Martens-Habbena W, Berube PM, Urakawa H, De La Torre JR, Stahl DA. Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature*. 2009;461:976–9.
42. Walker CB, de la Torre JR, Klotz MG, Urakawa H, Pinel N, Arp DJ, et al. *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc Natl Acad Sci*. 2010;107:8818–23.
43. Könneke M, Schubert DM, Brown PC, Hugler M, Standfest S, Schwander T, et al. Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO₂ fixation. *Proc Natl Acad Sci*. 2014;111:8239–44.
44. Satinsky BM, Crump BC, Smith CB, Sharma S, Zielinski BL, Doherty M, et al. Microspatial gene expression patterns in the Amazon River Plume. *Proc Natl Acad Sci*. 2014;111:11085–90.
45. Satinsky BM, Smith CB, Sharma S, Ward ND, Krusche AV, Richey JE, et al. Patterns of Bacterial and Archaeal Gene Expression through the Lower Amazon River. *Front Mar Sci*. 2017;4:1–15.
46. Richey JE, Hedges JJ, Devol AH, Quay PD, Victoria R, Martinelli L, et al. Biogeochemistry of carbon in the Amazon River. *Limnol Oceanogr*. 1990;35:352–71.
47. Dai A, Trenberth KE. Estimates of Freshwater Discharge from Continents: Latitudinal and Seasonal Variations. *J Hydrometeorol*. 2002;3:660–87.
48. Kang Y, Pan D, Bai Y, He X, Chen X, Chen CTA, et al. Areas of the global major river plumes. *Acta Oceanologica Sinica*. 2013;32:79–88.
49. Horner-Devine AR, Hetland RD, MacDonald DG. Mixing and Transport in Coastal River Plumes. *Annu Rev Fluid Mech*. 2015;47:569–94.
50. Lentz SJ. Seasonal variations in the horizontal structure of the Amazon plume inferred from historical hydrographic data. *J Geophys Res*. 1995;100:2391–400.
51. Lentz SJ. The Amazon River plume during AMASEDS: subtidal current variability and the importance of wind forcing. *J Geophys Res*. 1995;100:2377–90.
52. Körtzinger A. A significant CO₂ sink in the tropical Atlantic Ocean associated with the Amazon River plume. *Geophys Res Lett*. 2003;30:2–5.
53. Cooley SR, Coles VJ, Subramaniam A, Yager PL. Seasonal variations in the Amazon plume-related atmospheric carbon sink. *Global Biogeochem Cycles*. 2007;21:1–15.
54. Moura RL, Amado-Filho GM, Moraes FC, Brasileiro PS, Salomon PS, Mahiques MM, et al. An extensive reef system at the Amazon River mouth - Supplementary Material. *Sci Adv*. 2016;2:e1501252.
55. Tatusov RL. A Genomic Perspective on Protein Families. *Science*. 1997;278:631–7.
56. Massana R, Delong EF, Pedrós-Alió C. A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. *Appl Environ Microbiol*. 2000;66:1777–87.
57. Silva BSO, Coutinho FH, Gregoracci GB, Leomil L, de Oliveira LS, Fróes A, et al. Virioplankton assemblage structure in the lower river and ocean continuum of the Amazon. *mSphere*. 2017;2:e00366–17.
58. Mayorga E, Aufdenkampe AK, Masiello CA, Krusche AV, Hedges JJ, Quay PD, et al. Young organic matter as a source of carbon dioxide outgassing from Amazonian rivers. *Nature*. 2005;436:538–41.
59. Sawakuchi HO, Neu V, Ward ND, de LC BM, Valerio AM, Gagne-Maynard W, et al. Carbon Dioxide Emissions along the Lower Amazon River. *Front Mar Sci*. 2017;4:76.
60. Tank JL, Rosi-Marshall EJ, Griffiths NA, Enterekin SA, Stephen ML. A review of allochthonous organic matter dynamics and metabolism in streams. *J North Am Benthol Soc*. 2010;29:118–46.
61. Gubry-Rangin C, Hai B, Quince C, Engel M, Thomson BC, James P, et al. Niche specialization of terrestrial archaeal ammonia oxidizers. *Proc Natl Acad Sci*. 2011;108:21206–11.
62. Mosier AC, Francis CA. Relative abundance and diversity of ammonia-oxidizing archaea and bacteria in the San Francisco Bay estuary. *Environ Microbiol*. 2008;10:3002–16.
63. Restrepo-Ortiz CX, Auguet JC, Casamayor EO. Targeting spatiotemporal dynamics of planktonic SAGMG-1 and segregation of ammonia-oxidizing thaumarchaeota ecotypes by newly designed primers and quantitative polymerase chain reaction. *Environ Microbiol*. 2014;16:689–700.
64. Hallam SJ, Mincer TJ, Schleper C, Preston CM, Roberts K, Richardson PM, et al. Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *PLoS Biol*. 2006;4:520–36.
65. Beman JM, Popp BN, Francis CA. Molecular and biogeochemical evidence for ammonia oxidation by marine Crenarchaeota in the Gulf of California. *ISME J*. 2008;2:429–41.

66. Santoro AE, Casciotti KL. Enrichment and characterization of ammonia-oxidizing archaea from the open ocean: Phylogeny, physiology and stable isotope fractionation. *ISME J.* 2011;5:1796–808.
67. Sintez E, Bergauer K, De Corte D, Yokokawa T, Herndl GJ. Archaeal amoA gene diversity points to distinct biogeography of ammonia-oxidizing Crenarchaeota in the ocean. *Environ Microbiol.* 2013;15:1647–58.
68. Luo H, Tolar BB, Swan BK, Zhang CL, Stepanauskas R, Ann Moran M, et al. Single-cell genomics shedding light on marine Thaumarchaeota diversification. *ISME J.* 2014;8:732–6.
69. Tolar BB, Ross MJ, Wallsgrove NJ, Liu Q, Aluwihare LI, Popp BN, et al. Contribution of ammonia oxidation to chemoautotrophy in Antarctic coastal waters. *ISME J.* 2016;10:2605–19.
70. Richter M, Rossello-Mora R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci.* 2009;106:19126–31.
71. Chan JZM, Halachev MR, Loman NJ, Constantinidou C, Pallen MJ. Defining bacterial species in the genomic era: Insights from the genus *Acinetobacter*. *BMC Microbiol.* 2012;12:1.
72. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol.* 2007;57:81–91.
73. Moore WEC, Stackebrandt E, Kandler O, Colwell RR, Krichevsky MI, Truper HG, et al. Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *Int J Syst Evol Microbiol.* 1987;37:463–4.
74. Jung MY, Islam MA, Gwak JH, Kim JG, Rhee SK. *Nitrosarchaeum koreense* gen. nov., sp. nov., an aerobic and mesophilic, ammonia-oxidizing archaeon member of the phylum thaumarchaeota isolated from agricultural soil. *Int J Syst Evol Microbiol.* 2018;68:3084–95.
75. Doxey AC, Kurtz DA, Lynch MDJ, Sauder LA, Neufeld JD. Aquatic metagenomes implicate Thaumarchaeota in global cobalamin production. *ISME J.* 2015;9:461–71.
76. Otte J, Mall A, Schubert DM, Kizilman M, Berg IA. Malonic semialdehyde reductase from the archaeon *Nitrosopumilus maritimus* is involved in the autotrophic 3-hydroxypropionate/4-hydroxybutyrate cycle. *Appl Environ Microbiol.* 2015;81:1700–7.
77. Hansman RL, Griffin S, Watson JT, Druffel ERM, Ingalls AE, Pearson A, et al. The radiocarbon signature of microorganisms in the mesopelagic ocean. *Proc Natl Acad Sci.* 2009;106:6513–8.
78. Richey JE, Melack JM, Aufdenkampe AK, Ballester VM, Hess LL. Outgassing from Amazonia rivers and wetlands as a large tropical source of atmospheric CO₂. *Nature.* 2002;416:617–20.
79. Jia Z, Conrad R. Bacteria rather than Archaea dominate microbial ammonia oxidation in an agricultural soil. *Environ Microbiol.* 2009;11:1658–71.
80. Kobayashi S, Hira D, Yoshida K, Toyofuku M, Shida Y, Ogasawara W, et al. Nitric Oxide Production from Nitrite Reduction and Hydroxylamine Oxidation by Copper-containing Dissimilatory Nitrite Reductase (NirK) from the Aerobic Ammonia-oxidizing Archaeon, *Nitrososphaera viennensis*. *Microbes Environ.* 2018;33:428–34.
81. Spang A, Poehlein A, Offre P, Zumbrägel S, Haider S, Rychlik N, et al. The genome of the ammonia-oxidizing candidate *Nitrososphaera gargensis*: Insights into metabolic versatility and environmental adaptations. *Environ Microbiol.* 2012;14:3122–45.
82. Kerou M, Offre P, Valledor L, Abby SS, Melcher M, Nagler M, et al. Proteomics and comparative genomics of *Nitrososphaera viennensis* reveal the core genome and adaptations of archaeal ammonia oxidizers. *Proc Natl Acad Sci.* 2016;113:E7937–46.
83. Purwantini E, Mukhopadhyay B. Conversion of NO₂ to NO by reduced coenzyme F420 protects mycobacteria from nitrosative damage. *Proc Natl Acad Sci.* 2009;106:6333–8.
84. Bruce T, Meirelles PM, Garcia G, Paranhos R, Rezende CE, de Moura RL, et al. Abrolhos bank reef health evaluated by means of water quality, microbial diversity, benthic cover, and fish biomass data. *PLoS ONE.* 2012;7:e36687.
85. Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F. Laboratory procedures to generate viral metagenomes. *Nat Protoc.* 2009;4:470–83.
86. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics.* 2012;28:1420–8.
87. Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC, et al. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science.* 2012;337:1661–5.
88. Raveh-Sadka T, Thomas BC, Singh A, Firek B, Brooks B, Castelle CJ, et al. Gut bacteria are rarely shared by co-hospitalized premature infants, regardless of necrotizing enterocolitis development. *eLife.* 2015;2015:1–25.
89. Probst AJ, Castelle CJ, Singh A, Brown CT, Anantharaman K, Sharon I, et al. Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO₂ concentrations. *Environ Microbiol.* 2017;19:459–74.
90. Chen IMA, Markowitz VM, Chu K, Palaniappan K, Szeto E, Pillay M, et al. IMG/M: Integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.* 2017;45:D507–16.
91. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, et al. The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v4). *Stand Genomic Sci.* 2015;10:86.
92. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 2007;35(SUPPL.2):182–5.
93. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZY) in 2013. *Nucleic Acids Res.* 2014;42:490–5.
94. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 2014;12:59–60.
95. Development Core Team R. R: A Language and Environment for Statistical Computing. 2011.
96. Lebedeva EV, Hatzepichler R, Pelletier E, Schuster N, Hauzmayer S, Bulaev A, et al. Enrichment and genome sequence of the group I.1a ammonia-oxidizing archaeon “*Ca. Nitrosotenuis uzonensis*” representing a clade globally distributed in thermal habitats. *PLoS ONE.* 2013;8:1–12.
97. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol.* 2013;30:772–80.
98. Guindon S, Delsuc F, Dufayard J-F, Gascuel O. Estimating Maximum Likelihood Phylogenies with PhyML. In: Posada D, editor. *Bioinformatics for DNA Sequence Analysis*. Totowa: Humana Press; 2009. p. 113–37.
99. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavromatis K, Kyrpides NC, et al. Microbial species delineation using whole genome sequences. *Nucleic Acids Res.* 2015;43:6761–71.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

