



Genome-Resolved Metagenomic Insights into Massive Seasonal Ammonia-Oxidizing Archaea Blooms in San Francisco Bay

 Anna N. Rasmussen,^a  Christopher A. Francis^a

^aDepartment of Earth System Science, Stanford University, Stanford, California, USA

ABSTRACT Ammonia-oxidizing archaea (AOA) are key for the transformation of ammonia to oxidized forms of nitrogen in aquatic environments around the globe, including nutrient-rich coastal and estuarine waters such as San Francisco Bay (SFB). Using metagenomics and 16S rRNA gene amplicon libraries, we found that AOA are more abundant than ammonia-oxidizing bacteria (AOB) and nitrite-oxidizing bacteria (NOB), except in the freshwater stations in SFB. In South SFB, we observed recurrent AOA blooms of “*Candidatus Nitrosomarinus catalina*” SPOT01-like organisms, which account for over 20% of 16S rRNA gene amplicons in both surface and bottom waters and co-occur with weeks of high nitrite concentrations ($>10 \mu\text{M}$) in the oxic water column. We observed pronounced nitrite peaks occurring in the autumn for 7 of the last 9 years (2012 to 2020), suggesting that seasonal AOA blooms are common in South SFB. We recovered two high-quality AOA metagenome-assembled genomes (MAGs), including a *Nitrosomarinus*-like genome from the South SFB bloom and another *Nitrosopumilus* genome originating from Suisun Bay in North SFB. Both MAGs cluster with genomes from other estuarine/coastal sites. Analysis of *Nitrosomarinus*-like genomes show that they are streamlined, with low GC content and high coding density, and harbor urease genes. Our findings support the unique niche of *Nitrosomarinus*-like organisms which dominate coastal/estuarine waters and provide insights into recurring AOA blooms in SFB.

IMPORTANCE Ammonia-oxidizing archaea (AOA) carry out key transformations of ammonia in estuarine systems such as San Francisco Bay (SFB)—the largest estuary on the west coast of North America—and play a significant role in both local and global nitrogen cycling. Using metagenomics and 16S rRNA gene amplicon libraries, we document a massive, recurrent AOA bloom in South SFB that co-occurs with months of high nitrite concentrations in the oxic water column. Our study is the first to generate metagenome-assembled genomes (MAGs) from SFB, and through this process we recovered two high-quality AOA MAGs, one of which originated from bloom samples. These AOA MAGs yield new insight into the *Nitrosopumilus* and *Nitrosomarinus*-like lineages and their potential niches in coastal and estuarine systems. *Nitrosomarinus*-like AOA are abundant in coastal regions around the globe, and we highlight the common occurrence of urease genes, low GC content, and range of salinity tolerances within this lineage.

KEYWORDS estuary, ammonia-oxidizing archaea, *Thaumarchaeota*, nitrification, bloom, pelagic, metagenome-assembled genome

Nitrogen is an essential element for all life on Earth, but due to human activities, it has also become a major pollutant in estuarine and coastal systems (1). San Francisco Bay (SFB) is the largest estuary on the west coast of North America and is highly N polluted from agricultural and urban runoff, as well as large inputs from wastewater treatment plants (WWTPs). SFB consists of two distinct but connected arms, generally referred to as North and South SFB, and 5 major subembayments, including Suisun Bay, San Pablo Bay, Central

Editor Rachel Poretsky, University of Illinois at Chicago

Copyright © 2022 Rasmussen and Francis. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Christopher A. Francis, caf@stanford.edu.

The authors declare no conflict of interest.

Received 22 October 2021

Accepted 14 December 2021

Published 25 January 2022

Bay, South Bay, and Lower South Bay, each with different sources and forms of dissolved inorganic nitrogen (DIN) inputs (2). Wastewater treatment plants discharge high levels of ammonia into SFB, especially in South Bay (2, 3). This excess ammonia has many potential fates once it reaches SFB waters, including consumption/assimilation by microorganisms as well as oxidation to nitrite and ultimately nitrate via the chemoautotrophic process of nitrification. Nitrification generally occurs in two steps, ammonia and then nitrite oxidation. The organisms responsible for the first step, aerobic ammonia oxidation, include ammonia-oxidizing archaea (AOA) and ammonia-oxidizing bacteria (AOB), while the second step is carried out by nitrite-oxidizing bacteria (NOB). Comammox bacteria are capable of carrying out both ammonia and nitrite oxidation.

Understanding the microorganisms responsible for the transformation of ammonia to nitrate in the SFB water column has implications for primary production and potential N loss. High ammonia concentrations can lead to eutrophication in estuaries (4), though currently, phytoplankton dynamics in SFB are strongly influenced by light limitation (5–7) and grazing (8, 9), leading to a high-nutrient, low-chlorophyll state in parts of SFB. However, the form of DIN could also be one of several factors influencing phytoplankton dynamics and composition in SFB (10–16), potentially linking rates of nitrification to phytoplankton community composition. Nitrate produced by nitrification can also be denitrified in suboxic/anoxic sediments of the bay (17, 18) and lost from the system as N gases (e.g., N_2 and N_2O). While nitrification rate measurements are limited for SFB, studies have found rates to vary seasonally, geographically, and by water column depth (19, 20). For example, rates are highest in ammonia-rich waters near the Sacramento regional water treatment plant (20), tend to be higher in bottom waters than in surface waters (19), and are linked to suspended particulate matter (SPM) dynamics (19).

Despite the importance of nitrification to N cycling in SFB, only a few studies have investigated the ecology of benthic nitrifiers in SFB (21, 22), and nitrifiers in the water column remain virtually unstudied. Several AOA have been successfully enriched, isolated, and sequenced from North SFB sediments, yielding insights into low-salinity and brackish sediment organisms, including the genetic capacity for mechanosensitive channels for osmotic regulation and motility (23–26). Clone library and qPCR-based studies based on *amoA* genes (encoding ammonia monooxygenase subunit A) have had mixed results, with either AOB or AOA being numerically dominant in low-salinity sediments of North SFB (21, 22) and AOB being more abundant in higher-salinity sediments of the bay (21). Both AOB and AOA have been shown to be more abundant than the other depending on the estuary (27–34), though several studies have found AOA to be dominant in nutrient-enriched estuarine waters (29, 35). Of particular relevance to this study, seasonal AOA blooms and the decoupling of ammonia and nitrite oxidation have been observed in estuarine waters around the globe (36, 37); however, such blooms have yet to be examined using genome-resolved metagenomics. In this study, we used genomic binning and analysis approaches to assess nitrifier populations and generate the first metagenome-assembled genomes (MAGs) from SFB.

MAGs have greatly expanded our ability to examine the metabolic capacity of organisms from a wide variety of environments. For example, MAGs have already increased our understanding of different niches within the AOA (38, 39) and allow us to study organisms that have yet to be enriched or isolated in the laboratory and/or that may have been missed by previous primer-based studies. The discovery of an increasing number of AOA MAGs allows us to compare a large number of otherwise undiscoverable genomes to gain insight into organisms from the estuarine environment.

A previous 16S rRNA gene amplicon-based study of the SFB water column revealed a high abundance of *Nitrosopumilus*-like organisms in South SFB along with high nitrite concentrations (40). Here, we delved much deeper into this initial finding by including metagenomic sequencing, as well as additional 16S rRNA gene amplicon libraries analyzed via amplicon sequence variants (ASVs) versus a 97% operational taxonomic unit (OTU)-based approach. We assessed both surface and bottom water amplicon libraries from South Bay. We recovered two high-quality MAGs (>97% complete and <1%

contamination), including a “*Candidatus Nitrosomarinus catalina*” SPOT01-like population genome and a *Nitrosopumilus* lineage. We compared the genomic characteristics of our MAGs to AOA genomes from other estuarine and coastal environments and leveraged our 16S rRNA gene time series to better understand the spatial and temporal distribution of temporally abundant AOA in SFB.

RESULTS AND DISCUSSION

We recovered two high-quality AOA MAGs from SFB after dereplicating 21 *Nitrosopumilus*-like MAGs (>70% complete, <5% contamination) (Table S1) at 98% average nucleotide identity (ANI) (Fig. S1). The two representative MAGs are designated SFB_27D_13Oct24_05_ms_bin_1 (97.6% complete, 0.0% contamination) and SFB_3D_13Oct25_100_mh_bin_18 (100.0% complete, 0.97% contamination), referred to here as SFB_27_05_bin1 and SFB_3_bin18. Both of these MAGs have 16S rRNA and *amoA* genes. Additionally, a different MAG, SFB_27D_13Oct24_03_ms_bin1, was used in place of SFB_27_05_bin1 in the concatenated ribosomal tree because several ribosomal genes are missing from SFB_27_05_bin1. According to the GTDB-tk classification, SFB_27_05_bin1 is classified as “*Nitrosopumilus catalinensis*” with an ANI of 99.23% to the “*Ca. Nitrosomarinus catalina*” SPOT01 (GCF_002156965.1), which was enriched from the water column in the San Pedro Channel near Santa Catalina Island off the coast of Los Angeles, CA, USA (41). SFB_3_bin18 is only classified to the genus level of *Nitrosopumilus* and most closely related to “*Nitrosopumilus* sp006740685,” with an ANI of 85.26% to *Nitrosopumilus* sp. strain SW (GCF_006740685.1), which was isolated from coastal surface waters (20 m) of the Yellow Sea on the west coast of the Korean peninsula (42). This MAG likely represents a new species based on the GTDB-tk classification. Information on the quality of all AOA MAGs generated in this study is available in Table S1. No other nitrifier MAGs were recovered.

South Bay AOA MAG and ASV abundance correlated with high nitrite concentrations. We aligned the 16S rRNA V4-V5 region from our two representative AOA MAGs with ASVs from our 16S rRNA gene amplicon library to assess spatiotemporal distribution of these representative organisms and gain further insight into their environmental distribution. Both metagenomic and 16S rRNA gene ASV analysis revealed a highly abundant “*Ca. Nitrosomarinus catalina*”-like lineage. SFB_27_05_bin1 recruits 6.0% of metagenomic reads from station 27 and has a maximum of 329 RPKG (reads recruited per kilobase of MAG per gigabase of metagenome) at this station (Fig. 1). ASV 8 (the 8th most abundant ASV) in our 16S rRNA gene amplicon data set is identical to the 16S V4-V5 region of SFB_27_05_bin1, indicating that this MAG represents the same population of AOA observed in the amplicon libraries (Table S2). We observed a pronounced recurring bloom in South Bay of ASV 8 in the fall of 2012 and 2013 in both bottom and surface waters (Fig. 1 and 2), when salinities are generally over 30 practical salinity units (PSU) and temperatures are between 16 and 20°C (Fig. S2A). The abundance of ASV 8 is strongly correlated with nitrite concentrations in South Bay in both surface ($r^2 = 0.97$, $P < 0.001$) and bottom ($r^2 = 0.92$, $P < 0.001$) waters (Fig. 2). Ammonia concentrations are not significantly correlated with ASV 8 abundance ($r^2 = 0.06$, $P > 0.05$). The abundance of SFB_27_05_bin1 is also correlated with nitrite concentrations ($r^2 = 0.94$, $P < 0.001$). The abundance of AOA (measured via quantitative PCR [qPCR] analysis of marine group I [MGI] *Thaumarchaeota* 16S rRNA genes) was also correlated with nitrite concentrations ($r^2 = 0.88$, $P < 0.001$) and with the relative abundance of AOA in 16S rRNA amplicon data ($r^2 = 0.78$, $P < 0.001$). Based on qPCR data, AOA abundance increased 2 orders of magnitude from nonbloom to bloom samples (Fig. 2). Nutrient data collected by the U.S. Geological Survey (USGS) show that high surface nitrite concentrations in South Bay have recurred in 7 of the last 9 years (2012 through 2020), generally peaking in October through December and reaching almost 12 μM in some years (Fig. S2B). The high AOA abundance in both years of 16S rRNA gene amplicon data and qPCR data (2012 and 2013) and strong correlation to nitrite concentrations suggest that AOA blooms are a recurrent feature in South Bay during the midfall to late fall, leading to this semiregular nitrite accumulation (Fig. 2).

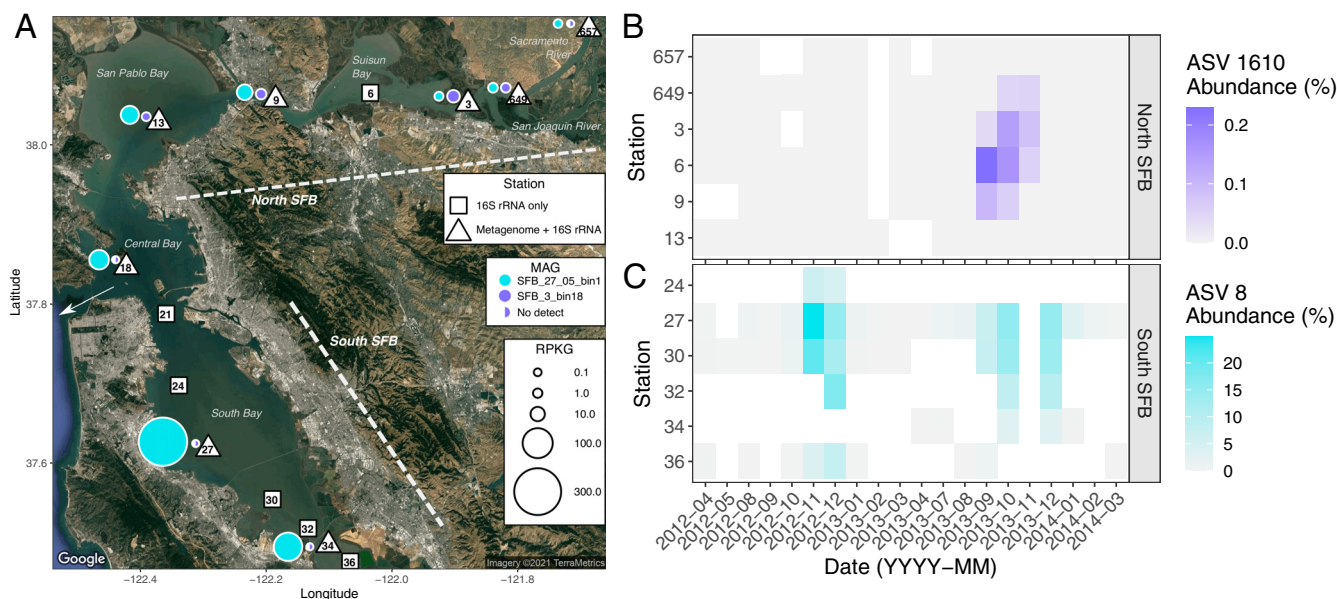


FIG 1 (A) Metagenome reads recruited to two high-quality representative AOA MAGs from SFB. White triangles or squares indicate station location for samples with 16S rRNA gene amplicons (squares) plus metagenomes (triangles). Circle size represents reads per kilobase of genome per gigabase of metagenome (RPKG) recruited to the respective MAG, which is indicated by circle color. SFB_3_bin18 was not detected in 4 samples, represented by half-filled circles. The white arrow indicates where SFB connects to the Pacific Ocean. (B and C) Tile plots indicate the relative abundance in 16S rRNA amplicon libraries of (B) ASV 1610 and (C) ASV 8, which correspond to the 16S rRNA gene of MAGs SFB_3_bin18 and SFB_27_05_bin1, respectively, over a 2-year monthly time series (April 2012 to March 2014). Panel B includes only bottom water samples, while panel C includes the mean abundance of surface and bottom water samples.

During this time of high AOA abundance, we found nitrite-oxidizing bacteria (NOB) to generally be 2 orders of magnitude less abundant than AOA, reaching a maximum relative abundance of 0.7% in 16S rRNA gene amplicon libraries in South SFB in December 2013 (after the onset of the AOA bloom in October 2013 [Fig. S3]). NOB detected in SFB belonged to the genera *Nitrospira*, *Nitrospina*, and LS-NOB (*Nitrospinae* clade 2 [43]). Despite high dissolved inorganic nitrogen (DIN) concentrations in SFB waters, no NOB from the phylum *Proteobacteria* were detected (e.g., "*Candidatus Nitrotoga*," *Nitrolancea*, *Nitrobacter*, or *Nitrococcus*). *Nitrospira* organisms were found throughout SFB, from low-salinity sites in the North SFB to near-marine-salinity sites in South SFB, while *Nitrospina* organisms were found in higher-salinity sites and had the highest abundances in South SFB. LS-NOB are found around the globe in temperate and tropical ocean waters and can be numerically dominant; within SFB, LS-NOB were the dominant NOB at the most marine-influenced site, station 18 (Fig. S3). Ammonia-oxidizing bacteria (AOB) consisted solely of *Nitrosomonas* and were of similarly low abundance to NOB, with a maximum of 0.2% in 16S rRNA gene amplicon libraries. In general, NOB and AOB had much lower abundance than AOA and had abundances only comparable to those of AOA in freshwater stations where all three nitrifier groups were of low relative abundance (Fig. S3).

High nitrite concentrations due to the decoupling of ammonia and nitrite oxidation in the water column have been reported in many other estuaries and coastal sites (36, 37, 44–47). Although low oxygen concentrations can contribute to the decoupling of ammonia and nitrite oxidation in natural and engineered systems (46–51), a meta-analysis of 29 estuaries and lagoons found that transient nitrite accumulation was driven primarily by higher temperatures rather than hypoxia (36). Schaefer and Hollibaugh (36) found peak nitrite occurring in late summer into early fall, driven by warmer temperatures between 20 and 30°C. AOA blooms have been observed in estuaries and coastal bays in the late summer (37), fall (40, 52), and winter (53) and after wind events that cause vertical mixing, bringing oxygenated waters into deeper, ammonium-rich waters (46). In another example, in two bays of the Yellow Sea, high *Nitrosopumilus*-like OTU abundance was concomitant with euryarchaeal blooms in October, suggesting that

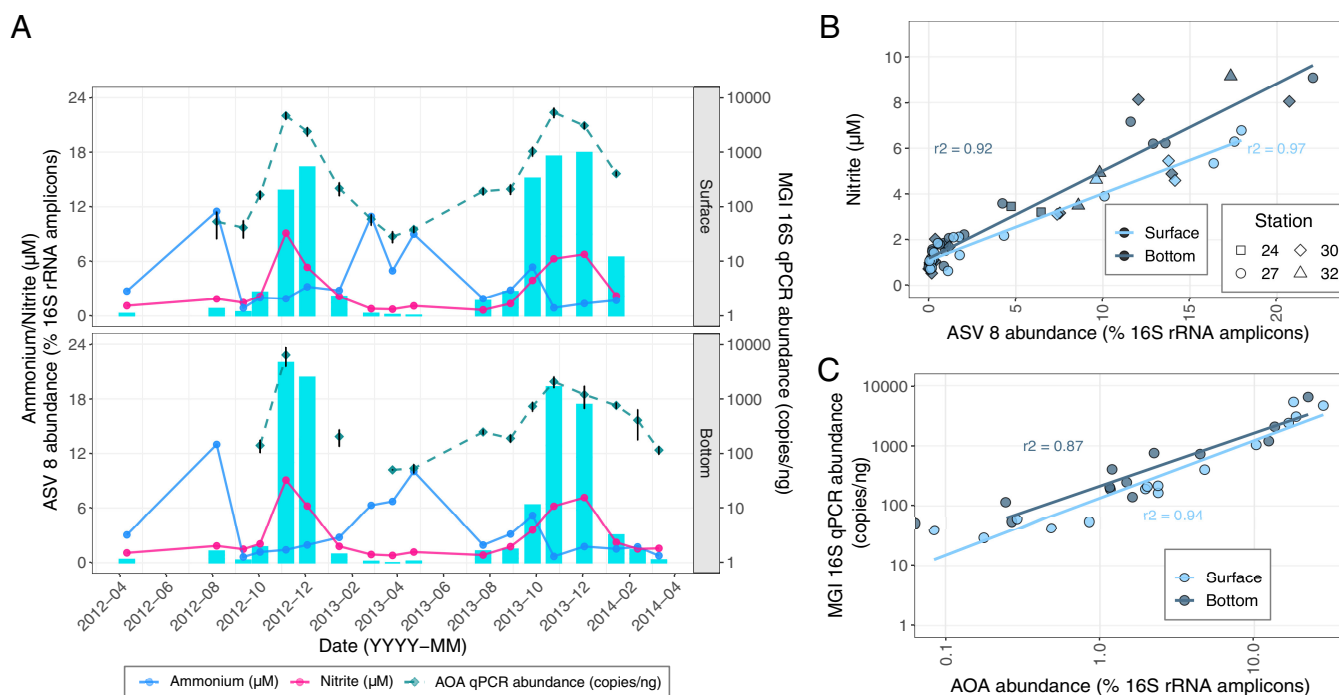


FIG 2 Abundance of ASV 8 in 16S rRNA gene amplicon libraries corresponds to temporal patterns in nitrite and ammonia concentrations in South Bay in both surface water (2 m) and bottom water (1 m above the estuary floor). (A) ASV 8 abundance (bars) at station 27 along with nitrite and ammonia concentrations and MGI *Thaumarchaeota* 16S rRNA gene qPCR abundance over the course of a 2-year approximately monthly time series (April 2012 to March 2014). qPCR values are on a log scale. (B) Nitrite concentrations versus ASV 8 abundance in several South Bay stations, including 24, 27, 30, and 32, is highly correlated in both surface ($r^2 = 0.97$) and bottom ($r^2 = 0.92$) waters. (C) AOA abundance measured via qPCR versus the relative abundance of AOA in 16S rRNA gene amplicon libraries is correlated in surface ($r^2 = 0.94$) and bottom ($r^2 = 0.87$) waters. Both axes are on a log scale.

eurycarchaeal heterotrophy could provide ammonia for ammonia oxidation (52). During this bloom in the Yellow Sea, *Nitrosopumilus*-like OTUs reached over 30% relative abundance and were correlated with nitrite concentrations, which reached $4 \mu\text{M}$ (52).

The decoupling of ammonia and nitrite oxidation in SFB does not obviously follow patterns observed in other estuaries. For example, in seasonally stratified systems, such as the Chesapeake Bay, transient wind events have led to nitrite accumulation when ammonia-rich and oxygen-poor waters are mixed, allowing significant ammonia but not nitrite oxidation (46). In contrast, SFB waters (especially in the channel where our samples were taken) are oxic and well mixed (Fig. S2A), suggesting that dissolved oxygen (DO) is not likely to be an explanatory factor. The AOA bloom in South SFB also occurs in fall, after peak summer temperatures have occurred (Fig. S2A), in sharp contrast to what has been observed in a global analysis of nitrite accumulation in estuaries (36). It is worth highlighting that “*Ca. Nitrosomarinus catalina*” SPOT01 has a lower optimal growth rate and outperforms other cultured marine *Thaumarchaeota* strains at temperatures of $<20^\circ\text{C}$ (41). Thus, the decoupling of ammonia and nitrite oxidation in SFB is not likely caused by low DO or warm temperatures; however, cooler temperatures could potentially lead to differences between AOA and NOB activity.

One potentially noteworthy environmental factor in SFB is the relatively low turbidity during the bloom period (Fig. S2A), which generally leads to greater light penetration (Fig. S2C). The relative photosensitivity of AOA versus AOB versus NOB varies and has been documented in the ocean, estuaries, and WWTPs (54–59). Although cultivated AOA and NOB can be more sensitive than AOB to photoinhibition (55, 59), some oceanic AOA are still active under high irradiance (60). Ammonia oxidation has also been shown to be less sensitive to light than nitrite oxidation in the sunlit upper ocean (where AOA predominate over AOB), contributing to the formation of primary nitrite maxima (61–63). Similarly, it is indeed possible that NOB in SFB are more photosensitive than the AOA lineage that blooms, leading to their decoupled abundance and

activity. At station 27, the relative abundance of both *Nitrospina* and *Nitrospira* is positively correlated with the natural log of SPM concentrations ($r^2 = 0.59$ and $r^2 = 0.70$, respectively; $P < 0.001$), which could be related to photoinhibition or particle association. Overall, the potential role of differential light inhibition during the AOA bloom is intriguing and warrants further physiological studies of relevant AOA and NOB. Finally, we also did not observe a concomitant euryarchaeal bloom with the AOA bloom, as observed in the Yellow Sea (52), suggesting that enhanced euryarchaeal heterotrophy is not a source of additional ammonia for AOA. A co-occurrence network identified 11 ASVs with associations with bloom-forming ASV 8, none of which belonged to the *Euryarchaea* or NOB. Of the 11 ASVs, we highlight three ASVs with peak abundances occurring during the AOA bloom belonging to *Bacteroidota*, *Dadabacteria*, and the NS11-12 marine group (*Sphingobacteriales*) (Fig. S4). The associations of these ASVs could indicate possible positive interactions during the bloom or similar ecophysiological responses to seasonal conditions.

Given that SPOT01-like organisms have been found to be dominant not only during the AOA bloom in SFB but in estuarine and coastal sites across the globe in 16S rRNA, *amoA*, and metagenomic studies (31, 38, 41, 64, 65), key environmental parameters in addition to cooler temperatures, as mentioned above, likely influence when and where SPOT01-like AOA thrive. Salinity could also be an important influence on the dominance of SPOT01-like organisms. In the Jiulong River estuary (JRE), a SPOT01-like *amoA* OTU is dominant at salinities ranging from 15.8 to 27.0 PSU (65); in the Pearl River estuary (PRE), SPOT01-like MAGs are also most abundant in saltier stations (64); and SPOT01 dominates at the San Pedro Ocean time series (SPOT) off the coast of California in September and October (41). These findings indicate that SPOT01-like organisms thrive in near-marine salinities, such as those found in South SFB, and in the cooler temperatures of fall. Residence time could also influence the ability of AOA to bloom. In the PRE, high nitrification rates are correlated with longer residence times (65). South SFB is expected to have long residence times in summer and fall, on the order of months (66). Long residence times could allow populations to grow and consume excess ammonia from WWTPs. However, why NOB are not also favored during this time, especially with weeks to months of high nitrite concentrations, is unclear yet highly intriguing.

SFB_3_bin18 has peak abundance in brackish salinities in North SFB. SFB_3_bin18 has a peak abundance at station 3, recruiting 0.12% of metagenome reads and having a maximum relative abundance of 5.8 RPKG. ASV 1610 is identical to the 16S rRNA V4-V5 region of SFB_3_bin18 (Table S2). We observed a peak abundance of ASV 1610 in Suisun Bay (stations 3 and 6) in the late summer/early fall in salinities ranging from 8 to 12 PSU. Unlike in South Bay, where we saw high AOA relative abundance (>20%), ASV 1610 accounts for a maximum of only 0.2% of the 16S rRNA gene amplicon library (Fig. 1), occurring in September 2013. There is no correlation between ASV 1610 and nitrite concentration ($r^2 = 0.0002$, $P > 0.05$), and nitrite concentrations are not elevated as in South Bay (Fig. S2B). Intriguingly, in 2014 and 2015 we observed high concentrations of nitrite in the late summer and fall in Suisun Bay (Fig. S2B), which could be evidence for an AOA bloom or some other type of decoupling of ammonia and nitrite oxidation. However, since this nitrite accumulation occurs in the warmer temperatures of late summer and early fall and at mesohaline salinities, the underlying cause(s) may be distinct from what we observed in South SFB. A co-occurrence network identified 10 ASVs associated with ASV 1610, 6 of which had similar seasonal patterns in abundance to ASV 1610 and belonged to the NS4 marine group (*Flavobacteriaceae*), IS-44 (*Nitrosomonadaceae*), MB11C04 marine group (*Verrucomicrobiae*), *Alphaproteobacteria*, *Methylophilaceae*, and clade III (SAR11) (Fig. S4). The co-occurrence of these ASVs and ASV 1610 could indicate potential interactions or could be related to similar ecophysologies. No NOB were associated with this AOA lineage, as has been observed for marine AOA and NOB ecotypes at different depths in Monterey Bay (67). Though the genus IS-44 falls into the *Nitrosomonadaceae*, it is unclear if organisms in this lineage are AOB.



FIG 3 Concatenated ribosomal protein tree from MAFFT alignment of 24 ribosomal proteins from 153 genomes. The tree was constructed using IQtree and model LG+R5+F with 1,000 bootstraps. *Aigarchaeota* was used as an outgroup. MAGs are in bold, and MAGs generated in this study have a yellow star beside them. Background colors highlight subgroups of interest. Node color indicates bootstrap support.

Similarity of SFB MAGs to other estuarine/coastal AOA. We compared the genomic content of the representative MAGs generated in this study to medium- to high-quality (>70% complete, <10% contamination) AOA genomes from NCBI and IMG, along with additional MAGs from other pelagic estuary and coastal samples, including those from the: North Sea (68), Baltic Sea (69), PRE (64), JRE (65), Gulf of Mexico (GoM) (70), Sapelo Island (71), Amazon River (72, 73), and Monterey Bay (MB). Based on concatenated ribosomal trees, we identified two “pelagic” estuarine clusters within the *Nitrosopumilus* genus (Fig. 3). SFB_27_05_bin1 clusters with MAGs from the North Sea, Mediterranean

TABLE 1 Summary statistics for the two estuarine AOA clusters (*Nitrosomarinus*-like and SCM1-like) in the pangenome

Pangenome cluster ^a	Avg GC content	Avg genome size (bp)	Avg no. of genes	Avg gene length (bp)	Avg no. of genes per kb	Avg no. of singletons
<i>Nitrosomarinus</i> -like	0.31	1,247,478	1,485	715	1.27	46.9
SCM1-like	0.34	1,559,476	1,762	740	1.23	77.0

^aIncludes only genomes that are >90% complete, which includes 21/39 genomes from the pangenome analysis.

Sea, PRE, JRE, and GoM (Fig. 3). SFB_3_bin18 clusters with MAGs from Sapelo Island, PRE, JRE, and GoM (Fig. 3). For simplicity, we refer to these two estuarine clusters as a “*Nitrosomarinus*-like” cluster and a *Nitrosopumilus* “SCM1-like” cluster based on well-known representative genomes from each cluster. The *amoA* phylogeny reveals the same clusters (Fig. S5).

We used a pangenomic workflow to compare a total of 39 AOA genomes, 23 and 16 from the *Nitrosomarinus*-like and SCM1-like clusters, respectively. This pangenome contained 5,140 gene clusters, of which 3,295 (64%) were annotated. In our comparison of genomes, we found that the *Nitrosomarinus*-like cluster had a relatively lower GC content (mean = 0.31) (Table 1) and smaller genome sizes (Fig. 4). In addition to this, we found that compared to the SCM1-like cluster, the *Nitrosomarinus*-like cluster was enriched in fewer genes (23 versus 44), possibly due to the smaller genome sizes. Most notably, a majority of *Nitrosomarinus*-like genomes contained urease and urease accessory genes, while most SCM1-like genomes did not (Fig. 5 and Fig. S6). While laboratory studies found that the SPOT01 strain could not grow on urea as a sole N source (41), urease may be of importance to other members of the clade. We also note that ectoine synthesis genes are missing in *Nitrosomarinus*-like organisms. Ectoine can function as an osmoprotectant in AOA (74) and the acquisition of ectoine synthesis genes may have been important for the expansion of AOA into the marine environment (75). Other genes enriched in *Nitrosomarinus*-like genomes include a sodium-proline symporter gene, *putP*, and a gene for an unknown transporter, *ynfA*.

Several coastal/estuarine genomes were classified as “*Nitrosopumilus catalinensis*” using GTDB (Table S3), and we refer to these as “SPOT01-like.” Organisms from this group can be transcriptionally and numerically dominant in meso- to euhaline estuarine and coastal sites (31, 38, 41). Organisms from eutrophic sites such as the GoM, JRE, and PRE cluster together

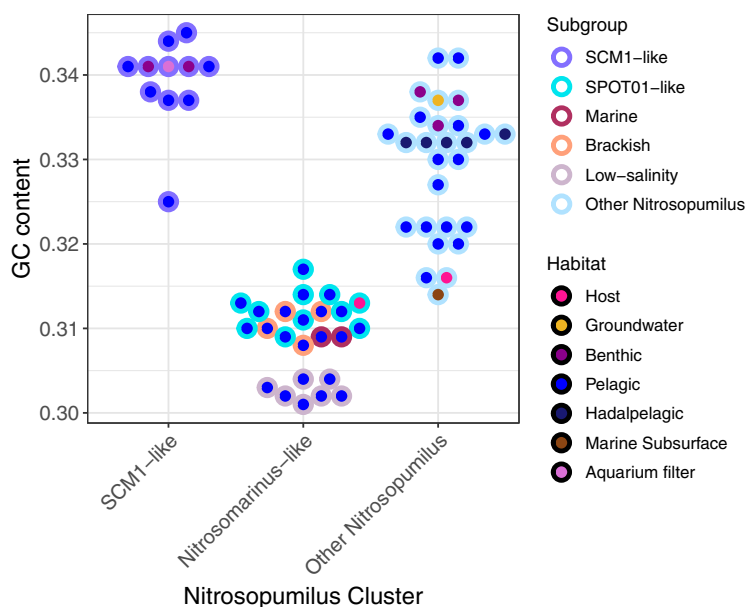


FIG 4 GC content for genomes in the SCM1-like, *Nitrosomarinus*-like, and other *Nitrosopumilus*-like clusters for genomes classified as *Nitrosopumilus* at the genus level using GTDB-tk and with >70% completeness and <5% contamination according to CheckM and GC content of >0.37 to remove higher-GC host-associated *Nitrosopumilus* genomes. Circle fill color indicates the habitat from which the genome originated, and circle border color indicates the subclade.

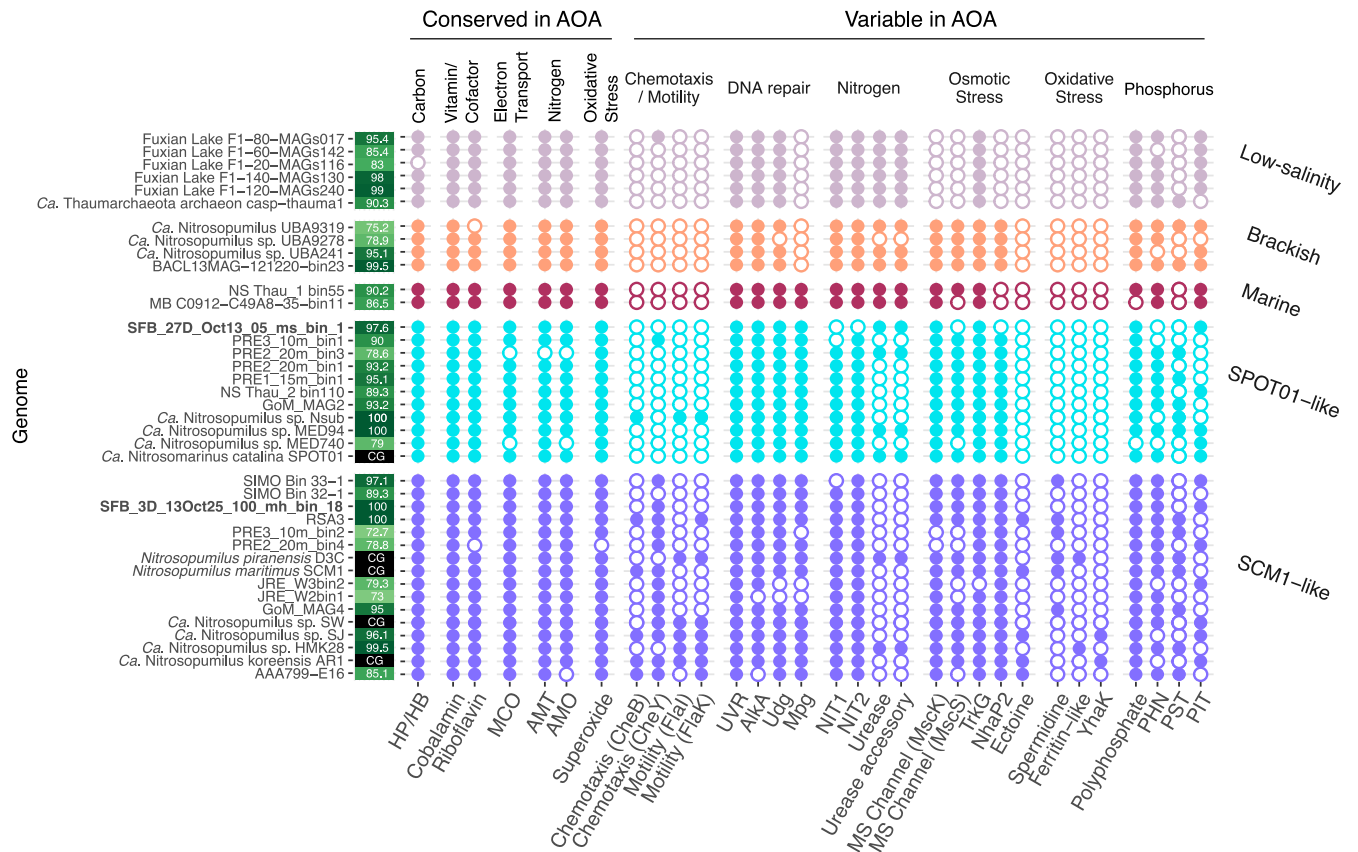


FIG 5 Gene presence or absence in pangenome based on COG, KO, or PFAM annotations. Solid circles represent gene presence, and open circles represent gene absence. Genome completeness is indicated in green squares next to genome names, while complete genomes are marked “CG” with a black background. MAGs generated in this study are in bold. Genes are clustered based on whether they are conserved or variable in AOA based on the work of Ren et al. (75) and Kerou et al. (108). Genes are further grouped by general function, labeled at the top of the plot. Genes displayed are as follows: HP/HB, Hcd (4-hydroxybutyryl-coenzyme A dehydrogenase); cobalamin, Cob5 (cobalamin synthase); riboflavin, RibB (3,4-dihydroxy-2-butanone 4-phosphate synthase); MCO, SufI (multicopper oxidase with three cupredoxin domains); AMT, AmtB (ammonia channel protein); AMO, *amoA* (ammonia monooxygenase subunit A); superoxide, SodA (superoxide dismutase); chemotaxis (CheB), CheB (chemotaxis response regulator); chemotaxis (CheY), CheY (chemotaxis response regulator receiver domain); motility (FlaI), archaeal flagellar protein; motility (FlaK), archaeal preflagellin peptidase; UVR, UvrA (UV-induced DNA lesion repair endonuclease ATPase subunit); Alka, Alka (DNA-3-methyladenine DNA glycosylase); Udg, Udg4/5 (uracil DNA glycosylase family 4/5); Mpg, 3-methyladenine DNA glycosylase; NIT1, YbeM (deaminated glutathione amidase); NIT2, YafV/Nit2 (nitrilase/omega-amidase); urease, UreC (urease alpha subunit); urease accessory, UreE (urease accessory protein); MS channel (MscK), small-conductance mechanosensitive channel; MS channel (MscS), small-conductance mechanosensitive channel; TrkG, TrkG (K⁺ transport system, membrane component); NhaP2, monovalent cation-proton antiporter; ectoine, EctD (ectoine hydroxylase); spermidine, SpeE (spermidine synthase); ferritin-like, Dps (DNA-binding ferritin-like protein); YhaK, YhaK (redox-sensitive bicupin protein); polyphosphate, Ppa (inorganic pyrophosphatase); PHN, PhnD (ABC-type phosphate/phosphonate transport system periplasmic component); PST, PstA (high-affinity ABC-type phosphate transport system permease component); PIT, PitA (low-affinity phosphate/sulfate permease).

(Fig. 3 and Fig. S7). PRE AOA genomes are enriched in heavy metal (e.g., manganese, zinc, and iron) transport genes as well as phosphate transport, indicating AOA are adapted to eutrophic conditions (64). Although PRE SPOT01-like MAGs were originally reported to not contain *ureC* (64), we recovered a *ureC* (along with all other urease subunits and accessory proteins) from PRE2_20m_bin3 (Fig. 5 and Fig. S5).

Within the *Nitrosomarinus*-like cluster, we found several clades of coastal/estuarine AOA found at marine salinities (referred to as “SPOT01-like” and “marine” in Fig. 3 and 5), a distinct cluster of organisms from the brackish waters of the Baltic Sea, and a cluster of organisms from freshwater to low-salinity sites (Fig. 3). The low-salinity organisms are missing genes related to osmoregulation (*mscK*, *mcsS*, and *nhaP2*) (Fig. 5) and have particularly low GC content (Fig. 4). These salinity “ecotypes” within the *Nitrosomarinus*-like cluster appear to be under streamlining pressure, given the small genome sizes and apparent loss of genes to deal with osmoregulation. It is intriguing that these seemingly stream-lined *Nitrosomarinus*-like organisms are dominant in estuarine and coastal waters, which are often nutrient rich, around the globe (38) and that they reach such high temporal abundances in SFB. While other marine AOA have streamlined genomes (41, 76), the

exceptionally low GC content and high coding density point to potentially different selection pressures on *Nitrosomarinus*-like organisms.

Genomes in the SCM1-like cluster were enriched in many genes of unknown or predicted function compared to the *Nitrosomarinus*-like genomes (Fig. S6), possibly due to their larger size. SCM1-like genomes generally contained a *cheY* receiver domain (related to chemotaxis), spermidine synthase, and several putative amino acid symporters (*gltP* and *potE*) (Fig. S6), while *Nitrosomarinus*-like genomes did not. While they were not significantly enriched, it is worth noting that other genes related to motility and archaeal flagellin are more abundant in the SCM1-like cluster of genomes than in *Nitrosomarinus*-like genomes (Fig. 5), although the SCM1 strain itself lacks archaeal flagellin genes. The only *Nitrosomarinus*-like organism with motility genes is host-associated *Nitrosopumilus* sp. strain Nsub. The GC content was low (mean = 0.34) for SCM1-like organisms but higher than for the *Nitrosomarinus* cluster (mean = 0.31). In general, metrics such as gene length and gene coding density point to even more streamlining in *Nitrosomarinus*-like than SCM1-like genomes (Table 1). The SCM1-like clade includes MAGs from the GoM, PRE, JRE, and Sapelo Island, as well as our North SFB MAG (Fig. 3; Fig. S5 and Table S1). Sapelo Island is known for having AOA blooms annually in the warm summer months, concurrent with nitrite accumulation and high nitrification rates (37, 77), highlighting that bloom organisms can come from both of the “pelagic” AOA clusters analyzed in our study. As observed in other studies, the form of phosphate transporter does not strictly follow phylogeny (38), and we note that most genomes have genes for either the high-affinity PST (PstA, a high-affinity ABC-type phosphate transport system permease component) transporter or the low affinity PIT (PitA, a low-affinity phosphate/sulfate permease) transporter (Fig. 5). The ranges of ANI between genomes are similar within both the *Nitrosomarinus*-like clade (0.83 to 0.99) and SCM1-like clade (0.77 to 0.99), as is the median ANI for each clade (0.85 and 0.86, respectively). The median ANI between the *Nitrosomarinus*-like and SCM1-like clades is 0.79 (Fig. S7).

Ahlgren et al. (41) designated the SPOT01 organism as its own genus, *Nitrosomarinus*, based on its 16S rRNA gene, concatenated core conserved genes, *amoA*, and *ureC* gene phylogeny. Qin et al. (38) used the range of ANI and amino acid identity (AAI) within *Nitrosopumilus* to suggest that these organisms fall within the genus *Nitrosopumilus* and should not be distinguished as *Nitrosomarinus*. It is difficult to assign a strict ANI cutoff for delineating a genus, and several methods exist for trying to define genera based solely on genomic information (78, 79); however, an argument can be made for also considering the phylogenetic, physiological, metabolic, and ecological context of a group of organisms when describing a genus. Our study examining MAGs from a variety of pelagic environments found that *Nitrosomarinus*-like genomes form their own separate phylogenetic cluster based on ribosomal and key marker genes, as previously reported (38, 41), but also that they have a distinct ecological niche, occurring primarily in the pelagic environment, having low GC content and even more streamlined genomes than other *Nitrosopumilus* species. The *Nitrosomarinus*-like subclade is enriched in urease genes compared to the SCM1-like subclade but is also enriched in fewer genes than SCM1-like organisms. Additionally, genomes within the *Nitrosomarinus*-like clade form distinct subclades that can be found along the salinity gradient, can occur in eutrophic environments, and can bloom in some estuaries. Several studies have also noted the predominance of *Nitrosomarinus*-like organisms in coastal and estuarine sites (31, 38, 41, 64, 65). Physiological experiments with SPOT01 showing a lower optimum growth temperature could help explain the propensity of blooms in cooler months and a distinct niche for these organisms to thrive in. Increased efforts to enrich and isolate organisms from this group are necessary to establish their potentially distinct metabolisms (i.e., urea consumption) and physiology (i.e., cooler growth optimum). Defining a new genus generally requires multiple lines of evidence; however, *Nitrosomarinus*-like organisms may very well be different enough from other *Nitrosopumilus* organisms to merit this distinction.

Blooms of these AOA in South SFB has important implications for N cycling and ecosystem health. South Bay is characterized by high ammonia inputs from large WWTPs,

but urea measurements are currently lacking in SFB. Urea could be an important source of ammonia for nitrification. While the SPOT01 strain could not grow on urea as a sole N source, further studies are required to see what role urea could play for the AOA population that blooms in SFB, represented by SFB_27_05_bin1, and other organisms in the *Nitrosomarinus*-like clade. Furthermore, what environmental conditions allow these AOA to form massive blooms while preventing NOB from following suit when nitrite concentrations are high in SFB for several weeks is a critical open question.

Conclusions. We recovered two high-quality AOA genomes from the waters of SFB, one of which corresponds to a bloom AOA population in South SFB associated with significant nitrite accumulation. These MAGs are related to other coastal/estuarine AOA MAGs falling into a lower-GC *Nitrosomarinus*-like cluster as well as a *Nitrosopumilus* SCM1-like cluster. SFB_3_bin18 likely represents a distinct species based on its GTDB classification, while SFB_27_05_bin1 is classified as “*Ca. Nitrosomarinus catalina*.” The conditions that allow the SFB_27_05_bin1 lineage of AOA to bloom require further investigation; temperature, oxygen, salinity, turbidity, residence time, urea concentrations, or competition with phytoplankton could all play a role in bloom formation. Furthermore, genomes within the *Nitrosomarinus*-like cluster originate from pelagic sites ranging from fully marine salinities to middle to low salinities, have low GC content and high coding density, often contain urease genes, form a distinct phylogenetic cluster, and are abundant primarily in coastal environments. These findings support the idea that *Nitrosomarinus*-like organisms have a niche distinct from that of other *Nitrosopumilus*-like organisms.

MATERIALS AND METHODS

Sample collection. Microbial biomass was collected for DNA extraction approximately monthly between April 2012 and March 2014 during United States Geological Survey (USGS) water quality monitoring cruises in the channel of the SFB estuary onboard the R/V *Polaris*. Microbial cells were collected from surface waters (2 m) and bottom waters (1 m above the estuary floor) by pressure filtering 150 to 1,000 mL of water from conductivity, temperature, and depth (CTD) instrument package casts through a 10- μ m-pore-size polycarbonate Isopore membrane filter (47-mm diameter; EMD Millipore, Darmstadt, Germany) in line with a 0.22- μ m polyethersulfone Supor-200 membrane filter (47-mm diameter; Pall, Port Washington, NY), followed by flash freezing of 0.22- μ m filters in liquid nitrogen prior to storage at -80°C .

Environmental data. Corresponding water quality data from the exact sampling cruises, stations, and water depths used in this study were downloaded from the USGS Water Quality of SFB database (80). Additional ammonium, nitrate, and nitrite measurements were performed on water samples using filtered (0.22 μ m pore size) water that was frozen on dry ice prior to storage at -20°C . Ammonium was measured using the salicylate-hypochlorite method (81). Nitrate and nitrite were measured using a SmartChem200 discrete analyzer (Unity Scientific, Brookfield, CT) following standard manufacturer's operating procedures. Nutrients were measured within 1 week of sample collection.

DNA extraction. DNA was extracted with the FastDNA spin kit for soil (MP Biomedicals, Santa Ana, CA), following the manufacturer's instructions, with the following modifications: filters were homogenized in bead tubes for 40 s at speed 6.0 in a FastPrep bead beater (MP Biomedicals, Santa Ana, CA), and final DNA was eluted in 75 μ l of 55 $^{\circ}\text{C}$ sterile DNase-free water. DNA was quantified using the Qubit double-stranded DNA (dsDNA) broad-range assay (Life Technologies, Grand Island, NY) and stored at -80°C .

16S rRNA gene amplicon sequencing and processing. 16S rRNA gene amplicon library preparation and sequencing were conducted through a Department of Energy Joint Genome Institute (JGI) Community Science Program (CSP) project and through the Georgia Genomics and Bioinformatics Core (GGBC) at the University of Georgia. In total, 177 bottom water samples and 20 surface samples were amplified with the 16S V4-V5 primers (82) (515F-Y, GTGYCAGCMGCCGCGGTAA; 926R, CCGYCAATYMTTTRAGTTT). Library preparation was performed by JGI and GGBC following their standard operating procedures. The JGI protocol is available at <https://jgi.doe.gov/user-programs/pmo-overview/protocols-sample-preparation-information/>. GGBC uses an annealing temperature of 66 $^{\circ}\text{C}$ and follows the Illumina sample preparation guide: https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf. Samples were pooled and sequenced on an Illumina MiSeq instrument.

Cutadapt (v3.1) (83) was used to remove primer sequences from amplicons. The DADA2 (84) pipeline was used to create amplicon sequence variants (ASVs) following the standard pipeline available at <https://benjjneb.github.io/dada2/tutorial.html>. Briefly, reads were trimmed and filtered prior to learning error rates for each plate separately. Inferred sequence variants were then created, forward and reverse reads were merged, and sequence variants were dereplicated. The four sequencing plates were then merged into a single sequence table, and chimeras were removed. Taxonomy was assigned using Silva SSU r138 with the DECIPHER Bioconductor package (85). Libraries with fewer than 10,000 sequences were removed from analysis, and ASVs with fewer than 4 reads in 3 samples were removed.

Co-occurrence networks. Co-occurrence networks were made for station 27 bottom water samples and for station 6 samples to investigate potential associations between the AOA and other community

members. ASVs were filtered to have at least 50 reads in at least 3 samples to remove low-abundance ASVs that might cause spurious associations. Networks were made using SPIEC-EASI (sparse inverse covariance estimation for ecological association and statistical inference) (86) following standard practices using the neighborhood selection (MB) model. Igraph objects were made based on StARS-refit network. Subgraphs with nodes sharing direct edges with ASV 8 and ASV 1610 were selected using Cytoscape (87).

Quantitative PCR analysis of marine group I *Thaumarchaeota* 16S rRNA gene abundance. qPCR was performed on microbial DNA samples from station 27 using a fluorescent TaqMan assay. The assay amplified a region of the 16S rRNA gene of marine group I (MGI) *Thaumarchaeota* (representing AOA abundance) (88), using primers GI_751F (GTCTACCAGAACAYGTTC) and GI_956R (HGGCGTTGACTCCAATTG) and TaqMan probe MGI_889 FAM-BHQ (5'-6-carboxyfluorescein [FAM]-AGTACGTGAA-BHQ1a-Q-3') (89). Samples and standards were run in triplicate. Reaction mixtures contained a total volume of 25 μ l, including 1 μ l of template DNA, 0.5 μ l of primer (10 μ M), 0.75 μ l of probe (10 μ M), 10.75 μ l of TaqMan environmental master mix 2.0 (Applied Biosystems), and 11.5 μ l of double-distilled water (ddH₂O) and was run on an ABI StepOnePlus real-time PCR system (Applied Biosystems) as follows: 95°C for 10 min followed by 40 cycles of 95°C for 15 s and 55°C for 1 min. Standards were prepared by a dilution series of 10⁻¹ to 10⁻⁷ copies per reaction of linearized plasmids. All standard curves had *r*² values of ≥ 0.99 .

Metagenome sequencing, assembly, and binning. Metagenomes from October 2013 were sequenced via JGI CSP (proposal ID 503022) on an Illumina HiSeq 2500-1TB instrument. Quality-controlled filtered raw metagenome data (JGI Project IDs 1167699 to 1167707) were downloaded from JGI Genome Portal for assembly, binning, and refining using the metaWRAP (v1.3.2) pipeline (90). Metagenomes were assembled using MEGAHIT (v1.1.3) (91) following default parameters and a range of kmers (*k* = 21, 29, 39, 59, 79, 99, 119, 141). Metagenomes were subset using seqtk (seed -s100) to allow targeted assembly of highly abundant AOA. Additional assemblies were made using both MEGAHIT and metaSPAdes (v3.13.0) (92) using 1%, 3%, 5%, 10%, and 20% of metagenome reads. Assemblies were then binned using contigs of >2,000 nucleotides (nt) using both MetaBAT2 (v2.12.1) (93) and MaxBin 2.0 (v2.2.6) (94). Coassemblies of stations based on salinity zone were also done to target lower-abundance nitrifiers and binned using contigs of >2,500 nt using MetaBAT2 and MaxBin 2.0 plus CONCOCT (v1.1.0) (95) for coassemblies. Bins were consolidated and filtered using metawrap bin_refinement to be >50% complete and have <10% contamination. These consolidated bins were then reassembled with metaSPAdes using strict or permissive algorithm using metawrap reassemble_bins. Completeness and contamination of MAGs were calculated using CheckM (v1.1.3) (96). Reads were competitively recruited to MAGs using Bowtie2 (v2.4.2) (97) and the default parameters. Abundances are displayed as unpaired reads recruited per genome size of MAG in kilobases of MAG divided by gigabase of metagenome (RPKG). Genes were called using Prodigal (v2.6.3) (98). Translated sequences from each AOA MAG were annotated using the RAST tool kit (RAST-tk v2.0) (99, 100) for SEED (101) annotation and GhostKOALA (102) for KEGG KO annotations. Taxonomic classification for each MAG was performed using the Genome Taxonomy Database toolkit (GTDB-tk) (103) with the database release 05-RS95. MAGs were dereplicated using dRep (v2.3.2) (104) at 98% ANI.

Phylogenomic and pangenomic analysis. All *Nitrososphaerales* genomes in the GTDB database (release 05-RS95, 17 July 2020) were downloaded from the National Center of Biotechnology Information (NCBI) database (*n* = 170), and all genomes annotated as *Thaumarchaeota* that were nonredundant with the NCBI data set were downloaded from the Integrated Microbial Genome (IMG) database (*n* = 177) in January 2021. Additional pelagic estuarine and coastal *Thaumarchaeota* MAGs were downloaded from several other sequence databases (see Table S1 for full details on database, study citations, and accession numbers). *anvi'o* (hope v7) (105) was used to annotate, analyze, and compare all of the downloaded *Thaumarchaeota* genomes, in addition to the two finalized AOA MAGs generated in this study, following the standard phylogenomics and pangenomics workflows. Contigs shorter than 500 bp were excluded from the analysis. Conserved ribosomal and housekeeping genes were annotated with the 7 default hmm databases used by *anvi-run-hmms* with the argument `-also-scan-trnas`, which uses tRNAscan-SE (v2.0.7) to annotate tRNAs. All genomes were annotated through the *anvi'o* pipeline using 3 databases, including COG (*anvi-run-ncbi-cogs*, Diamond set to fast), KEGG (*anvi-run-kegg-kofams*, using *hmmsearch*), and Pfam (*anvi-run-pfams*, using *hmmsearch*). Completion and contamination were calculated using CheckM (v 1.1.3). Concatenated and aligned amino acid sequences were retrieved from MAGs using *anvi-get-sequences-for-hmm-hits* for all genomes with <5% contamination and >70% complete and containing 30 of 36 archaeal ribosomal genes, leaving 153 genomes in the analysis. Genes not occurring in at least 145 of 153 genomes were removed, yielding 24 remaining ribosomal genes, including L16, S8, S3Ae, L29, L26, S9, S17, S11, S2, L6, S13, L32e, S8e, L3, L4, L13, S12_S23, S19, L14, L22, L21e, L23, L1, and S7. Trimal was used to remove gaps in amino acid alignments with less than 80% coverage. Then, IQ-TREE -m MFP was used for extended model selection for phylogeny of MAGs. IQ-TREE was used to construct the final phylogenomic tree using model LG+R5+F with 1,000 bootstraps. Based on clusters in the tree and the GTDB species annotations, a pangenomics analysis was performed to compare two different groups within the *Nitrosopumilaceae*. Genomes with >10% contamination and <70% completeness were removed from pangenomics workflow, leading to a total of 39 genomes included in the pangenome. The ANI of genomes was calculated using *pyANI* (106) through *anvi'o* using default parameters. Gene enrichment between groups was calculated using the *anvi-compute-functional-enrichment* function for all three annotation sources, and significant genes were selected based on the adjusted *P* value to correct for multiple testing. Gene sequences of interest (i.e., *amoA* and *ureC*) were collected from pangenomes using *anvi-get-sequences-for-gene-clusters*.

Statistical analysis. Regressions were made using Pearson correlation with the *cor()* function from base R (107) stats package.

Data availability. Amplicon libraries of 16S rRNA genes are available in the NCBI SRA under BioProject no. [PRJNA577706](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA577706). Metagenomes are available under NCBI BioProject no. [PRJNA439806](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA439806)

through PRJNA439813. The final set of two AOA MAGs were deposited in NCBI under Biosample no. SAMN22441885 and SAMN22441886. SFB water quality data are available from the USGS database: <https://sfbay.wr.usgs.gov/water-quality-database/>.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, EPS file, 0.1 MB.

FIG S2, EPS file, 0.7 MB.

FIG S3, EPS file, 0.3 MB.

FIG S4, EPS file, 0.1 MB.

FIG S5, EPS file, 1.2 MB.

FIG S6, EPS file, 1 MB.

FIG S7, EPS file, 1.3 MB.

TABLE S1, PDF file, 0.04 MB.

TABLE S2, PDF file, 0.04 MB.

TABLE S3, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

Thanks go to Julian Damashek for sample collection and Linta Reji for advice on metagenomics workflows. Thanks go to Jim Cloern and the Water Quality of San Francisco Bay monitoring group at USGS and the R/V *Polaris* crew for facilitating our participation in numerous cruises. Sequencing was carried out thanks to JGI CSP project 503022 to C.A.F. Special thanks go to Tijana Glavino del Rio at JGI. Some of the computing for this project was performed on the Sherlock cluster. We thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results.

This work was supported in part by NSF CAREER grant OCE-0847266 from the Biological Oceanography program (to C.A.F.) and in part by fellowship support from the NSF GRFP and Stanford Data Science Scholars program (to A.N.R.). The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-05CH11231.

We declare no conflict of interest.

REFERENCES

- Vitousek PM, Mooney HA, Lubchenco J, Melillo JM. 1997. Human domination of Earth's ecosystems. *Science* 277:494–499. <https://doi.org/10.1126/science.277.5325.494>.
- Novick E, Senn D. 2014. External nutrient loads to San Francisco Bay. Contribution 704. San Francisco Estuary Institute, Richmond, CA.
- Hager SW, Schemel LE. 1992. Sources of nitrogen and phosphorus to northern San Francisco Bay. *Estuaries* 15:40–52. <https://doi.org/10.2307/1352708>.
- Cloern JE. 2001. Our evolving conceptual model of the coastal eutrophication problem. *Mar Ecol Prog Ser* 210:223–253. <https://doi.org/10.3354/meps210223>.
- Alpine AE, Cloern JE. 1992. Trophic interactions and direct physical effects control phytoplankton biomass and production in an estuary. *Limnol Oceanogr* 37:946–955. <https://doi.org/10.4319/lo.1992.37.5.0946>.
- Cole BE, Cloern JE. 1984. Significance of biomass and light availability to phytoplankton productivity in San Francisco Bay. *Mar Ecol Prog Ser* 17:15–24. <https://doi.org/10.3354/meps017015>.
- Cloern JE. 1999. The relative importance of light and nutrient limitation of phytoplankton growth: a simple index of coastal ecosystem sensitivity to nutrient enrichment. *Aquat Ecol* 33:3–15. <https://doi.org/10.1023/A:1009952125558>.
- Cloern JE, Foster SQ, Kleckner AE. 2014. Phytoplankton primary production in the world's estuarine-coastal ecosystems. *Biogeosciences* 11:2477–2501. <https://doi.org/10.5194/bg-11-2477-2014>.
- Kimmerer W. 2004. Open water processes of the San Francisco Estuary: from physical forcing to biological responses. *San Franc Estuary Watershed Sci* 2. <https://doi.org/10.15447/sfews.2004v2iss1art1>.
- Dugdale RC, Wilkerson FP, Hogue VE, Marchi A. 2007. The role of ammonium and nitrate in spring bloom development in San Francisco Bay. *Estuar Coast Shelf Sci* 73:17–29. <https://doi.org/10.1016/j.ecss.2006.12.008>.
- Dugdale R, Wilkerson F, Parker AE, Marchi A, Taberski K. 2012. River flow and ammonium discharge determine spring phytoplankton blooms in an urbanized estuary. *Estuar Coast Shelf Sci* 115:187–199. <https://doi.org/10.1016/j.ecss.2012.08.025>.
- Parker AE, Dugdale RC, Wilkerson FP. 2012. Elevated ammonium concentrations from wastewater discharge depress primary productivity in the Sacramento River and the Northern San Francisco Estuary. *Mar Pollut Bull* 64:574–586. <https://doi.org/10.1016/j.marpolbul.2011.12.016>.
- Wilkerson FP, Dugdale RC, Hogue VE, Marchi A. 2006. Phytoplankton blooms and nitrogen productivity in San Francisco Bay. *Estuaries Coasts* 29:401–416. <https://doi.org/10.1007/BF02784989>.
- Glibert PM, Dugdale RC, Wilkerson F, Parker AE, Alexander J, Antell E, Blaser S, Johnson A, Lee J, Lee T, Murasko S, Strong S. 2014. Major—but rare—spring blooms in 2014 in San Francisco Bay Delta, California, a result of the long-term drought, increased residence time, and altered nutrient loads and forms. *J Exp Mar Biol Ecol* 460:8–18. <https://doi.org/10.1016/j.jembe.2014.06.001>.
- Glibert PM, Wilkerson FP, Dugdale RC, Parker AE, Alexander J, Blaser S, Murasko S. 2014. Phytoplankton communities from San Francisco Bay Delta respond differently to oxidized and reduced nitrogen substrates—even under conditions that would otherwise suggest nitrogen sufficiency. *Front Mar Sci* 1:17.
- Strong AL, Mills MM, Huang IB, van Dijken GL, Driscoll SE, Berg GM, Kudela RM, Monismith SG, Francis CA, Arrigo KR. 2021. Response of

- Lower Sacramento River phytoplankton to high-ammonium wastewater effluent. *Elem Sci Anthr* 9:040. <https://doi.org/10.1525/elementa.2021.040>.
17. Hammond DE, Fuller C, Harmon D, Hartman B, Korosec M, Miller LG, Rea R, Warren S, Berelson W, Hager SW. 1985. Benthic fluxes in San Francisco Bay. *Hydrobiologia* 129:69–90. <https://doi.org/10.1007/BF00048688>.
 18. Cornwell JC, Glibert PM, Owens MS. 2014. Nutrient fluxes from sediments in the San Francisco Bay delta. *Estuaries Coasts* 37:1120–1133. <https://doi.org/10.1007/s12237-013-9755-4>.
 19. Damashek J, Casciotti KL, Francis CA. 2016. Variable nitrification rates across environmental gradients in turbid, nutrient-rich estuary waters of San Francisco Bay. *Estuaries Coasts* 39:1050–1071. <https://doi.org/10.1007/s12237-016-0071-7>.
 20. Kraus TEC, O'Donnell K, Downing BD, Burau JR, Bergamaschi BA. 2017. Using paired in situ high frequency nitrate measurements to better understand controls on nitrate concentrations and estimate nitrification rates in a wastewater-impacted river. *Water Resour Res* 53:8423–8442. <https://doi.org/10.1002/2017WR020670>.
 21. Mosier AC, Francis CA. 2008. Relative abundance and diversity of ammonia-oxidizing archaea and bacteria in the San Francisco Bay estuary. *Environ Microbiol* 10:3002–3016. <https://doi.org/10.1111/j.1462-2920.2008.01764.x>.
 22. Damashek J, Smith JM, Mosier AC, Francis CA. 2014. Benthic ammonia oxidizers differ in community structure and biogeochemical potential across a riverine delta. *Front Microbiol* 5:743. <https://doi.org/10.3389/fmicb.2014.00743>.
 23. Mosier AC, Lund MB, Francis CA. 2012. Ecophysiology of an ammonia-oxidizing archaeon adapted to low-salinity habitats. *Microb Ecol* 64: 955–963. <https://doi.org/10.1007/s00248-012-0075-1>.
 24. Mosier AC, Allen EE, Kim M, Ferriera S, Francis CA. 2012. Genome sequence of “Candidatus Nitrosopumilus salaria” BD31, an ammonia-oxidizing archaeon from the San Francisco Bay estuary. *J Bacteriol* 194: 2121–2122. <https://doi.org/10.1128/JB.00013-12>.
 25. Mosier AC, Allen EE, Kim M, Ferriera S, Francis CA. 2012. Genome sequence of “Candidatus Nitrosoarchaeum limnia” BG20, a low-salinity ammonia-oxidizing archaeon from the San Francisco Bay estuary. *J Bacteriol* 194:2119–2120. <https://doi.org/10.1128/JB.00007-12>.
 26. Blainey PC, Mosier AC, Potanina A, Francis CA, Quake SR. 2011. Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLoS One* 6:e16626. <https://doi.org/10.1371/journal.pone.0016626>.
 27. Li J, Nedwell DB, Beddow J, Dumbrell AJ, McKew BA, Thorpe EL, Whitby C. 2015. amoA gene abundances and nitrification potential rates suggest that benthic ammonia-oxidizing bacteria and not archaea dominate N cycling in the Colne Estuary, United Kingdom. *Appl Environ Microbiol* 81:159–165. <https://doi.org/10.1128/AEM.02654-14>.
 28. Zhang Y, Xie X, Jiao N, Hsiao S-Y, Kao S-J. 2014. Diversity and distribution of amoA-type nitrifying and nirS-type denitrifying microbial communities in the Yangtze River estuary. *Biogeosciences* 11:2131–2145. <https://doi.org/10.5194/bg-11-2131-2014>.
 29. Hou L, Xie X, Wan X, Kao S-J, Jiao N, Zhang Y. 2018. Niche differentiation of ammonia and nitrite oxidizers along a salinity gradient from the Pearl River estuary to the South China Sea. *Biogeosciences* 15:5169–5187. <https://doi.org/10.5194/bg-15-5169-2018>.
 30. Urakawa H, Martens-Habbena W, Huguet C, de la Torre JR, Ingalls AE, Devol AH, Stahl DA. 2014. Ammonia availability shapes the seasonal distribution and activity of archaeal and bacterial ammonia oxidizers in the Puget Sound Estuary. *Limnol Oceanogr* 59:1321–1335. <https://doi.org/10.4319/lo.2014.59.4.1321>.
 31. Happel E, Bartl I, Voss M, Riemann L. 2018. Extensive nitrification and active ammonia oxidizers in two contrasting coastal systems of the Baltic Sea. *Environ Microbiol* 20:2913–2926. <https://doi.org/10.1111/1462-2920.14293>.
 32. Santos JP, Mendes D, Monteiro M, Ribeiro H, Baptista MS, Borges MT, Magalhães C. 2018. Salinity impact on ammonia oxidizers activity and amoA expression in estuarine sediments. *Estuar Coast Shelf Sci* 211: 177–187. <https://doi.org/10.1016/j.ecss.2017.09.001>.
 33. Hampel JJ, McCarthy MJ, Aalto SL, Newell SE. 2020. Hurricane disturbance stimulated nitrification and altered ammonia oxidizer community structure in Lake Okeechobee and St. Lucie Estuary (Florida). *Front Microbiol* 11:1541. <https://doi.org/10.3389/fmicb.2020.01541>.
 34. Caffrey JM, Bano N, Kalanetra K, Hollibaugh JT. 2007. Ammonia oxidation and ammonia-oxidizing bacteria and archaea from estuaries with differing histories of hypoxia. *ISME J* 1:660–662. <https://doi.org/10.1038/ismej.2007.79>.
 35. Xie W, Luo H, Murugapiran SK, Dodsworth JA, Chen S, Sun Y, Hedlund BP, Wang P, Fang H, Deng M, Zhang CL. 2018. Localized high abundance of Marine Group II archaea in the subtropical Pearl River Estuary: implications for their niche adaptation. *Environ Microbiol* 20:734–754. <https://doi.org/10.1111/1462-2920.14004>.
 36. Schaefer SC, Hollibaugh JT. 2017. Temperature decouples ammonium and nitrite oxidation in coastal waters. *Environ Sci Technol* 51: 3157–3164. <https://doi.org/10.1021/acs.est.6b03483>.
 37. Hollibaugh JT, Gifford SM, Moran MA, Ross MJ, Sharma S, Tolar BB. 2014. Seasonal variation in the metatranscriptomes of a Thaumarchaeota population from SE USA coastal waters. *ISME J* 8:685–698. <https://doi.org/10.1038/ismej.2013.171>.
 38. Qin W, Zheng Y, Zhao F, Wang Y, Urakawa H, Martens-Habbena W, Liu H, Huang X, Zhang X, Nakagawa T, Mende DR, Bollmann A, Wang B, Zhang Y, Amin SA, Nielsen JL, Mori K, Takahashi R, Virginia Armbrust E, Winkler M-KH, DeLong EF, Li M, Lee P-H, Zhou J, Zhang C, Zhang T, Stahl DA, Ingalls AE. 2020. Alternative strategies of nutrient acquisition and energy conservation map to the biogeography of marine ammonia-oxidizing archaea. *ISME J* 14:2595–2609. <https://doi.org/10.1038/s41396-020-0710-7>.
 39. Wang B, Qin W, Ren Y, Zhou X, Jung M-Y, Han P, Eloe-Fadrosh EA, Li M, Zheng Y, Lu L, Yan X, Ji J, Liu Y, Liu L, Heiner C, Hall R, Martens-Habbena W, Herbold CW, Rhee S, Bartlett DH, Huang L, Ingalls AE, Wagner M, Stahl DA, Jia Z. 2019. Expansion of Thaumarchaeota habitat range is correlated with horizontal transfer of ATPase operons. *ISME J* 13:3067–3079. <https://doi.org/10.1038/s41396-019-0493-x>.
 40. Rasmussen AN, Damashek J, Eloe-Fadrosh EA, Francis CA. 2021. In-depth spatiotemporal characterization of planktonic archaeal and bacterial communities in North and South San Francisco Bay. *Microb Ecol* 81: 601–616. <https://doi.org/10.1007/s00248-020-01621-7>.
 41. Ahlgren NA, Chen Y, Needham DM, Parada AE, Sachdeva R, Trinh V, Chen T, Fuhrman JA. 2017. Genome and epigenome of a novel marine Thaumarchaeota strain suggest viral infection, phosphorothioation DNA modification and multiple restriction systems. *Environ Microbiol* 19: 2434–2452. <https://doi.org/10.1111/1462-2920.13768>.
 42. Kim J-G, Park S-J, Damsté JSS, Schouten S, Rijpstra WIC, Jung M-Y, Kim S-J, Gwak J-H, Hong H, Si O-J, Lee S, Madsen EL, Rhee S-K. 2016. Hydrogen peroxide detoxification is a key mechanism for growth of ammonia-oxidizing archaea. *Proc Natl Acad Sci U S A* 113:7888–7893. <https://doi.org/10.1073/pnas.1605501113>.
 43. Daims H, Lüscher S, Wagner M. 2016. A new perspective on microbes formerly known as nitrite-oxidizing bacteria. *Trends Microbiol* 24:699–712. <https://doi.org/10.1016/j.tim.2016.05.004>.
 44. Helder W, De Vries RTP. 1983. Estuarine nitrite maxima and nitrifying bacteria (Ems-Dollard estuary). *Neth J Sea Res* 17:1–18. [https://doi.org/10.1016/0077-7579\(83\)90002-9](https://doi.org/10.1016/0077-7579(83)90002-9).
 45. McCarthy JJ, Kaplan W, Nevins JL. 1984. Chesapeake Bay nutrient and plankton dynamics. 2. Sources and sinks of nitrite1. *Limnol Oceanogr* 29: 84–98. <https://doi.org/10.4319/lo.1984.29.1.0084>.
 46. Laperriere SM, Nidzieko NJ, Fox RJ, Fisher AW, Santoro AE. 2019. Observations of variable ammonia oxidation and nitrous oxide flux in a eutrophic estuary. *Estuaries Coasts* 42:33–44. <https://doi.org/10.1007/s12237-018-0441-4>.
 47. Horrigan SG, Montoya JP, Nevins JL, McCarthy JJ, Ducklow H, Goericke R, Malone T. 1990. Nitrogenous nutrient transformations in the spring and fall in the Chesapeake Bay. *Estuar Coast Shelf Sci* 30:369–391. [https://doi.org/10.1016/0272-7714\(90\)90004-B](https://doi.org/10.1016/0272-7714(90)90004-B).
 48. Casciotti K, Buchwald C. 2012. Insights on the marine microbial nitrogen cycle from isotopic approaches to nitrification. *Front Microbiol* 3:356. <https://doi.org/10.3389/fmicb.2012.00356>.
 49. Guisasaola A, Jubany I, Baeza JA, Carrera J, Lafuente J. 2005. Respirometric estimation of the oxygen affinity constants for biological ammonium and nitrite oxidation. *J Chem Technol Biotechnol* 80:388–396. <https://doi.org/10.1002/jctb.1202>.
 50. Philips S, Laanbroek HJ, Verstraete W. 2002. Origin, causes and effects of increased nitrite concentrations in aquatic environments. *Rev Environ Sci Biotechnol* 1:115–141. <https://doi.org/10.1023/A:1020892826575>.
 51. Bristow LA, Sarode N, Cartee J, Caro-Quintero A, Thamdrup B, Stewart FJ. 2015. Biogeochemical and metagenomic analysis of nitrite accumulation in the Gulf of Mexico hypoxic zone. *Limnol Oceanogr* 60:1733–1750. <https://doi.org/10.1002/lno.10130>.
 52. Kim J-G, Gwak J-H, Jung M-Y, An S-U, Hyun J-H, Kang S, Rhee S-K. 2019. Distinct temporal dynamics of planktonic archaeal and bacterial

- assemblages in the bays of the Yellow Sea. *PLoS One* 14:e0221408. <https://doi.org/10.1371/journal.pone.0221408>.
53. Pitcher A, Wuchter C, Siedenberg K, Schouten S, Sinninghe Damsté JS. 2011. Crenarchaeol tracks winter blooms of ammonia-oxidizing Thaumarchaeota in the coastal North Sea. *Limnol Oceanogr* 56:2308–2318. <https://doi.org/10.4319/lo.2011.56.6.2308>.
 54. Guerrero MA, Jones RD. 1996. Photoinhibition of marine nitrifying bacteria. I. Wavelength-dependent response. *Mar Ecol Prog Ser* 141:183–192. <https://doi.org/10.3354/meps141183>.
 55. Merbt SN, Stahl DA, Casamayor EO, Martí E, Nicol GW, Prosser JI. 2012. Differential photoinhibition of bacterial and archaeal ammonia oxidation. *FEMS Microbiol Lett* 327:41–46. <https://doi.org/10.1111/j.1574-6968.2011.02457.x>.
 56. Vanzella A, Guerrero M, Jones R. 1989. Effect of CO and light on ammonium and nitrite oxidation by chemolithotrophic bacteria. *Mar Ecol Prog Ser* 57:69–76. <https://doi.org/10.3354/meps057069>.
 57. Horrigan SG, Springer AL. 1990. Oceanic and estuarine ammonium oxidation: effects of light. *Limnol Oceanogr* 35:479–482. <https://doi.org/10.4319/lo.1990.35.2.0479>.
 58. Kim K, Park Y-G. 2021. Light as a novel inhibitor of nitrite-oxidizing bacteria (NOB) for the mainstream partial nitrification of wastewater treatment. *Processes* 9:346. <https://doi.org/10.3390/pr9020346>.
 59. Wang L, Qiu S, Guo J, Ge S. 2021. Light irradiation enables rapid start-up of nitrification through suppressing *nrxB* gene expression and stimulating ammonia-oxidizing bacteria. *Environ Sci Technol* 55:13297–13305. <https://doi.org/10.1021/acs.est.1c04174>.
 60. Smith JM, Chavez FP, Francis CA. 2014. Ammonium uptake by phytoplankton regulates nitrification in the sunlit ocean. *PLoS One* 9:e108173. <https://doi.org/10.1371/journal.pone.0108173>.
 61. Olson RJ. 1981. Differential photoinhibition of marine nitrifying bacteria: a possible mechanism for the formation of the primary nitrite maximum. *J Mar Res* 39:227–238.
 62. Santoro AE, Sakamoto CM, Smith JM, Plant JN, Gehman AL, Worden AZ, Johnson KS, Francis CA, Casciotti KL. 2013. Measurements of nitrite production in and around the primary nitrite maximum in the central California Current. *Biogeosciences* 10:7395–7410. <https://doi.org/10.5194/bg-10-7395-2013>.
 63. Beman JM, Sachdeva R, Fuhrman JA. 2010. Population ecology of nitrifying Archaea and Bacteria in the Southern California Bight. *Environ Microbiol* 12:1282–1292. <https://doi.org/10.1111/j.1462-2920.2010.02172.x>.
 64. Zou D, Li Y, Kao S-J, Liu H, Li M. 2019. Genomic adaptation to eutrophication of ammonia-oxidizing archaea in the Pearl River estuary. *Environ Microbiol* 21:2320–2332. <https://doi.org/10.1111/1462-2920.14613>.
 65. Zou D, Wan R, Han L, Xu MN, Liu Y, Liu H, Kao S-J, Li M. 2020. Genomic characteristics of a novel species of ammonia-oxidizing archaea from the Jiulong River estuary. *Appl Environ Microbiol* 86:e00736–20. <https://doi.org/10.1128/AEM.00736-20>.
 66. Walters RA, Cheng RT, Conomos TJ. 1985. Time scales of circulation and mixing processes of San Francisco Bay waters. *Hydrobiologia* 129:13–36. <https://doi.org/10.1007/BF00048685>.
 67. Reji L, Tolar BB, Smith JM, Chavez FP, Francis CA. 2019. Differential co-occurrence relationships shaping ecotype diversification within Thaumarchaeota populations in the coastal ocean water column. *ISME J* 13:1144–1158. <https://doi.org/10.1038/s41396-018-0311-x>.
 68. Orellana LH, Ben Francis T, Krüger K, Teeling H, Müller M-C, Fuchs BM, Konstantinidis KT, Amann RL. 2019. Niche differentiation among annually recurrent coastal Marine Group II Euryarchaeota. *ISME J* 13:3024–3036. <https://doi.org/10.1038/s41396-019-0491-z>.
 69. Hugerth LW, Larsson J, Alneberg J, Lindh MV, Legrand C, Pinhassi J, Andersson AF. 2015. Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biol* 16:279. <https://doi.org/10.1186/s13059-015-0834-7>.
 70. Kitzinger K, Padilla CC, Marchant HK, Hach PF, Herbold CW, Kidane AT, Könneke M, Littmann S, Mooshammer M, Niggemann J, Petrov S, Richter A, Stewart FJ, Wagner M, Kuypers MMM, Bristow LA. 2019. Cyanate and urea are substrates for nitrification by Thaumarchaeota in the marine environment. *Nat Microbiol* 4:234–243. <https://doi.org/10.1038/s41564-018-0316-2>.
 71. Damashek J, Edwardson CF, Tolar BB, Gifford SM, Moran MA, Hollibaugh JT. 2019. Coastal ocean metagenomes and curated metagenome-assembled genomes from Marsh Landing, Sapelo Island (Georgia, USA). *Microbiol Resour Announc* 8:e00934–19. <https://doi.org/10.1128/MRA.00934-19>.
 72. Pinto OHB, Silva TF, Vizzotto CS, Santana RH, Lopes FAC, Silva BS, Thompson FL, Kruger RH. 2020. Genome-resolved metagenomics analysis provides insights into the ecological role of Thaumarchaeota in the Amazon River and its plume. *BMC Microbiol* 20:13. <https://doi.org/10.1186/s12866-020-1698-x>.
 73. Santos-Júnior CD, Kishi LT, Toyama D, Soares-Costa A, Oliveira TCS, de Miranda FP, Henrique-Silva F. 2017. Metagenome sequencing of prokaryotic microbiota collected from rivers in the Upper Amazon Basin. *Genome Announc* 5:e01450–16. <https://doi.org/10.1128/genomeA.01450-16>.
 74. Widderich N, Czech L, Elling FJ, Könneke M, Stöveken N, Pittelkow M, Riclea R, Dickschat JS, Heider J, Bremer E. 2016. Strangers in the archaeal world: osmostress-responsive biosynthesis of ectoine and hydroxyectoine by the marine thaumarchaeon *Nitrosopumilus maritimus*: ectoine and hydroxyectoine biosynthesis in *Archaea*. *Environ Microbiol* 18:1227–1248. <https://doi.org/10.1111/1462-2920.13156>.
 75. Ren M, Feng X, Huang Y, Wang H, Hu Z, Clingenpeel S, Swan BK, Fonseca MM, Posada D, Stepanauskas R, Hollibaugh JT, Foster PG, Woyke T, Luo H. 2019. Phylogenomics suggests oxygen availability as a driving force in Thaumarchaeota evolution. *ISME J* 13:2150–2161. <https://doi.org/10.1038/s41396-019-0418-8>.
 76. Santoro AE, Dupont CL, Richter RA, Craig MT, Carini P, McIlvin MR, Yang Y, Orsi WD, Moran DM, Saito MA. 2015. Genomic and proteomic characterization of “*Candidatus Nitrosopelagicus brevis*”: an ammonia-oxidizing archaeon from the open ocean. *Proc Natl Acad Sci U S A* 112:1173–1178. <https://doi.org/10.1073/pnas.1416223112>.
 77. Hollibaugh JT, Gifford S, Sharma S, Bano N, Moran MA. 2011. Metatranscriptomic analysis of ammonia-oxidizing organisms in an estuarine bacterioplankton assemblage. *ISME J* 5:866–878. <https://doi.org/10.1038/ismej.2010.172>.
 78. Barco RA, Garrity GM, Scott JJ, Amend JP, Nealson KH, Emerson D, Giovannoni SJ. 2020. A genus definition for bacteria and archaea based on a standard genome relatedness index. *mBio* 11:e02475–19. <https://doi.org/10.1128/mBio.02475-19>.
 79. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P. 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* 38:1079–1086. <https://doi.org/10.1038/s41587-020-0501-8>.
 80. Schraga TS, Cloern JE. 2017. Water quality measurements in San Francisco Bay by the U.S. Geological Survey, 1969–2015. *Sci Data* 4:170098. <https://doi.org/10.1038/sdata.2017.98>.
 81. Bower CE, Holm-Hansen T. 1980. A salicylate-hypochlorite method for determining ammonia in seawater. *Can J Fish Aquatic Sci* 37:794–798. <https://doi.org/10.1139/f80-106>.
 82. Parada AE, Needham DM, Fuhrman JA. 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol* 18:1403–1414. <https://doi.org/10.1111/1462-2920.13023>.
 83. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. 1. *EMBnet J* 17:10–12. <https://doi.org/10.14806/ej.17.1.200>.
 84. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583. <https://doi.org/10.1038/nmeth.3869>.
 85. Wright ES. 2016. Using DECIPHER v2.0 to analyze big biological sequence data in R. *R J* 8:352–359. <https://doi.org/10.32614/RJ-2016-025>.
 86. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. 2015. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput Biol* 11:e1004226. <https://doi.org/10.1371/journal.pcbi.1004226>.
 87. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504. <https://doi.org/10.1101/gr.1239303>.
 88. Mincer TJ, Church MJ, Taylor LT, Preston C, Karl DM, DeLong EF. 2007. Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. *Environ Microbiol* 9:1162–1175. <https://doi.org/10.1111/j.1462-2920.2007.01239.x>.
 89. Lund MB, Smith JM, Francis CA. 2012. Diversity, abundance and expression of nitrite reductase (*nirK*)-like genes in marine thaumarchaea. *ISME J* 6:1966–1977. <https://doi.org/10.1038/ismej.2012.40>.
 90. Uritskiy GV, DiRuggiero J, Taylor J. 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6:158. <https://doi.org/10.1186/s40168-018-0541-1>.

91. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>.
92. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27:824–834. <https://doi.org/10.1101/gr.213959.116>.
93. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. <https://doi.org/10.7717/peerj.7359>.
94. Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32:605–607. <https://doi.org/10.1093/bioinformatics/btv638>.
95. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. *Nat Methods* 11:1144–1146. <https://doi.org/10.1038/nmeth.3103>.
96. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
97. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
98. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>.
99. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9:75. <https://doi.org/10.1186/1471-2164-9-75>.
100. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 5:8365. <https://doi.org/10.1038/srep08365>.
101. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 42:D206–D214. <https://doi.org/10.1093/nar/gkt1226>.
102. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 428:726–731. <https://doi.org/10.1016/j.jmb.2015.11.006>.
103. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
104. Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* 11:2864–2868. <https://doi.org/10.1038/ismej.2017.126>.
105. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3:e1319. <https://doi.org/10.7717/peerj.1319>.
106. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods* 8:12–24. <https://doi.org/10.1039/C5AY02550H>.
107. R Core Team. 2020. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
108. Kerou M, Ponce-Toledo RI, Zhao R, Abby SS, Hirai M, Nomaki H, Takaki Y, Nunoura T, Jørgensen SL, Schleper C. 2021. Genomes of Thaumarchaeota from deep sea sediments reveal specific adaptations of three independently evolved lineages. *ISME J* 15:2792–2808. <https://doi.org/10.1038/s41396-021-00962-6>.