Ammonia-oxidizing archaea in the Baltic Sea: dynamics and enrichment-enabled comparative genomics

M.Sc. Thesis

Sookyoung Kim
MSc Biological Oceanography

March 2024

Supervised by Prof. Dr. David Needham

First examiner: Prof. Dr. David Needham
Second examiner: Prof. Dr. Ute Hentschel Humeida

GEOMAR
Helmholtz Centre for Ocean Research Kiel

C A U
Christian-Albrechts-Universität zu Kiel

ABSTRACT

Marine ammonia-oxidizing archaea (AOA) are among the dominant deep sea marine carbon-fixing organisms and thrive in diverse ecosystems, from the depths of the ocean to coastal environments, estuaries, Arctic regions, and sediments. Despite their rich phylogenetic diversity, the driving forces behind ecological patterns and niche differentiation among AOA strains remain largely enigmatic, yet their influence over marine microbial ecosystems and biogeochemical processes is substantial. In this master's project, I aimed to investigate seasonal dynamics and genomic characteristics of different AOA strains in the Baltic Sea. Our studies reveal that they can be among the dominant taxa during late fall and Winter in the surface waters of the Baltic Sea as well as dominant throughout the year in deeper waters. I enriched and genome-sequenced numerous novel AOA from across the *Nitrosopumilus* genus, representing new species including a dominant deep Baltic Sea strain. In addition, I acquired the genome sequence of the dominant AOA strain inhabiting the winter surface layer through bulk metagenomic assembly. To reveal the niche-defining attributes, I compared the genomes of the Baltic Sea *Nitrosopumilus* as well as other publicly available genomes. Remarkably, these comparative analyses unveiled a high degree of gene conservation across the *Nitrosopumilus* genus, specifically, 57.2% (±7.5%) of all genes are part of the 'conserved' genome (meaning 90% of genomes contain the gene), suggesting a fundamental core of genes shared among AOA strains. However, despite these shared genes, all strains also had many unique genes (63.3% of all gene clusters), hinting at potential adaptations, and shedding light on their diverse ecological roles. This research enriches our understanding of AOA diversity and genomics, potentially illuminating AOA's impact on the Baltic Sea ecosystem. By integrating time-series and distributional data with genomic analysis, this project deepens our understanding of the complex ecosystem in relation to AOA within this unique environment.

Table of Contents

1. INTRODUCTION

1.1 Ammonia oxidizing archaea (AOA)

Nitrogen, a crucial element for life, plays an important role in controlling microbial productivity. Among the processes within the nitrogen cycle, nitrification is responsible for the oxidation of ammonium to nitrate, which accounts for ~96% of the total bioavailable nitrogen in the marine ecosystem (Gruber and Galloway 2008). The initial step of nitrification, the oxidation of ammonium to nitrite, was long attributed predominantly to Ammonia-oxidizing bacteria (AOB) broadly which are distributed in marine environments (Mincer et al. 2007). However, this understanding was revised with the discovery and isolation of Ammonium Oxidizing Archaea (AOA), specifically after the isolation of *Nitrosopumilus maritimus* (Könneke et al. 2005). At the time, it was observed that *N. maritimus* was affiliated with the same lineage that had been found abundant in the deep ocean, which were referred to as Marine Group I Archaea or Marine Group I Crenarchaea, and later Thuamarchaea (DeLong 1992; Fuhrman et al. 1992). This discovery highlighted the potentially substantial role of AOA in marine nitrification processes, and indeed research suggests that AOA make a substantial contribution to ammonia (NH_3) oxidation compared to AOB (Wuchter et al. 2006). Currently, AOA are classified under the phylum Thaumarchaeota (formally Crenarchaeota), class Nitrososphaeria, order Nitrososphaerales, and family Nitrosopumilaceae according to the GTDB; Genome Taxonomy Database. (Brochier-Armanet et al. 2008; Parks et al. 2020).

After the first discovery of AOA, it turns out that AOA are one the most abundant and ubiquitous microbes on Earth (Francis et al. 2005) including both marine (Wuchter et al. 2006) and terrestrial (Leininger et al. 2006) ecosystems. AOA have been identified as a dominant marine group, accounting for nearly 40% of all bacterioplankton in the mesopelagic ocean (Karner et al. 2001). Notably, AOA are also prevalent in various aquatic environments, such as freshwater lakes (Vissers et al. 2013), estuaries (Liu et al. 2018), as well as the Baltic Sea (Berg et al. 2015; Happel et al. 2018), North Sea (Herfort et al. 2007; Pitcher et al. 2011; Bale et al. 2013), and Arctic (Müller et al. 2018). Additionally, they have been observed in sediments (Kim et al. 2011; Mosier et al. 2012), as well as host-associated environments, especially sponges (Preston et al. 1996; Webster et al. 2001; Lee et al. 2003). These observations emphasize the significant importance and ubiquity of AOA within multiple ecosystems.

The isolation of *N. maritimus* and, subsequently, other AOA have facilitated extensive investigations into the metabolic flexibility and the physiological characteristics of AOA. Studies have revealed that AOA utilize the energy-efficient 3-hydroxypropionate/4-hydroxybutyrate pathway for carbon fixation which could be a possible explanation for AOA's success in low-nutrient environments (Walker et al. 2010; Könneke et al. 2014). While it is well-known that AOA are chemoautotrophic organisms

producing energy by oxidizing ammonia to nitrite, this metabolic pathway is only partially elucidated. The first step of ammonia oxidation is conducted by archaeal ammonia monooxygenase (AMO), a homolog of bacterial AMO, that converts ammonia to an intermediate, ammonia hydroxylamine (NH_2OH). NH_2OH is converted to nitrite then after but the presence of genes that could do this has not yet been identified until now (Vajrala et al. 2013). Meanwhile, archaeal AMO appears to be more complex than bacterial AMO. While it was believed that both bacterial and archaeal AMOs are composed of three subunits, recent findings have revealed that archaea possess three additional AMO subunits (Hodgskiss et al. 2023). Like AOB, AOA can produce nitrous oxide (N_2O) gas (https://onlinelibrary.wiley.com/doi/full/10.1111/gcb.14877), which is 300 times more potent as a greenhouse gas than carbon dioxide (Santoro et al. 2011; Löscher et al. 2012). AOA also exhibit the ability to generate oxygen and dinitrogen in oxygen-depleted conditions, highlighting their adaptability to such environments (Kraft et al. 2022).

The evolutionary history of AOA within the phylum Thaumarchaeota is believed to have commenced over a billion years ago, initially emerging from terrestrial environments and later diversifying into marine habitats around 500 million years ago (Betts et al. 2018; Yang et al. 2021). These studies indicate AOA's deep evolutionary roots and their remarkable adaptability over geological timescales. Exploring multiple AOA genomes through phylogenomic studies has revealed an array of adaptive traits among Thaumarchaeota strains, such as osmolality regulation, UV irradiation tolerance, and phosphate transport mechanisms. These traits have significantly contributed to their global distribution and evolutionary success (Ren et al. 2019).

Given the long evolutionary path of AOA, it's unsurprising that they exhibit metabolic versatility, adapting to various environments across the globe. Their ability to utilize ammonia, urea, and cyanate as energy sources underscores this versatility (Palatinszky et al. 2015; Bayer et al. 2016). Some isolates have shown the capacity to take up organic carbon showing mixotrophic characteristics (Qin et al. 2014). This characteristic enables AOA to compete in environments where ammonia is limited, or where the other substrates are relatively more available, broadening their ecological niche (Palatinszky et al. 2015; Bayer et al. 2016). Further research into *Nitrosopumilus* strains reveals fascinating adaptations, such as the possession of archaella, which suggests an active mechanism for nutrient acquisition (Bayer et al. 2016), and a phosphorothioation (PT) system which is a type of DNA modification, providing a defense against genomic threats (Ahlgren et al. 2017). These features, found despite their streamlined genomes, emphasize the ecological adaptability of Thaumarchaeota, enabling them to inhabit diverse environmental niches.

Major discoveries of genomic potentials and key metabolic pathways in AOA have predominantly been achieved through studies on strains that can be cultured in the laboratory. This approach facilitates more intensive and detailed physiological research.

Meanwhile, culture-independent methods, such as single amplified genomes (SAGs) and metagenomic assembled genomes (MAGs), often yield genomes of poor quality. Obtaining genomes through SAGs is particularly challenging due to difficulties in cell lysis, sorting small cells using flow cytometry, and the histone binding of archaeal DNA. Similarly, MAGs also have limited success in recovering AOA genomes and many worldwide sampling projects have successfully assembled only a few high-quality AOA genomes (Santoro et al. 2019; Nishimura and Yoshizawa 2022). This is due to the high microdiversity within AOA populations, making genome assembly difficult. Therefore, enrichment-enabled genomics plays a crucial role in studying the diversity and genomic capabilities of these microorganisms.

1.2 What's known about AOA in the Baltic Sea?

The Baltic Sea, with its brackish nature and limited connection to the North Sea, though a 'young' environment (Snoeijs-Leijonmalm et al. 2017) presents an opportunity to study AOA and nitrogen cycling in a relatively well characterized, ecosystem undergoing rapid change (Reusch et al. 2018). The Baltic Sea environment ranges from highly oceanic in the Skagerrak to very low salinities of freshwater in the Baltic Sea proper. The Baltic Sea has an average depth of 58 m, with the deepest depths reaching approximately 459 m. Depending on the flushing of the Baltic Sea with North Atlantic waters, which is happening less frequently due to climate change (Mohrholz et al. 2015), the deeper depths of Baltic Sea Proper is often hypoxic or anoxic in its aphotic zone depths. In the aphotic zone, AOA have been studied and shown to occupy discrete depths where there is enough oxygen to maintain their primarily aerobic metabolism, and also enough ammonia from below to maintain their growth (Labrenz et al. 2010).

In addition to the the Baltic Sea proper, the biogeochemistry of the southwest Baltic Sea, has been reasonably well characterized. Insights from the Boknis Eck time series station in Eckernförde Bay have significantly enriched our understanding of the area's environmental dynamics. Seasonal stratification, driven by surface warming and a pronounced salinity gradient between inflowing and outflowing water masses, limits vertical mixing from March to September (Gindorf et al. 2022)). This pattern shifts during winter as strong winds and cooling of surface waters facilitate complete mixing of water column (Bange et al. 2010). The region has strong phytoplankton blooms in early spring and autumn (Bange et al. 2010; Dale et al. 2011), leading to increased sedimentation of organic matter and heightened microbial respiration, which in turn causes pronounced hypoxia and occasional anoxia in the bottom waters by late summer (Bange et al. 2010; Lennartz et al. 2014; Steinle et al. 2017). Nutrient dynamics show that inorganic nutrients tend to be low in the summer months, with higher concentrations in the fall until the spring phytoplankton bloom.

Within these complex environmental settings, the Baltic Sea presents challenging yet novel opportunities for the exploration of the ecology of ammonia-oxidizing archaea.

1.3 The scope of this thesis

In this thesis, AOA of the Baltic Sea were studied using time-series observations and biogeographic sampling of their diversity coupled with isolation and genomic characterization. Through preliminary analysis of time-series data from the southwest Baltic Sea, we have observed that AOA numerically dominate the surface waters, especially during winter. This study aims to identify which AOA populations dominate in the Baltic Sea and to explore the reasons behind their predominance. We plan to investigate how environmental parameters such as temperature, nutrients, and dynamics of other nitrifiers—ammonia-oxidizing bacteria (AOB) and nitrite-oxidizing bacteria (NOB)—may influence their seasonal predominance. Furthermore, we aim to enrich various AOA strains present in the Baltic Sea to obtain high-quality genomes from these cultures, thus providing detailed insight into their unique characteristics and adaptations to local environmental conditions. Utilizing phylogenomics and comparative genomics, we seek to elucidate the relationships among these strains and to identify potential adaptive gene sets by comparing genomes we have acquired with those in existing databases. By integrating time-series and distributional data with genomic analysis, this study deepens our understanding of the realized niche and potential drivers between diverse AOA strains, serving as a model for understanding their distribution and function in the deep sea and beyond.

2. METHODS

2.1 Location and sampling

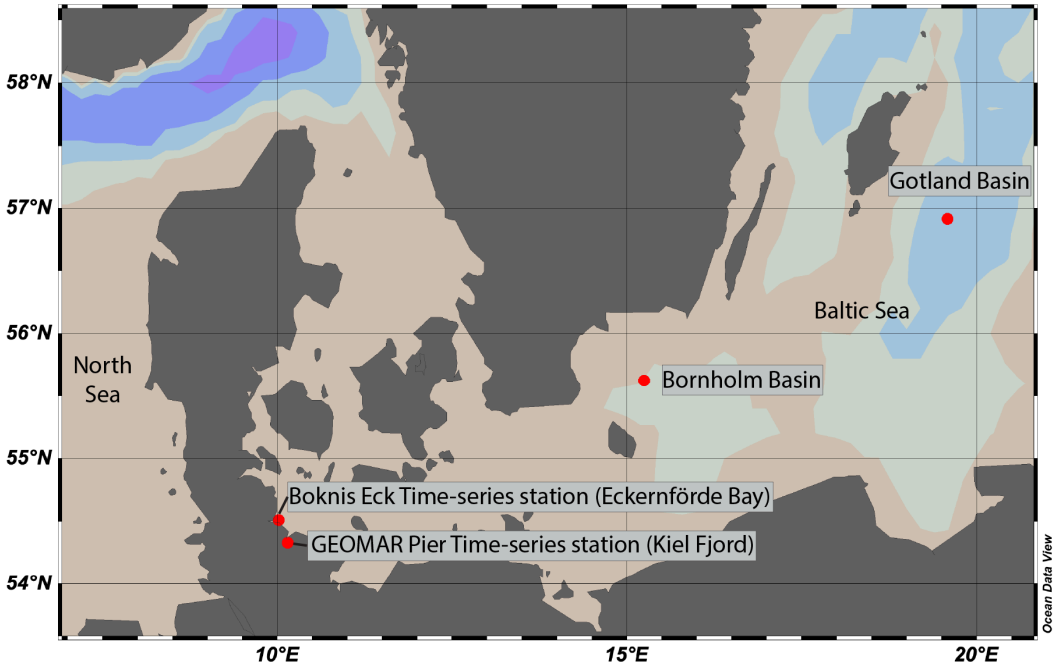


Figure 1: Map of Major Sampling Locations. Boknis Eck and GEOMAR Pier Time-series stations represent the southwest Baltic Sea environment. Bornholm Basin and Gotland Basin are selected to study the deep and expansive areas of the Baltic Sea.

This study is primarily based on the sampling collections from three key locations. First, The GEOMAR pier time series (GPT; 54°19'48.3"N 10°09'01.9"E) provided an accessible spot for twice-weekly surface water sampling for one and a half years (Oct 2021 to May 2023), with only a few samples missing due to sampling constraints. Water was collected by bucket and the seawater was directly brought back to a laboratory for full processing within 2 hours. Due to the convenient location, this site benefits us in monitoring AOA ecology of the southwest Baltic Sea, showing similar dynamics and diversity of AOA in Boknis Eck time-series station (Fig 7). Temperature, oxygen, and salinity were measured at GPT continuously via the Kiel Marine Organism Culture Centre (KIMOCC; https://www.geomar.de/mitarbeiter/fb3/eoe-b/chiebenthal/kimocc-co2-data-kiel-fjord) monitoring various environmental parameters at the GEOMAR. Additional measurements were taken using a handheld sensor (Conductivity portable meter ProfiLine Cond 3110 with Special conductivity measuring cells TetraCon® 325 S, WTW), which was used to verify KIMOCC data, but was not typically used. For nutrient data, seawater was frozen at -20 until the measurement of nitrite, nitrate, silicate, and phosphate. Samples were run in the central analytical laboratory at GEOMAR using a standard colorimetric, Continuous Flow Analyzer (Strickland and Parsons 1968). For graphing nutrient data, we excluded dates missing measurements, substituted values below the detection limit with the limit value itself, and

averaged replicates for consistency. The refined data was then plotted using R (R Core Team , 2022).

Second, the Boknis Eck time series (KBP; Kieler Bucht Project, 54°30'46.1" N 10°01'20.5" E) provided access to monthly resolution across the full water column of a representative southwest baltic location (entrance of Eckernförde Bay). Boknis Eck is one of the longest-running time-series stations around the world, contributing monthly depth-profiled samples. The deepest depth of this site is 25 m, and the samples that are used in this study are from 1 m, 10 m, 20 m, and 25 m for one and a half years (Jan 2022 to June 2023, except July 2022). The seawater was collected by a rosette sampler equipped with Niskin bottles and a conductivity-temperature-depth (CTD) sensor, stored in a cooler box, and brought back to the laboratory within 1 to 1.5 hours for sample processing.

Broader Baltic Sea samples were collected from two cruise campaigns, AL571 (Apr 2022) and AL580 (Aug 2022) on the research vessel ALKOR. Data for the vertical profile of relative dominance of AOA, AOB, and NOB (Nitrite-oxidizing bacteria) was collected during AL580 from Bornholm Basin (70m deep) and Gotland Basin (150m deep) by Niskin rosette sampler equipped with a CTD sensor. Each seawater sample was processed immediately on the vessel. Further details on the expeditions can be found on the PANGAEA database (https://www.pangaea.de/expeditions/bybasis/Alkor%20(1990)). All water samples from GPT, KBP, and broader Baltic Sea were vacuum-filtered under less than 200 mbar pressure through a 0.2 µm pore size filter (Pall, Supor™ 200 Membrane Disc Filters) and the filter was then stored at - 80°C until further processing for DNA extraction.

2.2 Enrichment and cultivation

For the enrichment and cultivation of archaea from the GPT and KBP samples, and during the two cruises, 10% of collected seawater was inoculated into archaeal medium (v/v) and incubated in the dark at room temperature or 10 °C. The recipe for the archaeal medium was adapted from the Oligotrophic North Pacific (ONP) medium (Santoro and Casciotti 2011), with several modifications: the use of 0.2 µm-filtered, aged Baltic Sea surface water collected from the GEOMAR pier, 100 µmol l^{-1} NH_4Cl, 1 ml l^{-1} chelated trace element solution, 15 µmol l^{-1} KH_2PO_4 and 50 µg ml^{-1} ampicillin. To reduce the diversity and abundance of bacterial cells, two selected strains were further supplemented with additional antibiotics, specifically streptomycin and kanamycin. Initially, non-autoclaved filtered seawater was utilized for the cultures, due to concerns about potential adverse effects on the media, such as altering organic content. However, as archaeal growth became more steady, the medium's seawater was subsequently autoclaved and filtered to ensure sterility.

Typically, after a period of several months, nitrite concentrations were monitored to indicate AOA growth in the cultures. The absence of nitrogen-oxidizing bacteria or ammonia-oxidizing bacteria was later verified through analysis of 16S rRNA amplicon data. To assess nitrite concentration, 150 µl of the culture, 0.1% NED, and 1% sulfanilamide were added in a flat-bottomed 96-well plate, with absorbance measured at 540 nm using a plate reader (Strickland and Parsons 1968). Cultures showing the production of nitrite—a product of ammonia oxidation—were subject to more frequent monitoring. Upon nitrite levels plateauing or exceeding 70 µM, 10% of the culture was transferred into fresh medium (v/v), and 1 to 2 mL of the culture was filtered through a 0.2 µm pore size filter (Supor™ 200 Membrane Disc Filters, Pall) at a pressure of under 200 mbar, then stored at -80 °C for subsequent DNA extraction.

2.3 DNA extraction for molecular diversity

DNA was extracted from each filter using the DNeasy Plant Kit (Qiagen) with a modified lysis protocol, as follows. Filters were thawed from -80°C storage and transferred into screw cap microtubes with mixed glass beads (0.1 mm and 0.5 mm). Lysis buffer from the kit was added, followed by three freeze-thaw cycles using liquid nitrogen and a warm water bath. Subsequent mechanical lysis was conducted via bead-beating using a Laboratory Mixer (MM 400, Retsch). The protocol from the supplier was then resumed from the chemical lysis step through to elution, using either molecular grade water or tris-EDTA (TE) buffer. The extracted DNA was stored at -80°C.

For most of the filters from GPT and KBP, to process numerous samples efficiently, lysates after mechanical lysis were transferred to a 96-well plate for subsequent steps using a multichannel pipette. To address the limitations of amplicon sequencing, which only shows the relative abundance of microbial communities, a known quantity of cells (ZymoBIOMICS Spike-in Control I (High Microbial Load), Zymo Research) was added to each sample for future comparative analyses. However, in this study, the sequencing reads originating from the spiked-in controls were excluded from the data analysis.

2.4 DNA library preparation and sequencing

The 16S rRNA gene amplicon libraries were made using dual-indexed custom-designed primers synthesized by Integrated DNA Technologies (IDT). The forward primer consisted of an Illumina P5 adaptor sequence (AATGATACGGCGACCACCGAGATCTACAC), a unique 8-base index with every index differing by at least two base pairs, a forward sequencing primer (ACACTCTTTCCCTACACGACGCTCTTCCGATCT), a random spacer of 7 random bases (NNNNNNN) to enhance diversity for the initial base-pair reads on the Illumina MiSeq, and the universal V4-V5 primers 515F (Parada et al. 2016). Similarly, the reverse primer included the Illumina P7 adaptor (CAAGCAGAAGACGGCATACGAGAT), a different 8-base index, a

reverse sequencing primer (GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT), a 7-base random spacer (NNNNNNN), and the universal V4-V5 primer 926R (Parada et al. 2016). Polymerase chain reactions (PCRs) were performed with the KAPA HiFi HotStart PCR Kit (Cat. KK2102, Roche), using 1 U of KAPA HiFi HotStart DNA Polymerase, 1× KAPA Hifi Fidelity buffer, 0.3 mM dNTPs, and 0.4 µM of each primer. For environmental bulk seawater samples, the concentration of DNA was quantified using the Quant-iT PicoGreen dsDNA Assay Kit, and each DNA sample was diluted to 1-2 ng μl^{-1} for the PCR, while DNA extract from enrichment cultures was directly added to a PCR mixture due to the low concentration. PCR thermocycling began with an initial denaturation at 95 °C for 5 m, followed by 30 cycles of denaturation at 98 °C for 20 s, annealing at 55 °C for 15 s, and extension at 72 °C for 1 m. It ended with a final extension at 72 °C for 5 m. PCR products were confirmed through agarose gel electrophoresis, including no-template controls, and then cleaned and size-selected via 0.8× (vol: vol) Ampure XP magnetic beads (Agencourt® AMPure® XP, Beckman Coulter). DNA concentration after purification was re-quantified, pooled for desired read counts, and checked for size distribution using a 2100 Bioanalyzer system (Agilent). In case there is a low molecular weight peak indicating the presence of primer dimers, the pooled DNA was purified again using 0.8× (vol: vol) of the magnetic beads.

2.5 Metagenome libraries construction and sequencing

To construct metagenome libraries cost-effectively, so-called Hackflex method (Gaio et al. 2022), a modification of the protocol of Illumina DNA Prep (M) Tagmentation Library Preparation was used. Metagenomic libraries were prepared from KBP samples, and cruise samples, as well as for genome sequencing from AOA enrichments. Given the low DNA yield from cultures, the concentration wasn't measured, but 1 µL of DNA extract was added to 50 µL of tagmentation mix. In the tagmentation step, bead-linked transposomes cut and tag DNA fragments. Amplification of tagged DNA used Phusion® Hot Start Flex DNA Polymerase, (New England Biolabs, M0535) and manually designed primers that consist of an Illumina adaptor, 8-base index, and overlapping sequence with transposome adaptor with an initial denaturation at 98°C for 30 seconds, followed by 12-16 cycles of: 98°C for 10 s, 62°C for 30 s, 72°C for 30 s, and 72°C for 5 m for a final extension. After PCR, products were checked by electrophoresis and underwent 0.6× purification and size selection via Ampure XP magnetic beads. The products were then re-evaluated using a bioanalyzer, followed by an additional 0.8× bead clean-up to remove short DNA fragments. Final quantification was conducted using the Quant-iT PicoGreen dsDNA Assay Kit, with pooling adjusted to achieve the desired sequencing depth.

All DNA libraries were sequenced at the Competence Centre for Genomic Analysis (CCGA) Kiel. Sequencing of amplicon libraries was performed on Illumina MiSeq with either 2 × 250 or 2 × 300 read lengths and environmental genomes (KBP569, BB04) were sequenced on Illumina NovaSeq_6000 2 × 150, while genomes from enrichment cultures (32a, 571-38, 54a,

68KS, 6a, 580-32, and 7KS) were also processed on MiSeq 2 × 250 or 2 × 300 (Table 1). All data was demultiplexed by CCGA providing individual fastq files for each sample.

2.6 16S rRNA gene amplicon data analysis

In the analysis of 16S rRNA gene amplicon data, the primers 515F and 926R were trimmed from raw reads using Cutadapt v4.4 (Martin 2011). The mixed amplicon sequences were then categorized into separate 16S and 18S pools with bbsplit.sh, part of the BBMap package v39.01 (Bushnell 2014). This categorization was based on the curated databases from SILVA 132 (Quast et al. 2013) and PR2 (Guillou et al. 2013) for accurate differentiation (Yeh et al. 2021). The subsequent analysis focused exclusively on prokaryotic reads, using QIIME2 v2022-2 and v2022-11 (Bolyen et al. 2019) for processing. Denoising was done by DADA2 within QIIME2 (Callahan et al. 2016), and truncation lengths for forward and reverse reads were determined based on sequence quality scores from FastQC v0.12.1 (Andrews 2010). Taxonomic classification was performed based on the SILVA 138.1 database (Quast et al. 2013) after trimming manually to just include the amplified region. Six individual sequencing runs were separately analyzed and subsequently merged. Chloroplast, mitochondrial, and unassigned reads were then removed using the taxa filter-table function in QIIME2 (Bolyen et al. 2019). The merged data was converted first into BIOM format and subsequently into TSV. The TSV file was imported into Excel for manual data curation. In Excel, reads originating from the spiked-in controls were manually identified and removed. Then, samples were categorized based on their collection source (GPT, KBP, cruises, and enrichments), and the relative abundance of each sample was calculated. Additionally, samples with low sequencing depth (< 900 reads) were filtered out.

2.7 Heatmap and abundance of each taxa

Initially, 16S rRNA gene sequencing data for each site were processed to calculate the relative abundance of ASV (amplicon sequence variant) by dividing the read counts for each ASV by the total read counts on each sample. The dataset was then filtered based on taxonomy at the family level. For AOA, only reads from the Nitrosopumilaceae family were retained. There were two reads from the Nitrososphaeraceae family in one sample, but since Nitrosophaeraceae was so rare, it was excluded. In the case of AOB, data were filtered to include families Nitrosomonadaceae and Nitrosococcaceae, while NOB data included families Nitrospinaceae and Nitrospiraceae. Rare ASVs for AOB and NOB were removed to focus on the more prevalent strains and the relative abundance of each ASV was visualized. Meanwhile, for the dataset of KBP, which has 1, 10, 20, and 25 meters, reads attributed to AOA, AOB, and NOB were separately summed up for each group without distinguishing between specific ASVs. The curated datasets were imported into R (R Core Team, 2022), and three separate heatmaps for AOA, AOB, and NOB were produced by ggplot2 (Wickham

2016). These heatmaps were combined and adjusted in Adobe Illustrator. Due to the absence of specific names for all strains, the hash codes of each ASV were utilized as labels.

2.8 Correlation between relative abundance and environmental parameters

To analyze the correlation between the relative abundance of each microbe and environmental parameters, the GPT data was refined by excluding samples lacking nutrient measurements or 16S rRNA data, and missing temperature data were interpolated. The Hmisc package in R was employed to compute the Spearman correlation across the data matrix, which was then visualized as a heatmap, incorporating p-values to assess the statistical significance of the correlations observed.

2.9 Genome assembly and binning

Raw paired-end reads were processed using Trimmomatic v0.39 (Bolger et al. 2014) for quality trimming. Reads at the first base with a quality score below 3 were truncated, followed by further truncation if the moving average quality score across 25 base pairs fell below 30 (settings: LEADING:3 TRAILING:3 SLIDINGWINDOW:25:30 MINLEN:50). The resulting paired reads were then assembled with SPAdes v3.13.0 (Bankevich et al. 2012), employing a spectrum of k-mer sizes (21, 33, 55, 77, 99, 127) and with '-sc' option. For downstream analysis, only contigs exceeding 5,000 base pairs were retained. These contigs served as the basis for creating Bowtie2 v2.3.5 (Langmead and Salzberg 2012) databases for each respective sample. Following database construction, the quality-trimmed reads were aligned back to their corresponding contigs using Bowtie2. The binning process was conducted using the Anvi'o v7.1 interactive interface (Murat Eren et al. 2015; Delmont and Eren 2018; Eren et al. 2021), leveraging information such as tetranucleotide frequency, GC-content, and contig coverage. The quality of the resulting bins was evaluated using CheckM v1.2.2 (Parks et al. 2015). Out of various archaeal bins, 9 high-quality bins (minimum completeness 89.32% and minimum redundancy 2.27%) were chosen. Their taxonomic classification, performed using GTDB-TK v2.1.1 (Chaumeil et al. 2019), identified them all as belonging to the genus *Nitrosopumilus*.

2.10 Selection and curation of reference genome dataset

To acquire a comprehensive set of reference genomes for phylogenetic analysis, accession numbers for all representative *Nitrosopumilus* genomes were retrieved from GTDB (Parks et al. 2020) using GToTree v1.8.1(Lee 2019). These genomes, along with *Nitrosopelagicus brevis* as an outgroup, were downloaded in fasta format using Bioinformatics Tools (bit) v1.8.61 (Lee 2022). Each reference genome underwent a quality assessment using CheckM v1.2.2(Parks et al. 2015) to ensure a minimum of 80% completion and less than 5% redundancy. The final dataset encompassed 64 genomes, including 54 *Nitrosopumilus*

representatives from GTDB, 9 from this study, and the *N. brevis* outgroup (Supplementary Table 2). Genome metadata, including sources and sampling locations, was collated from GTDB, NCBI, and CheckM results. For incomplete genomes, estimated genome sizes were calculated using the methods described by (Nayfach et al. 2019).

2.11 Phylogenomic analysis and functional gene annotation

The construction of the phylogenomic tree was utilized by GToTree v1.8.1 (Lee 2019) to select a single-copy genes set (76 target genes) for archaea. Gene prediction on fasta files was performed with Prodigal v2.6.3 (Hyatt et al. 2010), and target genes were identified using HMMER3 v3.3.2 (Eddy 2011), aligned individually with Muscle 5.1.linux64 (Edgar 2004), and then refined with trimAl v1.4.rev15 (Capella-Gutiérrez et al. 2009) before concatenation. The phylogenetic tree was estimated using IQ-TREE (Kalyaanamoorthy et al. 2017; Hoang et al. 2018; Minh et al. 2020), supported by a 1000 bootstrap value. The tree was visualized with R 4.2.2 using ggtree v3.6.2 (Yu et al. 2017) and other tools to process tree data (Yu 2022; Wickham 2023). For functional gene annotation, selected gene sets (amoABC, uvrABC, ectABC and thpD, pstABC, ureABCDEFG and cheCBRWY, and flaB) were manually identified from various *Nitrosopumilus* species, with gene IDs manually gathered from NCBI and UniProt (Supplementary table 3). These genes were downloaded using the 'ncbi datasets' tool and analyzed using NCBI BLAST+ (Camacho et al. 2009). BLAST results were filtered (e-value > 1e-5 and bitscore > 100) and summarized in a gene presence-absence table. and gene patterns and metadata visualized using (see below) using ggplot2 (Wickham 2016). The alignment of these elements next to the tree was achieved using the aplot package.

2.12 Pangenomic analysis

To compare all genomes within the genus *Nitrosopumilus*, we employed OrthoFinder v2.5.5 (Emms and Kelly 2019), Anvi'o v8 (Delmont and Eren 2018; Eren et al. 2021), and FastANI v1.33 (Jain et al. 2018). Both OrthoFinder and Anvi'o perform similar functions by identifying gene similarities across genomes through sequence alignment tools and clustering homologous genes. However, OrthoFinder was specifically used to analyze the presence/absence of orthologs (sets of similar genes) across genomes and to correlate genomes based on their gene presence/absence patterns. Anvi'o, on the other hand, was utilized to visualize the extent of gene cluster conservation or variability within each *Nitrosopumilus* genome, offering the advantage of incorporating additional metadata for an intuitive overview of large datasets.

For OrthoFinder analysis, Prodigal (Hyatt et al. 2010) was initially used to predict proteins from each genome, generating files with the DNA sequences of predicted genes. These files were then inputted into OrthoFinder, where gene sequences were aligned using DIAMOND

(Buchfink et al. 2015) and clustered using the MCL (Markov Clustering) algorithm (van Dongen and Abreu-Goodger 2012) with an inflation value of 1.5. The gene count per ortholog for each genome generated by OrthoFinder was used to establish Pearson correlation between genomes.

In parallel, Anvi'o analysis involved creating a database for each genome and annotating some important gene sets (rRNA and single-copy core genes of taxa) using built-in databases, with functions assigned using the Clusters of Orthologous Genes (COGs) database (Tatusov et al. 2000). Protein sequences were aligned using BLASTP (Camacho et al. 2009), and clustered to create gene clusters by MCL (van Dongen and Abreu-Goodger 2012) with an inflation value of 10, indicating a finer division of clusters. Anvi'o's interactive interface for pangenomics (Delmont and Eren 2018) provided a detailed view of which genes are conserved or variable across the genomes.

3. RESULTS

3.1 Environmental distribution and diversity of AOA/AOB/NOB

3.1.1 Kiel Fjord time-series (Geomar pier time-series; GPT)

In order to examine the dynamics and diversity of the taxa involved in nitrification in the Baltic Sea, data from 1.) twice-weekly surface Baltic Sea samples (GEOMAR Pier Time-series; GPT, Kiel Fjord), 2.) monthly depth-integrated time-series analysis (Kieler Bucht Project; KBP, Eckernförde Bay, also known as Boknis Eck), and 3.) Baltic Sea biogeography cruises were utilized (Fig. 1). To accomplish this, the cellular size fraction (>0.2 µm) was collected, extracted, and sequenced with universal rRNA gene amplicon sequencing. Our results revealed that AOA are among the dominant taxa present in surface waters in the late fall and early winter and throughout the deep baltic sea.

The twice-weekly GPT time-series revealed strong seasonal dynamics in AOA with peaks in relative abundance in late fall-early winter and decreased relative abundances in spring and summer (Fig. 2B). This successional pattern also had strong correspondence to AOB, and NOB, where AOA and AOB peak first (late fall, early winter), followed by NOB (mid-winter).

Among the AOA in the Kiel Fjord, distinct strains of *Nitrosopumilus* exhibited significant variations in dominance (Fig. 2B). During the first year of observation, AOA presence increased in late October, reaching a peak in November, then decreased before peaking once again in late December. In the subsequent year, the dominance of AOA was less marked than the previous year. While it is premature to conclusively attribute these patterns to a certain environmental parameter, given the limited two-year dataset, a consistent pattern was observed, with AOA showing two separate, apparent peaks each year in fall-winter. Specifically, the b9f strain, the most dominant among them, exhibited an average relative abundance that was 60-fold greater than that of the second most dominant strain (Fig. 3A).

Like the AOA, the AOB population showed an increase during the winter, almost concurrently with AOA but slightly delayed, albeit one-tenth of relative abundance compared to AOA (Fig. 2B). In contrast to AOA, AOB were consistently observed in the summer and a clear succession of different AOB strains over time was observed, with three different ASVS (a91, 53a, and 32f) becoming the most relatively abundant ASV at some point (Fig. 3B). Alongside 53a, a fourth, more rare taxon, fb5, also became relatively abundant. Notably, the AOB strain 32f was present from spring to early summer at the surface water of the GEOMAR pier station and then became undetected, yet it persisted in the deeper waters of the broader Baltic Sea as observed in September (Fig. 6).

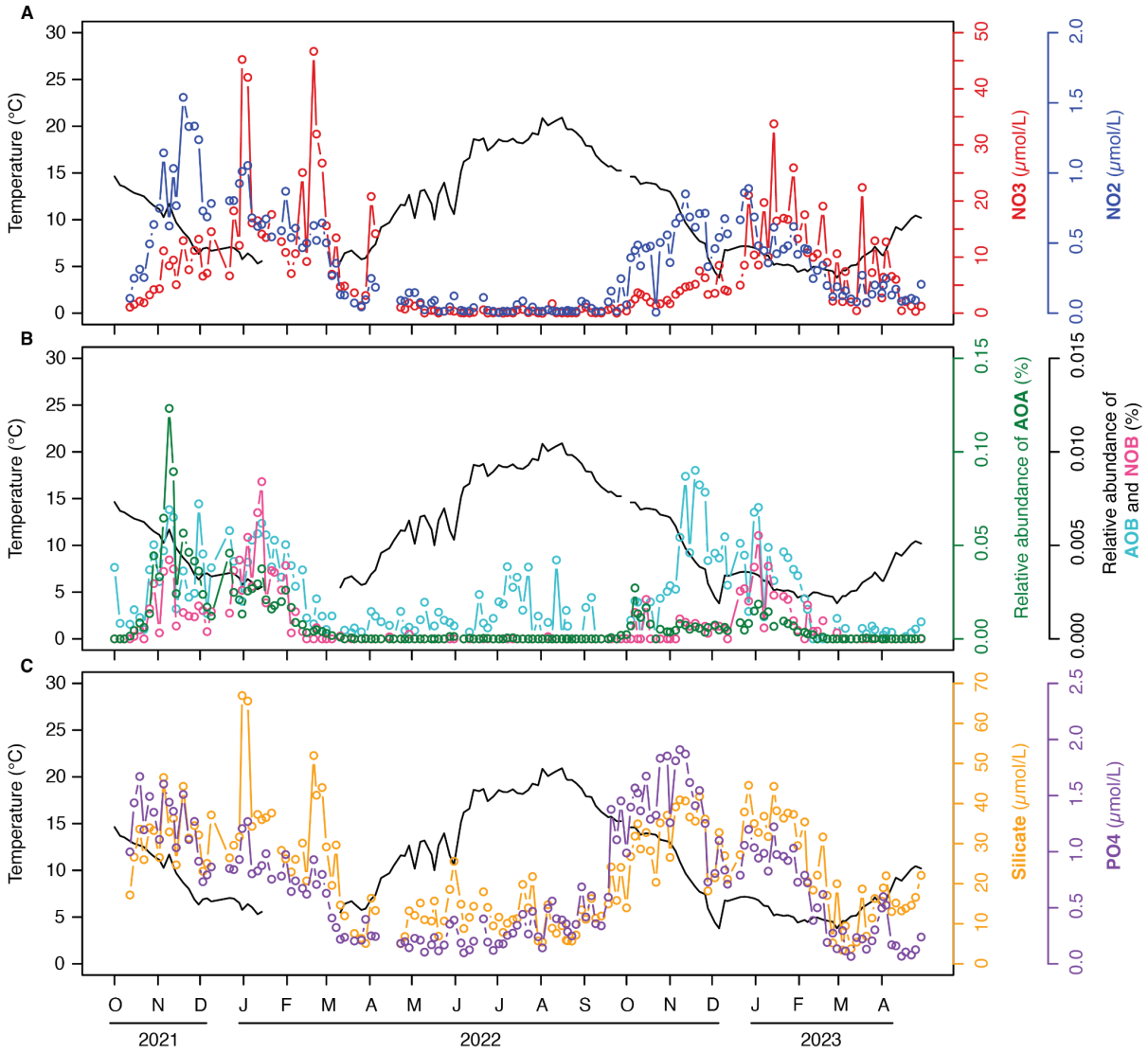


Figure 2: Seasonal variations of key nitrifying microbes nitrogen cycle microbial communities (AOA, AOB, and NOB) and environmental parameters from GEOMAR pier time-series (Fig .1). The top panel displays nitrate (NO_3, red) and nitrite (NO_2, blue) concentrations, the middle panel shows the relative abundance of AOA (green), AOB (cyan) and NOB (magenta), and the bottom panel depicts silicate ($Si(OH)_4$, yellow) and phosphate (PO_4, purple) concentrations, with the temperature (black line) superimposed on each panel for reference. Data points represent measurements taken bi-weekly over a period spanning from October 2021 to April 2023.

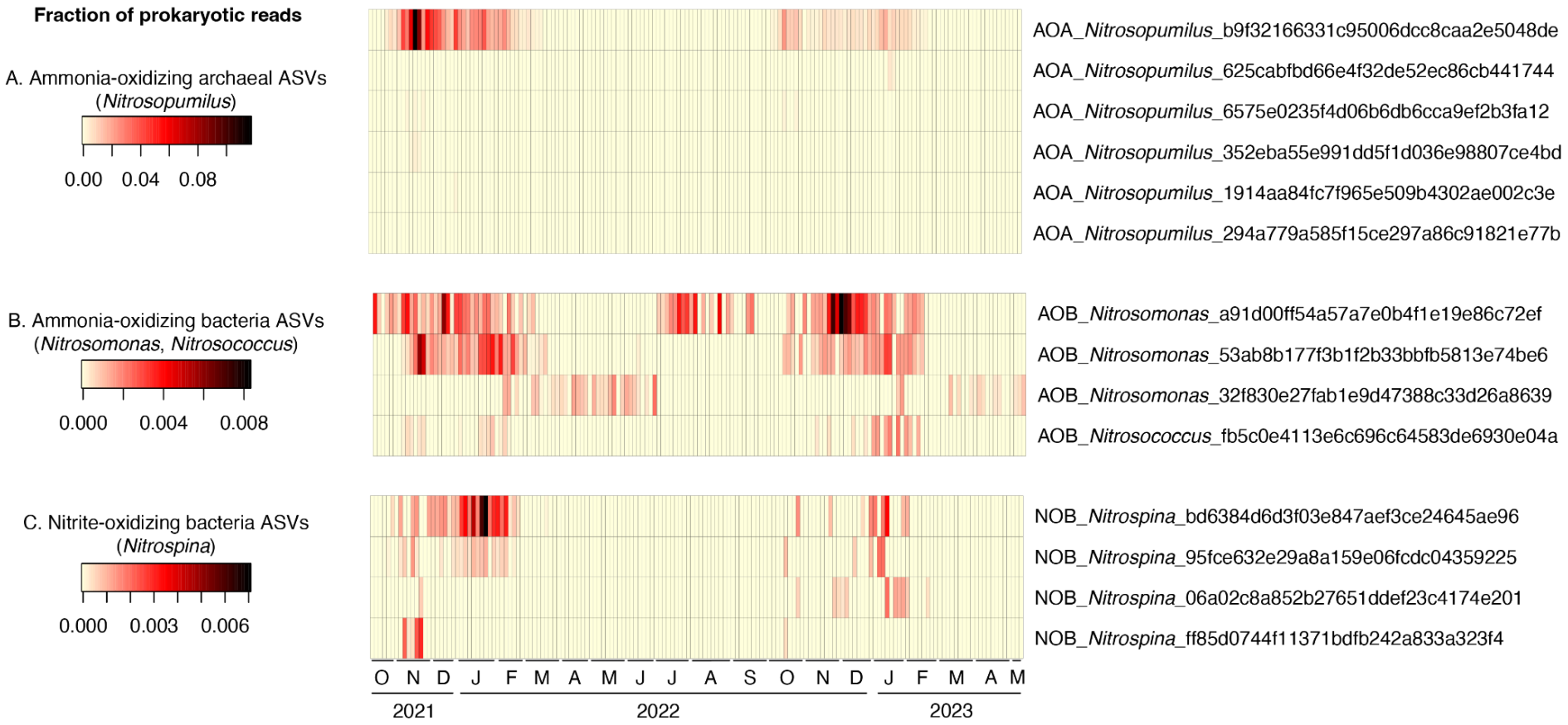


Figure 3: Dynamics of AOA, AOB, and NOB at an ASV level from the GEOMAR pier time-series. To facilitate comparison, read counts for each ASV across samples were converted to ratios. Different ASVs are distinguished using unique identifiers (hashes). Data presented here focus on the predominant ASVs, as rare AOB and NOB ASVs were excluded from this analysis.

In contrast to AOA and AOB, NOB reached maximal relative abundances in mid-Winter (i.e., later than AOA and AOB), though the initial increases in relative abundance coincided with AOA and AOB (Fig. 2B). Similarly to AOA and AOB though, NOB seemingly had two annual peaks with the difference being that the second peak, in the case of NOB, was larger. Additionally, similar to AOA, NOB disappears from the surface water during the warmer seasons. Interestingly, at the ASV level, NOB differs as there isn't a single dominant ASV like AOA, and it doesn't exhibit a succession of different strains, unlike AOB (Fig. 3B). Instead, all NOB strains appear simultaneously.

The dynamics of the AOA, AOB, and NOB were highly coupled to environmental parameters such as temperature and inorganic nutrients like nitrite and nitrate, which are intimately connected to their metabolic activities (Fig .2A). Similar to AOA and AOB, nitrite—a by-product of ammonia oxidation—increased in early fall, reaching maximum concentrations in mid-November for both observed years, before subsequently declining. This decrease in nitrite corresponds with a rise in nitrate concentrations, which peaks in mid-winter (approximately late January to mid-February), aligning with the increase in NOB. This nitrate peak slowly decreases from February towards, with both nitrite and nitrate concentrations significantly dropping throughout the summer and at times even falling below detectable limits.

Unlike nitrite and nitrate, silicate (SiO_4) and phosphate (PO_4) have early fall peaks (Fig. 2C). Notably, there are strong, pulsed peaks in concentration of nitrate and silicate, particularly in January and March of 2022. These peaks seem to be overlaid on the seasonal trends without altering the seasonal patterns of increase and decrease. Such observations suggest these spikes may result from brief, localized nutrient inputs, possibly from freshwater input from the Schwentine River, which flows into Kiel Fjord, or from coastal runoff in the Kiel area that is rich in these nutrients (Voss et al. 2011).

The time series plot visually demonstrates the interplay between microbial dynamics and environmental parameters, prompting us to further investigate these relationships through statistical analysis. Employing Spearman correlation analysis, we found that the relative abundance of the dominant AOA strain negatively correlates with temperature while it shows a strong positive correlation to inorganic nutrients, particularly nitrite (Fig. 4). Similarly, the most abundant AOB strain, a91, shows a positive correlation with inorganic nutrients like the dominant AOA. However, unlike AOA, this strain does not correlate with temperature likely because it appears during the summer as well (Fig. 3B). Another AOB strain, 32f, which is present from late winter to early summer (Fig. 3B), shows a negative correlation to the dominant AOB ASV that indicates clear niche differentiation among AOB strains in the environment (Fig. 4). NOB strains, appearing alongside AOA, positively correlate with other taxa and negatively with temperature, but not as strongly as AOA, possibly due to NOB being more influenced by nitrite and the nitrite-producing AOA.

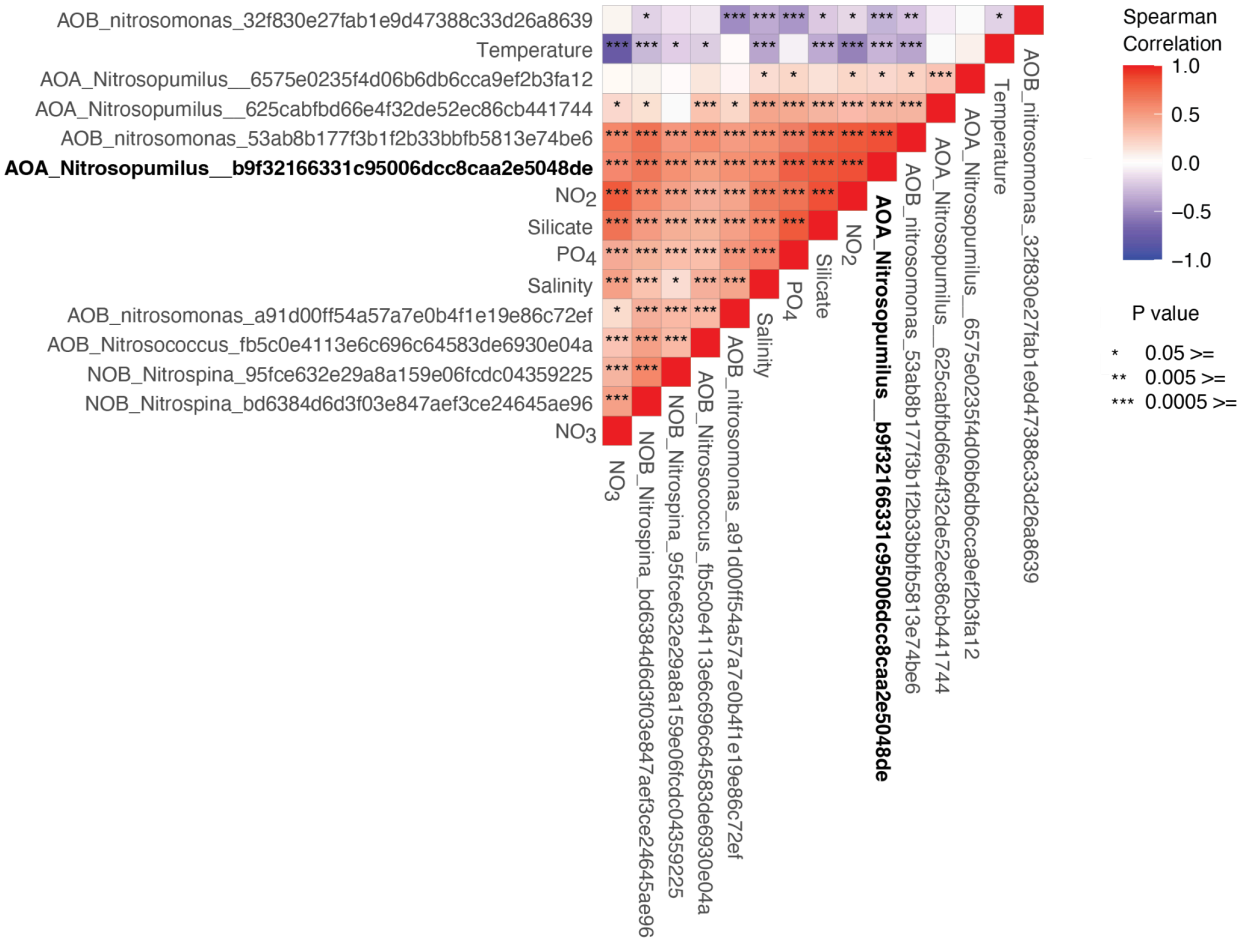


Figure 4: Spearman correlation heatmap showing the relationship between ASVs of nitrifiers (AOA, AOB, and NOB) and various environmental parameters from GEOMAR pier time-series station. The most dominant southwest Baltic Sea surface AOA ASV is highlighted in bold. Each cell represents the correlation coefficient indicated by the color gradient, with the asterisks denoting the significance levels of the correlations.

3.1.2 Boknis Eck Time-series (Kieler Bucht Project; KBP)

In addition to the high temporal resolution at the GEOMAR pier time-series, the abundances of AOA, AOB, and NOB were also examined at monthly resolution across the full water column at the Boknis Eck time-series station (Fig. 1).

Similar to the pattern observed at the GPT site, the surface abundance of AOA begins to rise in October, reaches its peak during the winter months, and then declines with the onset of spring (Fig. 5A). However, at the KBP site, AOA constitute a smaller proportion of prokaryotic reads on the surface—about 2%, in contrast to nearly 10% observed at the Kiel Fjord during the winter of 2022. Deeper in the water column, AOA represent about 7% of prokaryotic reads, corroborating established findings that AOA are more abundant in deeper waters (Karner et al. 2001). Their presence is significantly reduced in the second year compared to the first. This trend corresponds with the GPT data, suggesting a potential correlation with yearly variations in the region (Supplementary Figure 1). Additionally, a peak in bottom water was noted for all three groups during the summer, but with only one and a half years of data, it is premature to determine whether this is a recurring pattern. Despite these annual fluctuations and quantitative differences observed when compared to GPT, a consistent pattern is observed in the southwest Baltic Sea: AOA populations expand during the winter across the entire water column and decrease with the onset of spring.

On average, AOB exhibit a less pronounced seasonal pattern compared to AOA, remaining present during summer at the surface, a trend also noted at Kiel Fjord. However, with only one summer season captured in the one and a half years of data, observation for a second summer is necessary to verify this pattern. Interestingly, in December 2022, AOB were more prevalent than AOA in deeper waters at depths of 20 and 25 meters, though AOA remained more abundant at the surface. As for NOB, their occurrence seems to coincide with high AOA abundance. In the second year, NOB did not reach the levels of abundance observed in the first year like AOA did.

The strain diversity of AOA, AOB, and NOB at both sites is considerably similar (Fig. 7 for AOA). While not depicted in a figure, AOB and NOB at these sites share the same ASVs as well. At the ASV level, AOB display similar patterns as at GPT, yet discerning a clear succession of different AOB ASVs through monthly sampling is challenging.

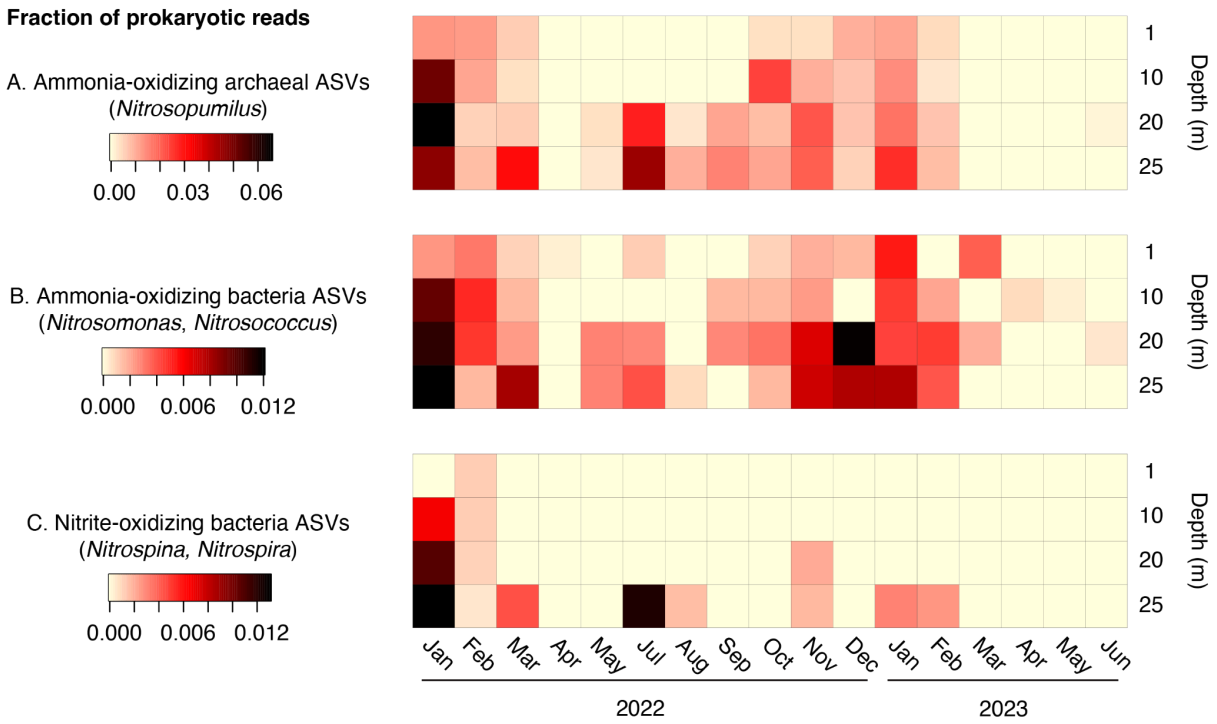


Figure 5: Temporal and depth distribution AOA, AOB, and NOB from January 2022 to June 2023 at the Boknis Eck time-series station. Due to the complexity of representing depth, time, and individual strain variations within a single plot, all strains within each group were combined for analytical purposes.

3.1.3 Broader Baltic Sea Biogeography and depth distributions

In addition to observations in the southwest Baltic Sea, two cruise campaigns in September 2022 enabled a detailed examination of the distribution of AOA, AOB, and NOB within the deeper water columns (Fig. 6). These microorganisms demonstrated a higher prevalence in deeper waters, with a notable dominance in the Bornholm Basin over the Gotland Basin. AOA populations were rare on the surface during the fall, mirroring patterns observed in the southwest Baltic Sea, but were prevalent in deeper waters (Fig. 6A). Notably, the newly dominant strain 897 was identified in these deep waters. This strain, which was successfully cultured in the laboratory, accounted for 15% of prokaryotic reads in the Bornholm Basin and was also present at a depth of 150 meters in the Gotland Basin.

Unlike AOA, AOB populations were more abundant in the middle water mass and seldom appeared at the bottom, with one exception in the Bornholm Basin (Fig. 6B). Conversely, NOB's presence closely matched that of AOA, with several NOB ASVs appearing simultaneously and in the same locations as AOA (Fig. 6C).

In temporal terms, AOA in the Baltic Sea become more abundant throughout the whole water column during winter, a pattern that, despite varying in intensity, seems to recur annually. Even in other seasons, AOA constitute a significant part of the prokaryotic community in deeper waters, where sunlight does not reach.

There are diverse AOA strains dwelling in our study region, yet when AOA become dominant, a single ASV often emerges as the most prevalent. This is reflected in the clear shift within the *Nitrosopumilus* community from the southwest towards the middle Baltic Sea (Fig. 7). In the southwest Baltic Sea, the ASV b9f accounts for 96% of all *Nitrosopumilus* reads at GPT, and 88% at KBP, but this proportion decreases to 15% and eventually to 0% as it extends into the middle Baltic Sea. In contrast, ASV 897 stands out in the mid-southern deep Baltic Sea, representing 65% of Nitrosopumilus reads in the Bornholm Basin and increasing to 78% in the Gotland Basin. A less diverse AOA and AOB community has also been observed in the Gulf of Finland according to a local study (Vetterli et al. 2016). These findings suggest a distinct niche differentiation for AOA across different regions of the Baltic Sea.

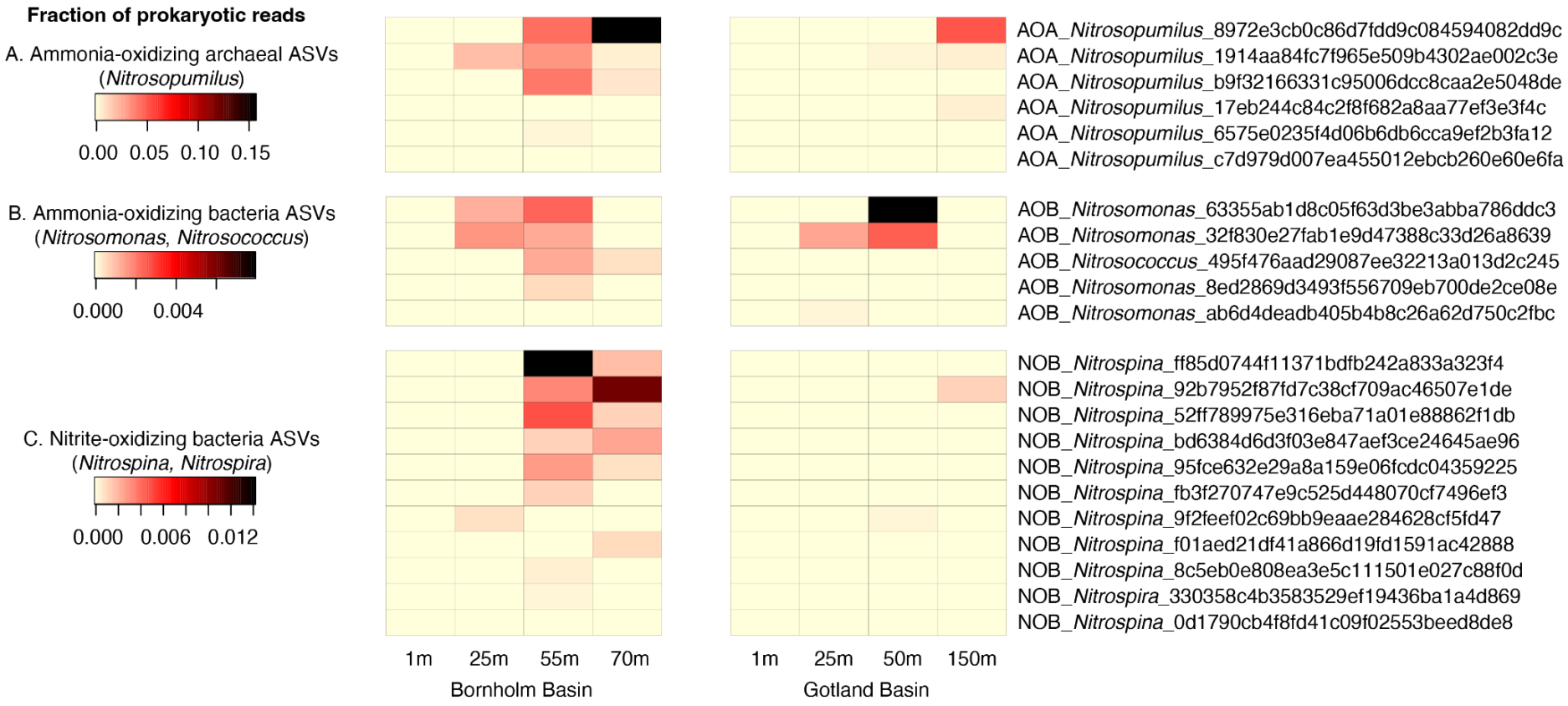


Figure 6: Distribution of AOA, AOB, and NOB at ASV level in Bornholm Basin and Gotland Basin, located in the mid-south Baltic Sea (Fig. 1), sampled in early September 2022.

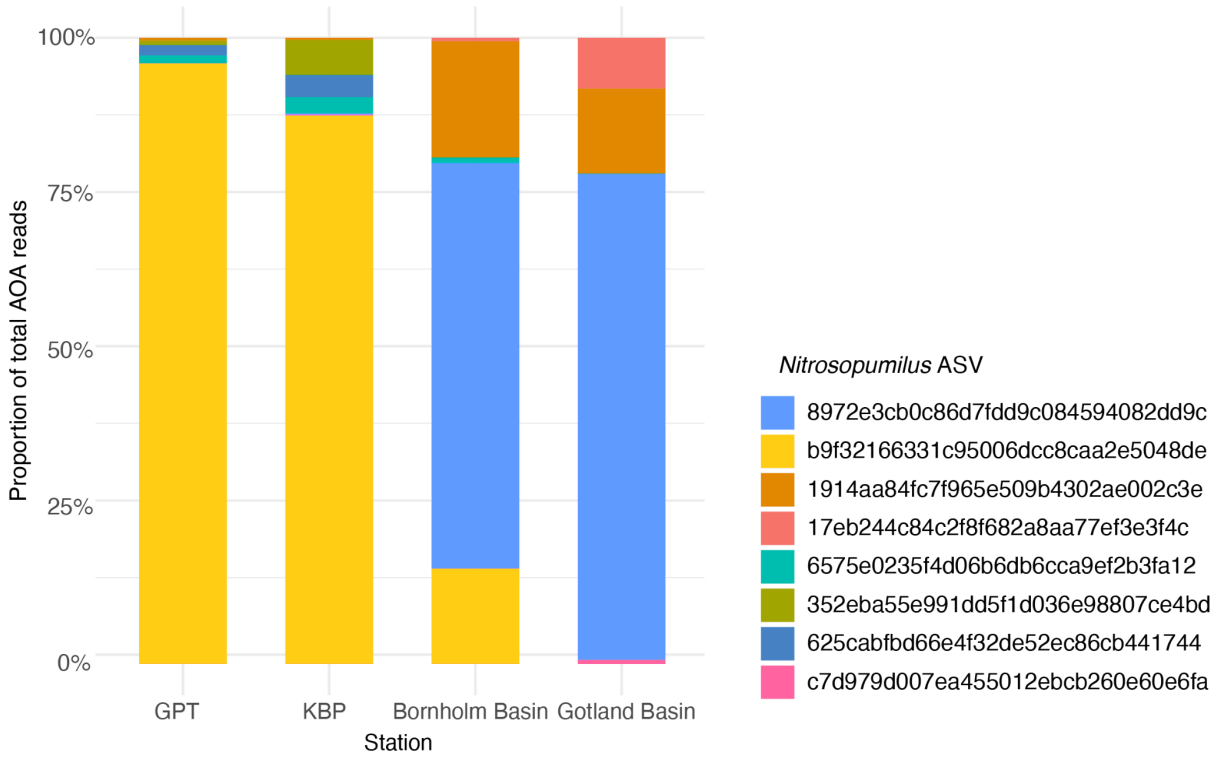


Figure 7: Proportion of *Nitrosopumilus* ASV relative to total AOA reads from various sampling stations. Each bar represents the percentage composition of *Nitrosopumilus* ASVs within the AOA community across different locations (Fig. 1).

3.2 Novel enrichment and genome characterization of AOA

3.2.1 Description of enrichments

Our studies on time-series and biogeographic distributions emphasized the diversity and importance of AOA, and nitrification across the Baltic Sea. This prompted a detailed investigation into AOA to enrich and characterize diverse strains and species, given that the databases are under-represented in these groups, especially from the Baltic Sea, with so far only one enrichment (Berg et al. 2014).

This research successfully enriched several Baltic Sea AOA strains, resulting in nine AOA genomes—seven through enrichment-enabled genomics and two via metagenomic sequencing of bulk water samples (Table 1). These genomes include the most dominant strain from the southwest Baltic Sea (at GPT and KBP), the prevalent deep sea strain, and other distinct strains. For the enrichment, archaea media made of natural Baltic Sea water was amended with ammonium, antibiotics, and other limiting compounds (see methods). Enrichment cultures were initiated from seawater from the GEOMAR Pier, Boknis Eck time series station, and deep Baltic Sea, incubated at room temperature or 10°C. The growth was

monitored by an accumulation of nitrite attributed to AOA activity confirmed by sequencing that showed enrichments ranging from 5% to 80% AOA presence (Fig. 8)

Each of the strains is denoted by names reflecting their isolation origin or timing. Cultures from the GPT and KBP sites were labeled 6a, 7KS, 32a, 54a, and 68KS, with the 'KS' suffix indicating supplemented antibiotics, kanamycin, and streptomycin. Two cultures from cruises: 571-38 and 580-32 were named by combining the cruise number with the sample's sequential number. Two additional genomes directly obtained from the environment were named after their collection sites: BB04 from the Bornholm Basin and KBP569 from the Kieler Bucht Project.

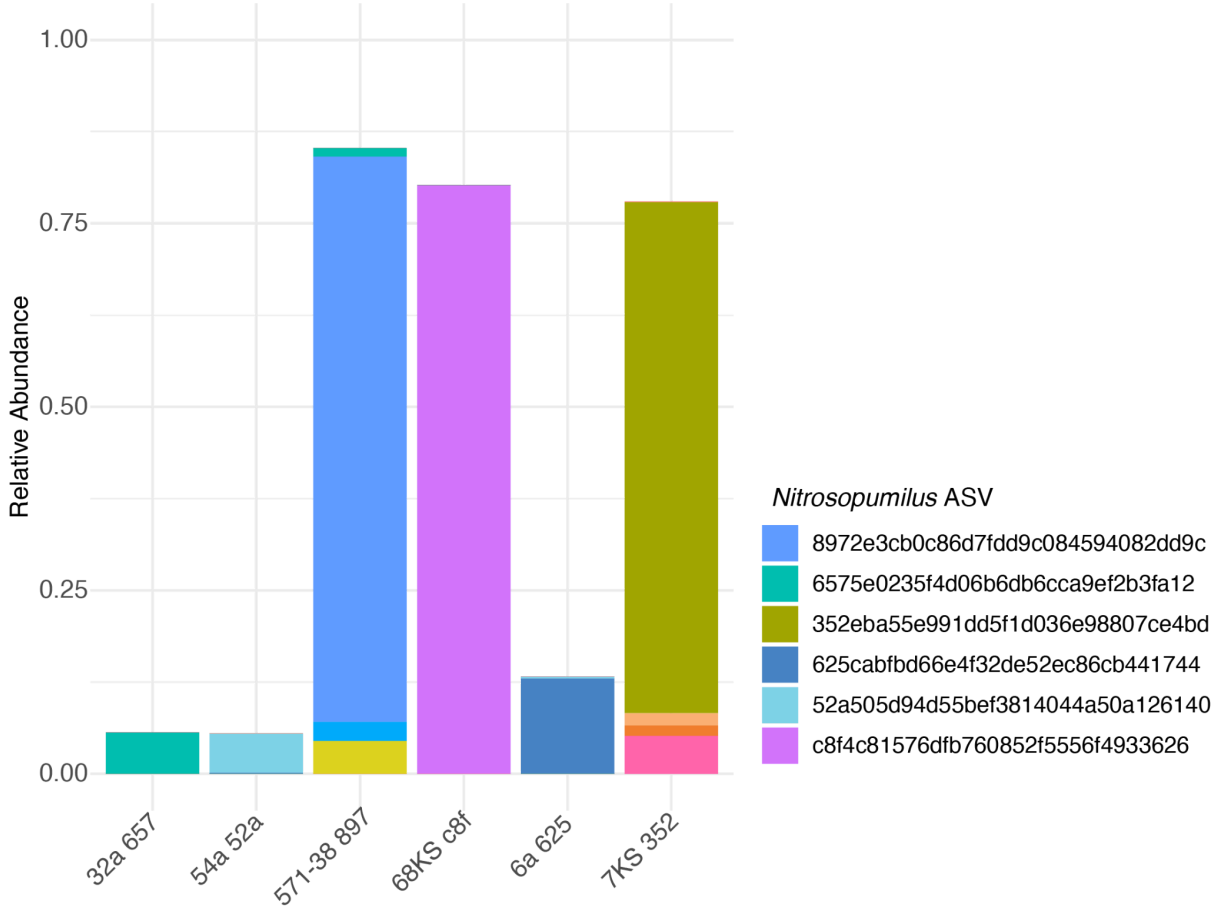


Figure 8: Percent of relative abundance of AOA in enrichment cultures derived from the Baltic Sea based on 16S rRNA gene sequencing. These samples are collected at one time-point prior to transfer and are also used for metagenome sequencing. The legend includes the top six *Nitrosopumilus* ASVs. Note: The enrichment culture 580-32 is not included in this figure due to unsuccessful 16S rRNA amplicon library preparation.

3.2.2 Relationship of strains to ASV environmental analysis

To connect the environmental data with newly sequenced AOA genomes, we linked MAGs to environmental ASVs by appending a three-character hash from corresponding ASVs to the names of cultured strains sharing the same 16S rRNA sequences (Table 1). For instance, an archaeal genome from the culture 571-38, matching the ASV hash starts with '897-' was named '571-38 897' for ease of data analysis. For genomes without matching ASVs, a hash from their 16S rRNA sequence was used. In this way, it would be easier to associate between genomic and environmental observations.

3.2.3 Genome sequencing, binning, and statistics of unique AOA strains

The genomes of AOA strains were sequenced by MiSeq and NovaSeq platforms, assembled individually, and binned computationally based on coverage and tetranucleotide frequency to distinguish an archaea strain from the co-occurring bacteria. The environmental MAGs were binned similarly but based also on the mapping of environmental samples from other time points.

High-quality genomes were obtained with completion rates ranging from 89.32% to 99.51% and redundancy estimates ranging from 0% to 2.27% (Table 1). The number of contigs ranged from 11 to 75 with estimated genome size ranging from: 1.28 to 1.86 Mb and homogeneous GC content ranged from 32.4% to 33.5%. The nine genomes acquired were all assigned to the genus *Nitrosopumilus* based on classification with GTDB-Tk (Chaumeil et al. 2019). Among these genomes, seven were not assigned to any species within the GTDB database based on Average Nucleotide Identity (ANI), indicating that these strains represent novel members of the genus *Nitrosopumilus*. ANI values calculated for all obtained genomes compared to reference genomes are detailed in Supplementary Table 1.

3.2.4 Description of genomes

The dominant Baltic Sea strain KBP569 b9f, representative of the southwest winter surface community, was obtained from the environmental metagenome of the Boknis Eck time series station at a depth of 25m despite its resistance to cultivation. This genome shows a 99% ANI with a GTDB reference genome from suspended organic matter of the Black Sea's western gyre, suggesting a shared ecological niche despite geographical separation. Given the stringent criterion for species delineation (>95% ANI), KBP569 b9f does not constitute a novel species. However, it implies the adaptation of this AOA strain to the brackish and low-oxygenated environment.

The 571-38 897 genome, sharing its 16S rRNA sequence with a predominant deep Baltic Sea strain, was assembled from an enrichment culture incubated at 10°C, originating from 63 m deep in the Bornholm Basin. Attempts to cultivate it at room temperature were

unsuccessful, indicating apparent preference for cold temperatures. Another genome from the environment, BB04, recovered from the Bornholm Basin in September, shows a 99% ANI match with 571-38 897, though it lacks a 16S rRNA sequence. Their closest relative (GCA_013203245.1), found in a meromictic Arctic lake in Ellesmere Island, Canada, shares 92% ANI with them, underscoring these strains' adaptation to brackish, cold, and hypoxic environments (Vigneron et al. 2023).

Beyond the dominant AOA strains, enrichment cultures allowed us to obtain genomes of four scarce strains in the environment. Genomes 580-32 657 and 32a 657 share 98% ANI between them and correspond to the ASV 657- found across the southern Baltic Sea yet remain very rare (Fig. 7). Their closest genome is also the Arctic lake genome previously mentioned, sharing 93% ANI with it. Another genome, 7KS 352 related closely with *N. oxyclinae* from brackish waters of Puget Sound, exhibiting a 93% ANI. Its ASV comprises 5% of all AOA reads at KBP and 0.6% at GPT (Fig. 7). Lastly, the 6a 625 genome, closest in ANI (87%) to *N. sediminis* from an enrichment culture derived from marine sediment off the coast of Svalbard, represents an ASV that is rarely detected at both GPT and KBP sites.

Among the strains characterized, two strains 54a 52a and 68KS c8f were not detectable in any of our environmental sampling, yet they were enriched in cultures. 54a 52a, shows the closest genetic affinity to a MAG from the Sapelo Island, GA, USA, north American coastal water. Notably, 54a 52a harbors a urease gene cluster, suggesting diverse nitrogen utilization capabilities, not observed in the other strains from this study. Another strain 68KS c8f, closest to *N. adriaticus* with an 89% ANI, has the biggest genome among nine genomes possessing partial flagella and chemotaxis-related genes indicative of potential motility and environmental responsiveness. Despite genomic traits that could confer adaptive advantages, these strains remain exceedingly rare, undetected in 16S rRNA surveys of the Baltic Sea, and only making their presence known within enrichment culture. These findings reveal the complexity and adaptability of marine AOA in the Baltic Sea, contributing valuable genomic data to the ecology of AOA.

Genome	32a	54a	68KS	6a	7KS	571-38	580-32	BB04	KBP569
ASV hash	657	52a	c8f	625	352	897	658	no 16S rRNA	b9f
Sampling site	KBP	GPT	GPT	GPT	GPT	Bornholm Basin	Gotland Basin	Bornholm Basin	KBP
Depth	5 m	Surface	Surface	Surface	Surface	63 m	20 m	55 m	25 m
Sampling date	02. Feb 22	12. Nov 21	10. Dec 21	03. Dec 21	10. Dec 21	27. Apr 02	03. Sep 22	05. Sep 22	02. Feb 22
Enrichment / Metagenome	Enrichment	Enrichment	Enrichment	Enrichment	Enrichment	Enrichment	Enrichment	Metagenome	Metagenome
Estimated genome size (Mb)	1.62	1.33	1.86	1.53	1.70	1.38	1.56	1.33	1.28
GC content (%)	33.4	32.9	32.6	33.3	32.9	32.5	33.5	32.4	32.5
Completion (%)	99.03	89.32	97.09	100	99.51	97.09	99.03	90.78	96.12
Redundancy (%)	0.97	0	0.97	0.97	0.97	0	0.97	0	2.27
Contig count	12	75	36	11	11	40	11	50	67
No. of expected proteins	1978	1459	2247	1922	2096	1670	1892	1552	1625
Sequencing platform	MiSeq	MiSeq	MiSeq	MiSeq	MiSeq	MiSeq	MiSeq	NovaSeq_6000	NovaSeq_6000
Read length	2x300	2x300	2x300	2x300	2x300	2x250	2x300	2x150	2x150
Note		Urease	Archaella			The dominant deep Baltic Sea strain			The dominant southwest-Baltic Sea strain

Table 1. Genomic characteristics and sampling information of AOA strains derived from the Baltic Sea. Genome names are paired with ASV hashes to correlate with environmental data in the following analysis. KBP; Kiel Bucht Project, GPT; Geomar Pier Time-series.

3.2.5 Phylogenomics and relationship to representative AOA

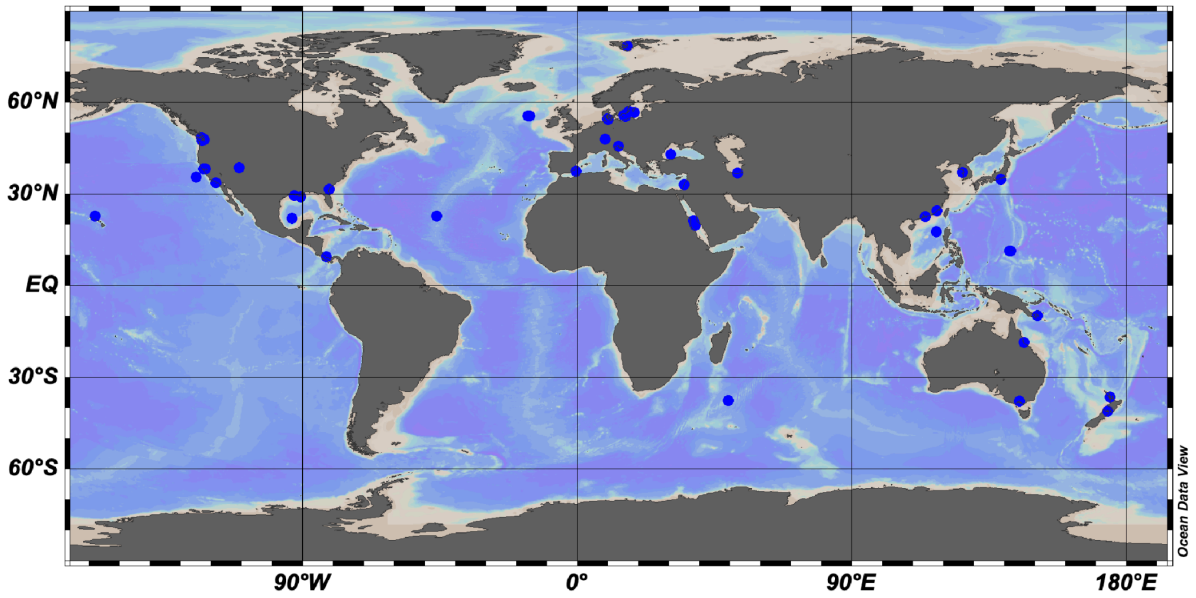


Figure 9: Global distribution of representative *Nitrosopumilus* genomes. Metadata for each genome was retained from the GTDB, NCBI, and reference papers (details in Supplementary Table 2)

Our phylogenomic analysis incorporated 55 representative *Nitrosopumilus* genomes from the GTDB, comprising 39 MAGs, 15 derived from isolate/enrichment cultures, and one SAG. The origin of genomes are diverse—29 from water columns, 11 from sediments, and 12 host-associated—predominantly from Northern Hemisphere coastal regions (Fig. 9). Within the phylogenomic tree, the nine AOA genomes from the Baltic Sea display considerable diversity, spanning several branches of the *Nitrosopumilus* lineage (Fig. 10). All *Nitrosopumilus* genomes present in our analysis range in size from 1.2 to 2.2 Mb, with G+C content from 30% to 65%.

KBP569 b9f is placed within a branch associated with brackish water near the freshwater strain *N. limneticus*. In another brackish clade, the dominant deep baltic sea genome 571-38 897 are closely related together with *N. cobalaminigenes* and *N. oxyclinae* from a brackish environment as well. This cluster also includes three other rare Baltic Sea strains 6a 625, 32a-657/580-32 657, and 7KS 352. Yet, the presence of strains from hydrothermal vent and Arctic sediment near these brackish groups adds complexity to understanding their evolution at a genus-level.

The strains exclusively identified in cultures exhibit a noticeable separation from other Baltic Sea strains, forming a unique cluster. The strain 68KS c8f is genetically close to brackish sediment strains such as *N. zosterae*, *N. ureiphilus*, and *N. adriaticus* from the Adriatic Sea. Remarkably, all five genomes in this cluster display genes associated with chemotaxis and

flagella, suggesting motility capabilities. The other culture-exclusive strain, 54a 52a, aligns closely with its nearest MAG from Sapelo Island, diverging distinctly from the remaining strains and emphasizing its unique phylogenetic positioning.

Our phylogenomic analysis of the genus *Nitrosopumilus* highlights some distinct clusters, with one host-associated group. This cluster, characterized by significant genomic diversity among its members and higher G+C content, suggests varied adaptations to sponge or coral hosts. Deep sea strains form another cohesive cluster, intriguingly grouping together regardless of their source—whether it be sediment, host, or the water column. This suggests that these strains, once established in the deep sea, have undergone evolutionary adaptation to this specific environment. In contrast, strains originating from marine sediments are dispersed throughout the tree, showing no tendency to cluster.

3.2.6 Genomic potential of novel Baltic Sea enrichments and representative AOA

To explore the genomic potential contributing to the dominance of certain strains in their environments and examine the presence and absence of genes of interest across different clades, we manually annotated various functional genes (Supplementary Table 3). It turns out that genes involved in ammonia oxidation and exinuclease activities appear to be highly conserved within the *Nitrosopumilus* genus (Fig. 10). Interestingly, the deep sea *Nitrosopumilus* clade retains exinuclease genes, typically associated with UV damage repair, despite their deep, UV-limited habitats. This suggests that these exinucleases may have roles beyond repairing UV damage, possibly in the maintenance of genome integrity or other DNA repair processes that are advantageous even in the absence of sunlight.

Every genome within the host-associated cluster harbors a urease gene cluster, and four out of six genomes in this cluster contain phosphorothioation (PT) genes (Fig. 10). These PT genes offer potential advantages, including antioxidant capabilities, restriction-modification systems, and viral resistance (Jian et al. 2021). Meanwhile, not all host-associated *Nitrosopumilus* genomes feature a urease gene cluster. The presence of the urase gene is observed in strains from diverse environments such as brackish, freshwater, deep sea, and sediment, indicating that there is no single environmental factor linked to the possession of the urease gene set. Host-related genomes also share a similar pattern of ortholog presence and absence (Supplementary Figure 2). And one deep sea sponge symbiont, *Nitrosopumilus sp013911135*, clusters together with the deep sea branch in the phylogenomic tree, but in the ortholog presence/absence heatmap, it clustered with other sponge symbionts. This could suggest that, even in different locations, evolving in conjunction with sponges has directed their evolution towards accumulating genes beneficial to them.

Regarding ectoin biosynthesis, ectoin serves as an osmolyte, helping organisms cope with osmotic stress. Initially, these gene sets are selected to screen because this might be a

critical adaptation for the Baltic Sea given the variable salinity of its environment. However, our findings show that only a few strains possess the complete ectoine synthesis gene cluster (Fig. 10). While the ectB is ubiquitous across *Nitrosopumilus*, its presence appears not to be directly associated with ectoine biosynthesis. Instead, its role is probably on other molecular functions, such as amino acid synthesis.

In the context of flagella and chemotaxis, one cluster including 68KS c9f strain, two sponge symbionts, and strains from the deep sea and sediment have flagella and chemotaxis genes (Fig. 10). There seems to be no specific trait shared among genomes that possess flagella genes. However, genomes with urease or flagella/chemotaxis genes tend to be larger in size. *N. ureiphilus* is the only genome that has both urease and flagella/chemotaxis gene sets, having the largest genome size among all reference genomes. This observation suggests that a larger genome may harbor a broad range of beneficial traits. Despite this, the genomes that predominate in the Baltic Sea (KBP569 b9f and 571-38 897) are comparatively smaller and lack these specific functional genes (Fig. 10). It may imply that their dominance is attributed to factors such as a more rapid reproduction rate due to their smaller genomes, however, it also could be the result of phenotypic variations such as difference kinetics of AMO genes that isn't discernible through genomic analysis.

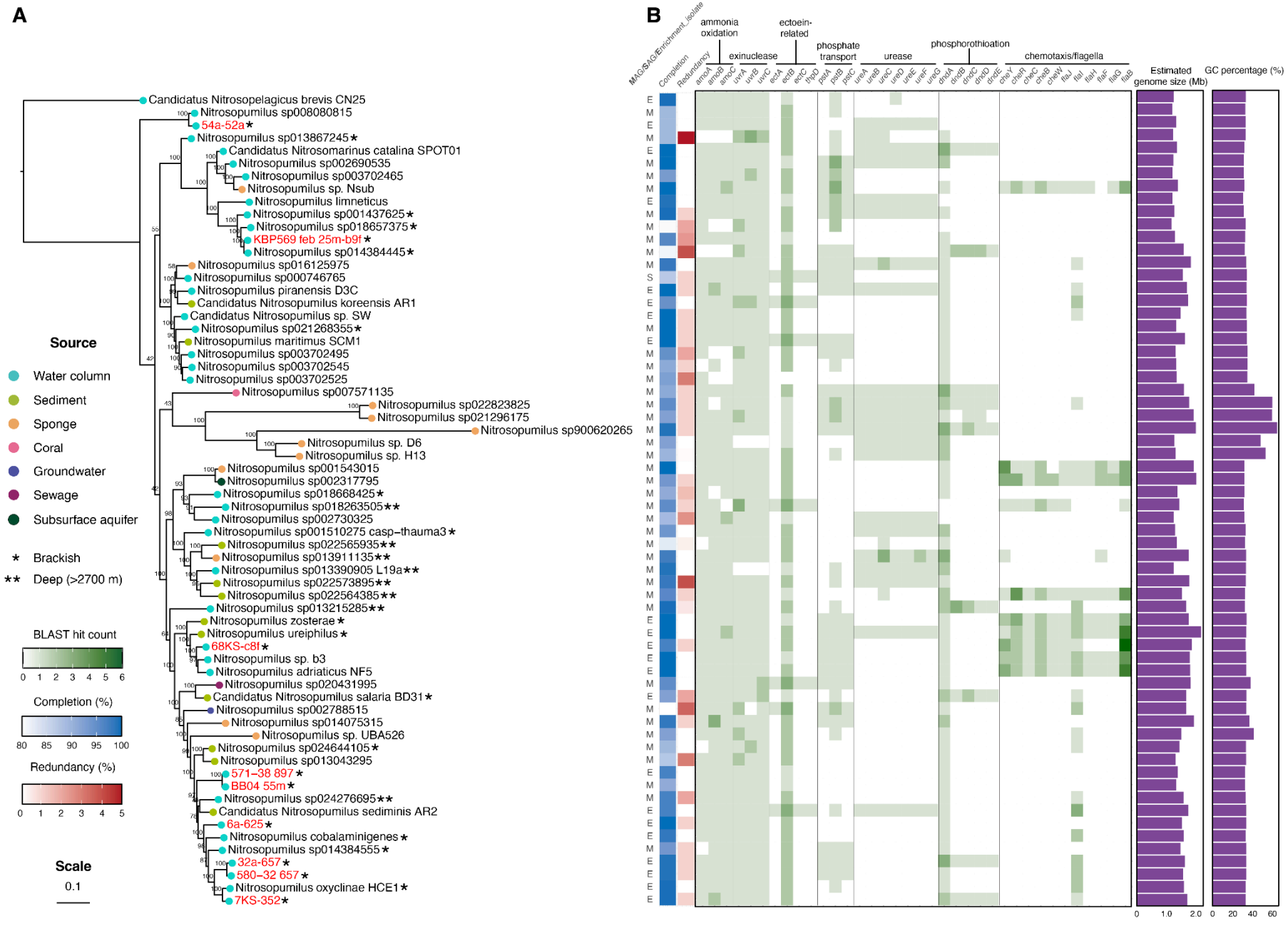

A
Candidatus Nitrosopelagicus brevis CN25
Nitrosopumilus sp008080815
54a-52a★
Nitrosopumilus sp013867245★
Candidatus Nitrosomarinus catalina SPOT01
Nitrosopumilus sp002690535
Nitrosopumilus sp003702465
Nitrosopumilus sp. Nsub
Nitrosopumilus limneticus
Nitrosopumilus sp001437625★
Nitrosopumilus sp018657375★
KBP569 feb 25m-b9f★
Nitrosopumilus sp014384445★
Nitrosopumilus sp016125975
Nitrosopumilus sp000746765
Nitrosopumilus piranensis D3C
Candidatus Nitrosopumilus koreensis AR1
Candidatus Nitrosopumilus sp. SW
Nitrosopumilus sp021268355★
Nitrosopumilus maritimus SCM1
Nitrosopumilus sp003702495
Nitrosopumilus sp003702545
Nitrosopumilus sp003702525
Nitrosopumilus sp007571135
Nitrosopumilus sp022823825
Nitrosopumilus sp021296175
Nitrosopumilus sp900620265
Nitrosopumilus sp. D6
Nitrosopumilus sp. H13
Nitrosopumilus sp001543015
Nitrosopumilus sp002317795
Nitrosopumilus sp018668425★
Nitrosopumilus sp018263505★★
Nitrosopumilus sp002730325
Nitrosopumilus sp001510275 casp−thauma3★
Nitrosopumilus sp022565935★★
Nitrosopumilus sp013911135★★
Nitrosopumilus sp013390905 L19a★★
Nitrosopumilus sp022573895★★
Nitrosopumilus sp022564385★★
Nitrosopumilus sp013215285★★
Nitrosopumilus zosterae★
Nitrosopumilus ureiphilus★
68KS-c8f★
Nitrosopumilus sp. b3
Nitrosopumilus adriaticus NF5
Nitrosopumilus sp020431995
Candidatus Nitrosopumilus salaria BD31★
Nitrosopumilus sp002788515
Nitrosopumilus sp014075315
Nitrosopumilus sp. UBA526
Nitrosopumilus sp024644105★
Nitrosopumilus sp013043295
571−38 897★
BB04 55m★
Nitrosopumilus sp024276695★★
Candidatus Nitrosopumilus sediminis AR2
6a-625★
Nitrosopumilus cobalaminigenes★
Nitrosopumilus sp014384555★
32a-657★
580−32 657★
Nitrosopumilus oxyclinae HCE1★
7KS-352★
Source
Water column
Sediment
Sponge
Coral
Groundwater
Sewage
Subsurface aquifer
★ Brackish
★★ Deep (>2700 m)
BLAST hit count
0 1 2 3 4 5 6
Completion (%)
80 85 90 95 100
Redundancy (%)
0 1 2 3 4 5
Scale
0.1
B
MAG/SAG/Enrichment_Isolate
Completion
Redundancy
ammonia oxidation
exinuclease
ectoein-related
phosphate transport
urease
phosphorothioation
chemotaxis/flagella
amoA amoB amoC uvrA uvrB uvrC ectA ectB ectC thpD pstA pstB pstC ureA ureB ureC ureD ureE ureF ureG dndA dndB dndC dndD dndE cheY cheR cheC cheB cheW flaJ flaI flaH flaF flaG flaB
Estimated genome size (Mb)
GC percentage (%)
0 1.0 2.0
0 20 40 60

Figure 10: Phylogenomic tree and functional gene distribution in 64 *Nitrosopumilus* genomes. A) The tree presents a phylogenomic tree constructed from 76 archaeal single-copy genes, incorporating both genomes obtained from this study (highlighted in red) and representative *Nitrosopumilus* strains from the Genome Taxonomy Database (GTDB). The environmental origins of the strains are denoted by colored circles at each branch tip, with strains from brackish environments marked by an asterisk (*) and those from deep sea environments (>2700m) indicated with two asterisks (**). B) Next to the tree, genome attributes and blast hit counts for selected functional genes are displayed. Sources of genomes are labeled as E for enrichment cultures or isolates, S for single-cell amplified genomes (SAGs), and M for metagenomic data. A blue and red heatmap illustrates genome completeness and redundancy assessed by CheckM. A green heatmap shows the gene count of selected functional genes identified through manual annotation. On the far right, purple columns provide estimates of genome size and GC content. Detailed information on each gene can be found in Supplementary Table 3.

3.2.7 The pan-genome of the genus Nitrosopumilus

In the examination of gene conservation and flexibility across *Nitrosopumilus* genomes, we engaged Anvi'o to visualize patterns of gene presence and absence across a set of genomes for a comprehensive overview and to identify genes unique to each genome. To check if the dominant strains in the Baltic Sea exhibit distinctive characteristics that make them successful in the environment, nine genomes from this study are visualized first (Fig. 11A). All 64 genomes in the phylogenomic analysis are screened as well to explore the differentiation between *Nitrosopumilus* genomes (Fig. 12).

Initially, a total of 3,648 gene clusters were identified across the nine *Nitrosopumilus* genomes from the Baltic Sea, and the contribution of each genome to each gene cluster was visualized through circularized bars (Fig. 11A). Gene clusters are binned manually by the number of contributing genomes, the gene clusters that have all nine genomes are grouped as 'Core genes 9' representing a conservative gene set. The conservative core genes, where every genome contributed genes to a particular gene cluster, amounted to 939. These core genes represent an average of 54.8% (±8.1%) of the genes in each genome. Considering incomplete genomes, soft core genes—those to which at least seven out of the nine genomes contributed— emerged as highly conserved across the 9 genomes studied. When combining soft core and core genes, they comprise an average of 76.3% (±7.3%) of the genes in each genome, indicating a high level of conservation across the studied genomes.

Gene clusters labeled 'Flexible genes 4-6' and 'Flexible genes 2-3', contributed to by parts of genomes, represent flexible or accessory gene sets that play a role in the evolution among different strains. It is possible to get insights into which genomes are more closely related by observing the presence and absence patterns of flexible genes, although no significant difference between dominant and rare strains was detectable through this way of visualization. Furthermore, the analysis revealed 1,556 singleton gene clusters that have one or multiple genes from a single genome, indicating a substantial diversity of unique genes across these genomes. On average, singletons constitute 9.42% (±5.97%) of the total protein count of each genome. Remarkably, the 68KS genome, which is the largest among those studied, features singleton genes comprising 20% of its total gene content. Such a high proportion of singleton genes in 68KS and others emphasizes the *Nitrosopumilus* genus's genetic variability and potential for specialized ecological adaptation.

With the same analysis produced with 64 genomes, we identified 14,259 gene clusters, with only 85 of these being shared across all genomes (Fig. 12). Since it was not easy to group the gene clusters into well-defined manners in this case, gene clusters are grouped and named by the number of contributing genomes. Given that many of these genomes are incomplete, asserting that only 85 gene clusters are common within the *Nitrosopumilus* genus seems impractical. Instead, by defining 'conserved' gene clusters as those present in approximately 90% of the genomes, we find that 1,001 gene clusters are conserved, accounting for about

57.2% (±7.5%) of the genes within each genome (gene clusters 61-64 in Fig. 10). This approach in analyzing gene presence-absence patterns, particularly in gene clusters 7-30 (gene clusters 7-30 in Fig. 10), enables the observation of diversification across different genomes within the genus. However, unlike the clear differentiation of ecotypes observed in other abundant marine microbes like *Prochlorococcus* (Delmont and Eren 2018), we did not identify a distinct set of flexible genes capable of classifying genomes from the Baltic Sea into specific groups.

Regarding singleton gene clusters (gene cluster 1 in Fig. 10), there are 9,023 singleton gene clusters, which represent 63.3% of all gene clusters. These singletons show considerable variability across genomes, accounting for an average of 7.7% (±7.0%) of the genes per genome. This substantial percentage highlights the extensive genomic diversity within the genus, underlining a vast reservoir of unique genes that potentially contribute to adaptive capabilities and genomic plasticity.

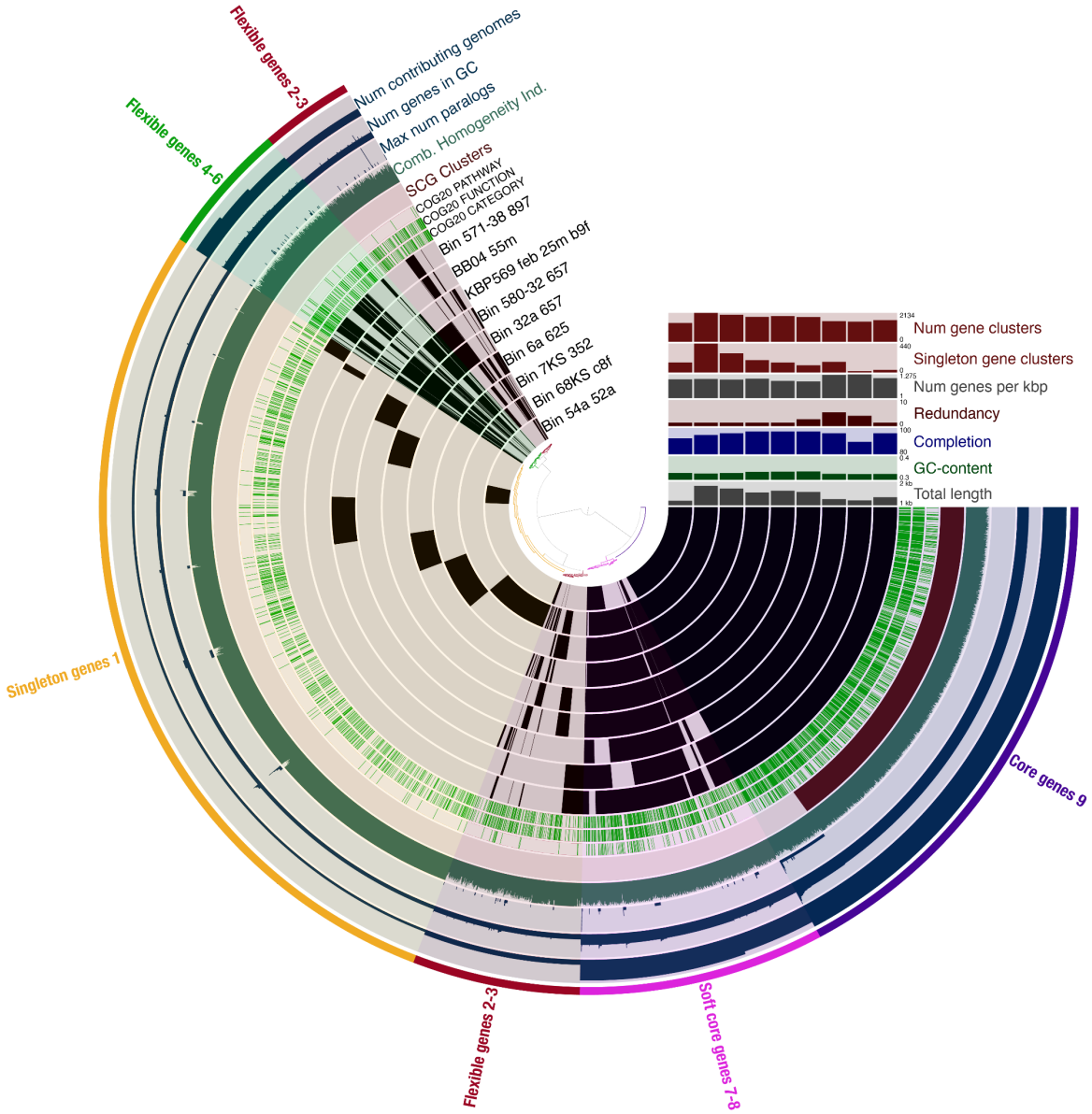


Figure 11: The Pan-genome of (A) the nine genomes from this study. and (B) the nine genomes plus *Nitrosopumilus* representatives. Each bar in the circularized plot represents a gene cluster and is filled if the corresponding genome contributes a gene to the gene cluster. The inner dendrogram displays homolog gene clusters in a pangenome, clustered by presence and absence patterns across the genomes. A panel located on the top right side (i.e., at "3 o'clock") shows information about each genome. Outer circles beside genomes add metadata for each gene cluster. From outermost to innermost these rows show 1.) the number of genomes that are contributing to the gene cluster, 2.) number genes in Gene Clusters, 3.) the number of genes in a gene cluster, 4.) maximum number of paralogs, 5.) maximum number of genes that come from the same genome, 6.) combined homogeneity index; indicative for homogeneity of sequences between genes in a gene cluster, 7.) SCG Clusters; single copy gene clusters, a gene cluster where every genome contributes one gene. 8.) COG20 PATHWAY, COG20 FUNCTION, COG20 CATEGORY; a gene cluster that has a gene annotated by COGs database.

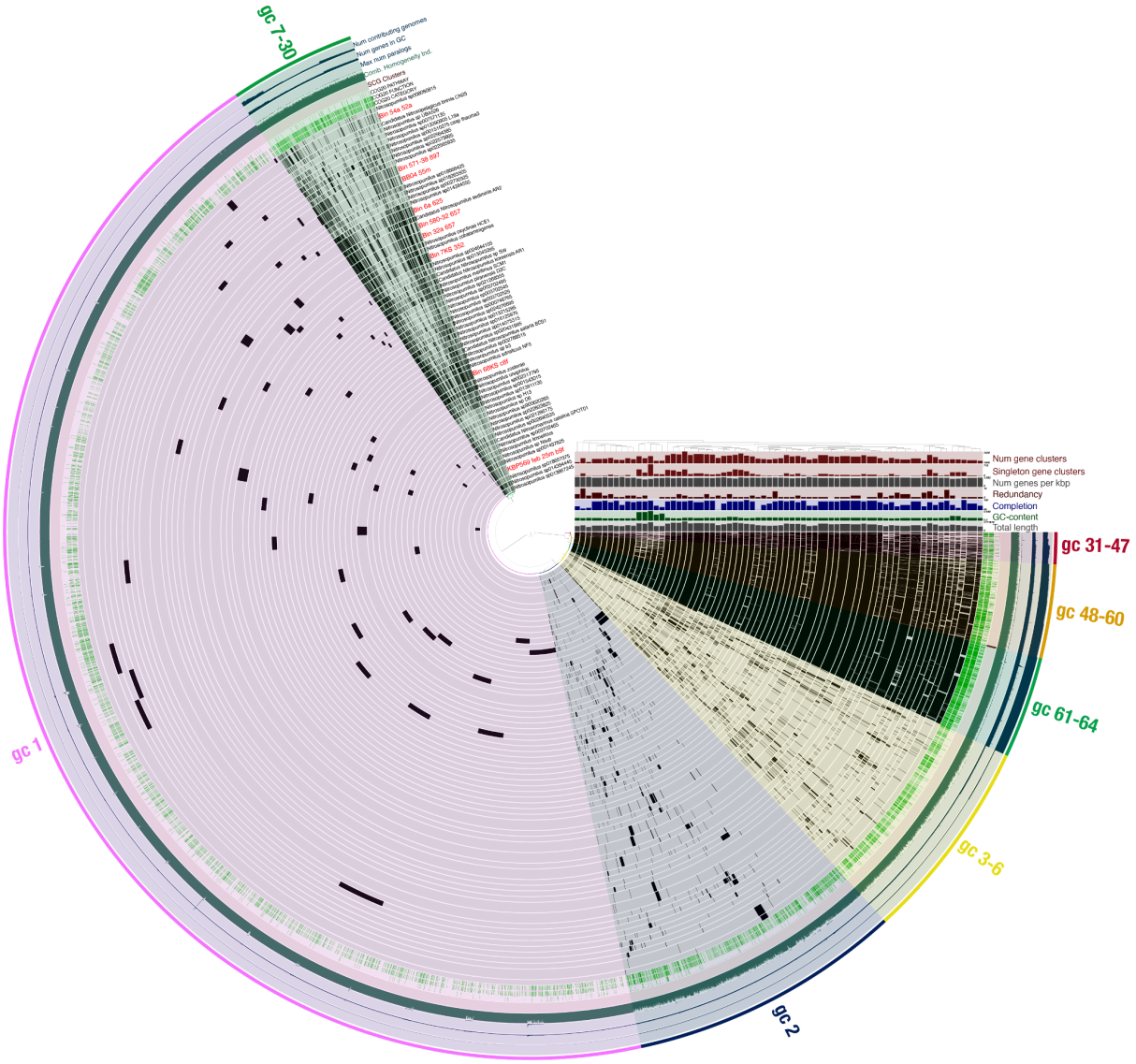


Figure 12: The Pan-genome of the nine genomes, like Figure 11, plus all *Nitrosopumilus* representatives from the GTDB database. Layers are the same as in Figure 11.

4. DISCUSSION

Our study provides a comprehensive analysis of the environmental distribution, diversity, and genomic characterization of AOA across two key time-series locations, Kiel Fjord and Boknis Eck, as well as numerous samples from the multiple depths of the broader Baltic Sea. The acquisition of enrichments of various AOA strains enabled us to obtain nine genomes including the most dominant deep baltic sea strain, and seven genomes are novel based on genomic whole genome nucleotide similarity metrics (average nucleotide similarity <95%). Phylogenomic and pan-genomic analyses within the genus *Nitrosopumilus* not only delineated the relationships between these new strains but also uncovered their genomic potential, illuminating AOA's complex ecology and adaptability in marine environments.

4.1 Dynamics of nitrifiers and environmental parameters in the Baltic Sea

The dynamics of nitrifiers in the southwest Baltic Sea demonstrate a seasonal dominance of AOA, especially the genus *Nitrosopumilus* in the surface water highlighting an apparent importance in nitrification processes (Fig. 2). AOA and AOB populations increase in late fall to early winter, with NOB peaking mid-winter, suggesting sequential nitrification intertwined by temperature and nutrients. One of the AOA strains (ASV b9f-) outnumbers all others, indicating a distinct ecological niche, whereas AOB exhibits a succession of dominant strains over time (Fig. 3). These dynamics seem closely aligned with environmental parameters like temperature and nutrients. At the Boknis Eck station, AOA populations reach their highest relative abundances in late fall-early winter, prevailing at the surface and with even higher relative abundances at deeper depth (Fig. 5). AOB, present alongside AOA, exhibit a less distinct seasonal pattern, showing consistent presence (albeit quite low in relative abundance) also in the summer. A clear shift in the dominant AOA strain from the southwest to the middle Baltic Sea, especially in the hypoxic deep Bornholm Basin, which is dominated by AOA ASV 897-, points to niche differentiation among AOA strains (Fig. 8).

While this study has provided foundational data for studying the dynamics of nitrifying microbes in the Baltic Sea, it faces several limitations that make it difficult to draw comprehensive conclusions. A significant constraint is the absence, so far, of ammonium and chlorophyll measurements and detailed environmental parameters from the Kieler Bucht Project (KBP) and broader Baltic Sea regions, which hinders our ability to directly correlate the observed microbial community dynamics with these specific environmental factors. Additionally, the time series data spanning only one and a half years does not adequately cover all seasons twice, limiting the robustness of our conclusions regarding seasonal patterns and their repeatability. Although we observed a winter dominance of AOA that suggests a clear seasonal trend, the current dataset does not provide sufficient evidence to connect microbial community changes to environmental parameters definitively. This limitation requires further years of observation and more comprehensive environmental

data collection to strengthen the link between microbial dynamics and the changing environmental conditions.

Analyzing microbial dynamics through 16S rRNA gene amplicon sequencing limits our ability to quantify the absolute abundance of microbes, allowing us only to assess their relative abundance within samples. Consequently, our observations are based on changes in the proportion of each taxon relative to others within the samples, which may not precisely reflect shifts in their actual population sizes in the environment. Despite this limitation, the seasonal increase in AOA, particularly *Nitrosopumilus*, from undetectable levels in some summer samples to a maximum of 10% of prokaryotic reads in winter samples, indicates their significant increase during the winter. In the future, approaches such as 'spike-in' of known microbes known not to be present in our samples can be used to estimate absolute abundances; the spike-in approach was done for the samples here, but the approach has not yet been fully validated, thus such estimates were left out and only relative abundances were used.

Addressing the compositional nature of these datasets, we calculated the relative abundance of each ASV as a proportion of the total count of the relative abundances into ratios. However, for more nuanced statistical analyses, applying log-ratio transformations, such as the centered log-ratio (CLR), would have been preferable (Gloor et al. 2017). Log-ratio transformations make the data symmetrical, allowing for more meaningful comparisons and correlations by treating proportional changes equally (Gloor et al. 2017). In addition, our analysis did not account for temporal autocorrelation within the microbial community dynamics, which could be addressed through model-based approaches or simpler methods like calculating dissimilarity at time-lag intervals. Recognizing these shortcomings underscores the need for more sophisticated statistical methodologies in future work to accurately capture the complex interactions between microbial communities and environmental parameters.

Furthermore, an essential aspect yet to be explored in our study involves quantifying carbon fixation and ammonia oxidation rates, particularly given the dominance of specific AOA strains during winter and in deeper waters. As AOA are known as chemoautotrophs and play an important role in the nitrogen cycle, the ability to directly measure these rates would offer invaluable insights into the ecological impact of AOA on nutrient cycling and primary productivity in the Baltic Sea and elsewhere. It could be presumed that AOA are among the dominant carbon fixers in the late-fall early winter in the Baltic Sea, based on their abundance, but this is not yet known, which further underscores their importance in the ecosystem. Therefore, metabolic rate measurements should be prioritized for future studies to elucidate the broader ecological implications of nitrifying microbes under varying environmental conditions.

The dynamics between AOA and AOB converting ammonia into nitrite, followed by NOB transforming nitrite into nitrate are consistent with expectations of a successional pattern of nitrification. These patterns have significant implications for the environment, particularly during the spring bloom, when rising temperatures and increased light levels fuel the growth of plankton and other organisms that consume nitrate—often their preferred energy source. This rapid uptake of nitrate typically results in its early depletion in the spring. These strong and rather consistent dynamics also set up a model ecosystem for studying these dynamics to advance an understanding of how the dynamics of the organisms are influenced by environmental parameters as well as one another.

Exploring the dynamics between environmental conditions and nitrifier populations in the Baltic Sea reveals complexities beyond nutrient concentration. Generally, AOA and AOB coexist across various regions in our study. However, in the hypoxic and sulfidic depths of the Gotland Basin, AOA notably outcompete AOB which might be attributed to AOA's tolerance to abiotic stressors including low oxygen and high sulfide concentration (Caffrey et al. 2007). Other studies on the Baltic Sea suggest AOA thrives in areas where sulfide and oxygen overlap, outcompeting AOB in such sulfidic, hypoxic environments (Labrenz et al. 2010). This dominance is highlighted in nitrification zones characterized by high nitrite concentrations amidst lower levels of oxygen or ammonium, within the pelagic layers (Labrenz et al. 2010).

Drawing from a decade of environmental data from the Boknis Eck time series station (Ma et al. 2020), we note that enhanced methane (CH_4) concentrations correlate with AOA peaks more closely than temperature or oxygen levels, particularly during February, May–June, and October. This pattern hints that AOA may thrive in environments where AOB is more sensitive than AOA, specifically in low oxygen areas (Yin et al. 2018), characterized by sulfidic and methane-rich conditions, using ammonia as their primary energy source excluding AOB in hypoxic regions. While our study didn't explore the interactions between nitrifiers and sulfate-reducing microbes or other anaerobic microbes, existing research underscores the importance of considering a broader range of environmental factors in understanding nitrifier dynamics.

Previous studies on the dynamics and distribution of ammonia-oxidizing archaea and bacteria in the Baltic Sea have largely relied on amoA primers, which focus solely on the dynamics of ammonia oxidizers providing valuable insights into niche differentiation between ammonia oxidizers (Kim et al. 2008; Labrenz et al. 2010; Vetterli et al. 2016; Happel et al. 2018). A recent study in the Northern Barents Sea (Thiele et al. 2023) adopted a universal primer set, assessing the seasonal prokaryote community composition. They observed patterns of nitrifier dynamics comparable to our findings, with AOA dominating from early winter to spring throughout the water column. Although their study spans a wide region of the Barents Sea, it limited its analysis to only four sampling time points. Our study could provide a more detailed portrayal of nitrifier dynamics, offering high-resolution

time-series data. Moreover, utilizing universal primers to encompass all microbial taxa, lays a foundational understanding of the microbial ecology of nitrifiers alongside other microbial groups.

4.2 Novel enrichment and genomic characterizations of AOA

Our study delved deeply into the diversity and ecological significance of AOA in the Baltic Sea, prompting us to enrich and characterize diverse AOA strains. Through enrichments and metagenomic sequencing, we obtained nine high-quality AOA genomes, unveiling strains from various Baltic Sea regions and identifying seven as novel members of the *Nitrosopumilus* genus (Table 1). These efforts revealed a considerable genetic diversity and potential for specialized ecological adaptation among AOA strains, as evidenced by our phylogenomic analysis which incorporated genomes from a broad array of environments and highlighted distinct clusters within the *Nitrosopumilus* lineage (Fig. 10). Our analysis also identified highly conserved core genes across the genomes, underscoring the genetic stability within the genus (Fig. 11 and Fig. 12). Furthermore, the discovery of a significant number of singleton gene clusters among these genomes emphasizes the extensive genomic variability within *Nitrosopumilus*, pointing towards a rich potential for adaptive evolution and ecological specialization. This comprehensive genomic characterization enhances our understanding of AOA's role in the Baltic Sea's microbial ecosystem, contributing valuable insights into the complexity and adaptability of marine ammonia-oxidizing archaea.

Enriching novel AOA strains from the Baltic Sea provides valuable opportunities for an in-depth exploration of their characteristics and adaptability to this environment, and may be able to serve as a model for the numerous other ecosystems in which they dominate. For a comprehensive understanding of these strains, it’s essential to characterize their optimal growth conditions such as temperature preferences, nutrient requirements, and affinities to substrates. Ideally, such detailed studies would be better conducted with isolated strains. However, achieving stable cultures, particularly for strains like the dominant deep Baltic Sea strain that thrives at cooler temperatures (4 and 10 C), is likely to be time-consuming and challenging. Among our enrichments, the dominant Baltic Sea strain 571-38, suggests the potential discovery of an AOA strain with a pronounced preference for cold temperatures, possibly the most cold-adapted within the characterized members of the *Nitrosopumilus* genus, though whether or not growth at such cold temperatures was attempted is unclear or unknown. In contrast to *N. maritimus* SCM1, *N. catalina*, and *N. adriaticus*, which exhibit growth optima at higher temperatures of 32°C, 23°C, and 30-32°C respectively, *N. oxyclinae* adapts to cooler temperatures with a growth minimum at 4°C and an optimum at 25°C (Ahlgren et al. 2017; Qin et al. 2017). Notably, *N. oxyclinae*, sharing a relatively close genetic relationship with 571-38 (86.4% ANI; Supplementary Table 1), hints at the cold-adaptation trait of 571-38. This is supported by its enrichment at 10°C, a capability that it does not exhibit at room temperature.

Our genomic data of AOA in the Baltic Sea shed light on the diversity of novel genomes and their relationships, with average completion rates of 96.4% and average redundancy estimates of 0.8% (Table 1). Yet, these genomes still present fragmentation, ranging from 11 to 75 contigs. Consequently, confidence in these findings is tempered by the incomplete nature of the genomes since they might have missed important genetic information. Moreover, some genomes might be hybrids from different species due to microdiversity among co-existing AOA strains in the environment and in some enrichment cultures (571-38 897 and 7KS 352 in Fig. 8). Therefore, for a more comprehensive description of their diversity and relationships, it is recommended to strive for single-contig assemblies. Utilizing long-read sequencing technology for key samples could significantly improve assembly difficulties associated with Illumina's short-read sequencing.

While we achieved relatively high-quality genomes for these dominant strains, understanding why certain strains prevail over others through genomic analysis alone proved challenging. This difficulty stems from several interconnected factors, the first of which is the challenge of obtaining complete genomes from all organisms of interest within the environment. In our study areas, a single AOA strain is dominant, and its variation with regional changes was evident. This dominance provides a unique opportunity for deeper investigation. However, if multiple strains coexist and fluctuate, deciphering the complex dynamics governing microbial ecosystem interactions becomes exponentially more challenging.

Furthermore, linking specific genomic traits to environmental adaptability is complex. Identifying the unique functional attributes that provide advantages to certain strains over their close relatives is not straightforward. Although genomic differences among strains are a given, these variations do not consistently align with regions that have been functionally annotated or are well-understood, making it hard to interpret their ecological implications. This is also true in our pan-genomic analysis with 64 genomes, only 32.2% of gene clusters (4,585 out of 14,259 gene clusters) have an annotated gene by COGs database.

Lastly, identifying unique genomic traits that could explain a strain's dominance is already a difficult task, but even if achieved, genomic data alone are insufficient. Further work, utilizing methods such as transcriptomics and proteomics, is necessary to prove how these genes are expressed and contribute to the organism's ecological success, demonstrating the need for a more comprehensive approach to understanding genetic potential and its translation into ecological dominance.

This study represents a pivotal step in understanding the diversity and ecological dynamics of AOA in the Baltic Sea, achieving successful enrichment of various strains, including novel members of the *Nitrosopumilus* genus, setting the stage for further isolation and

comprehensive characterization. These efforts mark the first extensive genomic representation of AOA strains in the Baltic Sea. While directly linking genomic evidence to the dominance of specific strains in the environment remains challenging, the acquisition of numerous high-quality genomes lays the groundwork for future studies. Looking ahead, isolating and closely examining these enriched strains will be key to confirming their role in the Baltic Sea ecosystem and nitrogen cycle.

5. CONCLUSIONS

This study has significantly advanced our understanding of the diversity and ecological dynamics of ammonia-oxidizing archaea (AOA) in the Baltic Sea, marking the first extensive genomic representation of AOA strains within this unique marine ecosystem. Through the successful enrichment and genomic sequencing of various AOA strains, including seven novel members of the *Nitrosopumilus* genus, we have unveiled a remarkable genetic diversity and potential for specialized ecological adaptation among these organisms. Comparative genomic analysis has further illuminated a high degree of gene conservation across the *Nitrosopumilus* genus, suggesting a core set of genes fundamental to AOA strains while also revealing many unique genes that hint at their diverse ecological roles and adaptability. While an important step forward, comparative genomics is challenged by annotation difficulties of novel Archaea as well as tying it directly to the observed ecological patterns. Nevertheless, our findings underscore the apparent critical role of AOA in the Baltic Sea's nitrification processes, highlighted by their seasonal dominance on the surface during the colder months and their substantial presence in deeper waters.

Moving forward, the isolation and detailed characterization of these enriched strains are imperative for confirming their status as key players in the Baltic Sea's microbial community and understanding their contribution to nitrogen cycling and marine biogeochemistry. Despite the challenges in directly linking genomic evidence to strain dominance in the environment, our study lays a solid foundation for future research. It paves the way for a more comprehensive exploration of AOA's ecological impact, leveraging additional methodologies such as transcriptomics, heterologous expression studies of novel or target genes, stable isotope probing experiments to evaluate their impact and ecological interactions with other organisms and to uncover how these organisms adapt and thrive in the Baltic Sea's changing conditions. As such, this work not only enriches our grasp of microbial ecology but also establishes the Baltic Sea as a pivotal model for studying marine ammonia oxidizers and their integral role in global nitrogen cycles.

Acknowledgments

This thesis would not have been achievable without the help from many at GEOMAR. I deeply appreciate all who were involved in our time-series projects, with special thanks to the Marine Microbial Ecology group for their dedication in sampling and processing for GPT and KBP: Elisa D'Agostino, Lara Schroeder, Alyzza Calayag, Marjan Ghotbi, and David Needham. I also thank the invaluable help from former interns for our time series sampling: Selina Ernst, Karolin Eisenschmid, Clara Fenge, Ana Belén Kuhlmann, and Maarten Kanitz. Special appreciation goes to Elisa D'Agostino for organizing the sampling projects and carrying out molecular work of the GPT, KBP, and cruise samples and Selina Ernst who helped enrich the initial AOA from the Baltic Sea.

I would like to extend my thanks to the chief scientists who facilitated the samplings from the broader Baltic Sea, Jan Dierking (AL571) and Felix Mittermayer (AL580), as well as to all participants in the cruises, including myself and David Needham who helped collect samples. For Boknis Eck time series sampling, thanks to Hermann Bange and Kastriot Qelaj for organizing the monthly cruises. Also, the valuable environmental data was from KIMOCC thanks to Claas Hiebenthal and Frank Melzer.

My gratitude goes to David Needham for his guidance and support throughout my master's study. His dedication and willingness to listen have been invaluable. Our group members and all colleagues from OEB have shown their unwavering support and encouragement.

It has already been five years since I turned from my previous career as a teacher to embark on a new journey as a scientist. Although the path was not always smooth, I am deeply grateful to have reached this point and to finalize my master's thesis. I extend my thanks to all the companions I've met along this journey, and to my dearest friends and family. Lastly, this is dedicated to J.J., who always stands by my side.

REFERENCES

Ahlgren NA, Chen Y, Needham DM, Parada AE, Sachdeva R, Trinh V, et al. Genome and epigenome of a novel marine Thaumarchaeota strain suggest viral infection, phosphorothioation DNA modification and multiple restriction systems. Environ Microbiol. 2017 Jun;19(6):2434–52.

Andrews S. FastQC: a quality control tool for high throughput sequence data. Available online. Retrieved May. 2010;17:2018.

Bale NJ, Villanueva L, Hopmans EC, Schouten S, Sinninghe Damsté JS. Different seasonality of pelagic and benthic Thaumarchaeota in the North Sea. Biogeosciences. 2013 Nov 12;10(11):7195–206.

Bange HW, Bergmann K, Hansen HP, Kock A, Koppe R, Malien F, et al. Dissolved methane during hypoxic events at the Boknis Eck time series station (Eckernförde Bay, SW Baltic Sea). Biogeosciences. 2010 Apr 19;7(4):1279–84.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012 May;19(5):455–77.

Bayer B, Vojvoda J, Offre P, Alves RJE, Elisabeth NH, Garcia JA, et al. Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation. ISME J. 2016 May;10(5):1051–63.

Berg C, Listmann L, Vandieken V, Vogts A, Jürgens K. Chemoautotrophic growth of ammonia-oxidizing Thaumarchaeota enriched from a pelagic redox gradient in the Baltic Sea. Front Microbiol. 2014;5:786.

Berg C, Vandieken V, Thamdrup B, Jürgens K. Significance of archaeal nitrification in hypoxic waters of the Baltic Sea. ISME J. 2015 Jun;9(6):1319–32.

Betts HC, Puttick MN, Clark JW, Williams TA, Donoghue PCJ, Pisani D. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. Nat Ecol Evol. 2018 Oct;2(10):1556–62.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019 Aug;37(8):852–7.

Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. Nat Rev Microbiol. 2008 Mar;6(3):245–52.

Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015 Jan;12(1):59–60.

Bushnell B. BBMap: A Fast, Accurate, Splice-Aware Aligner. 2014 Mar 19 [cited 2024 Feb 11]; Available from: https://escholarship.org/uc/item/1h3515gn

Byung Kwon, Kim, Man-Young, Jung, Dong Su, Yu, Soo-Je, Park, Tae Kwang, Oh, Sung-Keun, Rhee, et al. Genome Sequence of an Ammonia-Oxidizing Soil Archaeon, "Candidatus Nitrosoarchaeum koreensis" MY1. J Bacteriol. 2011 Sep 13;193(19):5539–40.

Caffrey JM, Bano N, Kalanetra K, Hollibaugh JT. Ammonia oxidation and ammonia-oxidizing bacteria and archaea from estuaries with differing histories of hypoxia. ISME J. 2007 Nov;1(7):660–2.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016 Jul;13(7):581–3.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009 Dec 15;10:421.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009 Aug;25(15):1972–3.

Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2019 Nov 15;36(6):1925–7.

Dale AW, Sommer S, Bohlen L, Treude T, Bertics VJ, Bange HW, et al. Rates and regulation of nitrogen cycling in seasonally hypoxic sediments during winter (Boknis Eck, SW Baltic Sea): Sensitivity to environmental variables. Estuar Coast Shelf Sci. 2011 Nov 1;95(1):14–28.

Delmont TO, Eren AM. Linking pangenomes and metagenomes: the Prochlorococcus metapangenome. PeerJ. 2018 Jan 25;6:e4320.

DeLong EF. Archaea in coastal marine environments. Proc Natl Acad Sci U S A. 1992 Jun 15;89(12):5685–9.

van Dongen S, Abreu-Goodger C. Using MCL to extract clusters from networks. Methods Mol Biol. 2012;804:281–95.

Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011 Oct;7(10):e1002195.

Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019 Nov 14;20(1):238.

Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, et al. Community-led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol. 2021 Jan;6(1):3–6.

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 U S A. 2005 Oct 11;102(41):14683–8.

Fuhrman JA, McCallum K, Davis AA. Novel major archaebacterial group from marine plankton. Nature. 1992 Mar 12;356(6365):148–9.

Gaio D, Anantanawat K, To J, Liu M, Monahan L, Darling AE. Hackflex: low-cost, high-throughput, Illumina Nextera Flex library construction. Microb Genom [Internet]. 2022 Jan;8(1). Available from: http://dx.doi.org/10.1099/mgen.0.000744

Gindorf S, Bange HW, Booge D, Kock A. Seasonal study of the small-scale variability in dissolved methane in the western Kiel Bight (Baltic Sea) during the European heatwave in 2018. Biogeosciences. 2022 Oct 27;19:4993–5006.

Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional:

And This Is Not Optional. Front Microbiol. 2017 Nov 15;8:2224.

Gruber N, Galloway JN. An Earth-system perspective of the global nitrogen cycle. Nature. 2008 Jan 17;451(7176):293–6.

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, et al. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Res. 2013 Jan;41(Database issue):D597–604.

Happel E, Bartl I, Voss M, Riemann L. Extensive nitrification and active ammonia oxidizers in two contrasting coastal systems of the Baltic Sea. Environ Microbiol. 2018 Aug;20(8):2913–26.

Herfort L, Schouten S, Abbas B, Veldhuis MJW, Coolen MJL, Wuchter C, et al. Variations in spatial and temporal distribution of Archaea in the North Sea in relation to environmental variables. FEMS Microbiol Ecol. 2007 Dec;62(3):242–57.

Hmisc. 2023 [cited 2024 Mar 3]. Available from: https://hbiostat.org/r/hmisc/

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol. 2018 Feb 1;35(2):518–22.

Hodgskiss LH, Melcher M, Kerou M, Chen W, Ponce-Toledo RI, Savvides SN, et al. Unexpected complexity of the ammonia monooxygenase in archaea. ISME J. 2023 Apr;17(4):588–99.

Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11(1):119–119.

Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018 Nov 30;9(1):5114.

Jian H, Xu G, Yi Y, Hao Y, Wang Y, Xiong L, et al. The origin and impeded dissemination of the DNA phosphorothioation system in prokaryotes. Nat Commun. 2021 Nov 4;12(1):6382.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017 Jun;14(6):587–9.

Karner MB, DeLong EF, Karl DM. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. Nature. 2001 Jan 25;409(6819):507–10.

Kim JG, Park SJ, Sinninghe Damsté JS, Schouten S, Rijpstra WIC, Jung MY, et al. Hydrogen peroxide detoxification is a key mechanism for growth of ammonia-oxidizing archaea. Proc Natl Acad Sci U S A. 2016 Jul 12;113(28):7888–93.

Kim OS, Junier P, Imhoff JF, Witzel KP. Comparative analysis of ammonia monooxygenase (amoA) genes in the water column and sediment-water interface of two lakes and the Baltic Sea. FEMS Microbiol Ecol. 2008 Nov;66(2):367–78.

Klotz F, Kitzinger K, Ngugi DK, Büsing P, Littmann S, Kuypers MMM, et al. Quantification of archaea-driven freshwater nitrification from single cell to ecosystem levels. ISME J. 2022 Jun;16(6):1647–56.

Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA. Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature. 2005 Sep 22;437(7058):543–6.

Könneke M, Schubert DM, Brown PC, Hügler M, Standfest S, Schwander T, et al. Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO2 fixation. Proc Natl Acad Sci U S A. 2014 Jun 3;111(22):8239–44.

Kraft B, Jehmlich N, Larsen M, Bristow LA, Könneke M, Thamdrup B, et al. Oxygen and nitrogen production by an ammonia-oxidizing archaeon. Science. 2022 Jan 7;375(6576):97–100.

Labrenz M, Sintes E, Toetzke F, Zumsteg A, Herndl GJ, Seidler M, et al. Relevance of a crenarchaeotal subcluster related to Candidatus Nitrosopumilus maritimus to ammonia oxidation in the suboxic zone of the central Baltic Sea. ISME J. 2010 Dec;4(12):1496–508.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4;9(4):357–9.

Lee EY, Lee HK, Lee YK, Sim CJ, Lee JH. Diversity of symbiotic archaeal communities in marine sponges from Korea. Biomol Eng. 2003 Jul;20(4-6):299–304.

Lee M. Bit: A multipurpose collection of bioinformatics tools. F1000Res. 2022 Jan 31;11:122.

Lee MD. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics. 2019 Oct 15;35(20):4162–4.

Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol GW, et al. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature. 2006 Aug 17;442(7104):806–9.

Lennartz ST, Lehmann A, Herrford J, Malien F, Hansen HP, Biester H, et al. Long-term trends at the Boknis Eck time series station (Baltic Sea), 1957–2013: does climate change counteract the decline in eutrophication? Biogeosciences. 2014 Nov 24;11(22):6323–39.

Liu X, Pan J, Liu Y, Li M, Gu JD. Diversity and distribution of Archaea in global estuarine ecosystems. Sci Total Environ. 2018 Oct 1;637-638:349–58.

Löscher CR, Kock A, Könneke M, LaRoche J, Bange HW, Schmitz RA. Production of oceanic nitrous oxide by ammonia-oxidizing archaea. Biogeosciences. 2012 Jul 4;9(7):2419–29.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011 May 2;17(1):10–2.

Ma X, Sun M, Lennartz ST, Bange HW. A decade of methane measurements at the Boknis Eck Time Series Station in Eckernförde Bay (southwestern Baltic Sea). Biogeosciences. 2020 Jul 6;17(13):3427–38.

Mehrshad M, Amoozegar MA, Ghai R, Shahzadeh Fazeli SA, Rodriguez-Valera F. Genome Reconstruction from Metagenomic Data Sets Reveals Novel Microbes in the Brackish Waters of the Caspian Sea. Appl Environ Microbiol. 2016 Jan 4;82(5):1599–612.

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 May;9(5):1162–75.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020 May 1;37(5):1530–4.

Moeller FU, Webster NS, Herbold CW, Behnam F, Domman D, Albertsen M, et al. Characterization of a thaumarchaeal symbiont that drives incomplete nitrification in the tropical sponge Ianthella basta. Environ Microbiol. 2019 Oct;21(10):3831–54.

Mohrholz V, Naumann M, Nausch G, Krüger S, Gräwe U. Fresh oxygen for the Baltic Sea — An exceptional saline inflow after a decade of stagnation. J Mar Syst. 2015 Aug 1;148:152–66.

Mosier AC, Allen EE, Kim M, Ferriera S, Francis CA. Genome sequence of "Candidatus Nitrosopumilus salaria" BD31, an ammonia-oxidizing archaeon from the San Francisco Bay estuary. J Bacteriol. 2012 Apr;194(8):2121–2.

Müller O, Wilson B, Paulsen ML, Rumińska A, Armo HR, Bratbak G, et al. Spatiotemporal Dynamics of Ammonia-Oxidizing Thaumarchaeota in Distinct Arctic Water Masses. Front Microbiol. 2018 Jan 23;9:24.

Murat Eren A, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ. 2015 Oct 8;3:e1319.

Nakagawa T, Koji M, Hosoyama A, Yamazoe A, Tsuchiya Y, Ueda S, et al. Nitrosopumilus zosterae sp. nov., an autotrophic ammonia-oxidizing archaeon of phylum Thaumarchaeota isolated from coastal eelgrass sediments of Japan. Int J Syst Evol Microbiol [Internet]. 2021 Aug;71(8). Available from: http://dx.doi.org/10.1099/ijsem.0.004961

Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC. New insights from uncultivated genomes of the global human gut microbiome. Nature. 2019 Apr;568(7753):505–10.

Nishimura Y, Yoshizawa S. The OceanDNA MAG catalog contains over 50,000 prokaryotic genomes originated from various marine environments. Sci Data. 2022 Jun 17;9(1):305.

Palatinszky M, Herbold C, Jehmlich N, Pogoda M, Han P, von Bergen M, et al. Cyanate as an energy source for nitrifiers. Nature. 2015 Aug 6;524(7563):105–8.

Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ Microbiol. 2016 May;18(5):1403–14.

Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. A complete domain-to-species taxonomy for Bacteria and Archaea. Nat Biotechnol. 2020 Sep;38(9):1079–86.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015 Jul;25(7):1043–55.

Park SJ, Kim JG, Jung MY, Kim SJ, Cha IT, Ghai R, et al. Draft genome sequence of an ammonia-oxidizing archaeon, "Candidatus Nitrosopumilus sediminis" AR2, from Svalbard in the Arctic Circle. J Bacteriol. 2012a Dec;194(24):6948–9.

Park SJ, Kim JG, Jung MY, Kim SJ, Cha IT, Kwon K, et al. Draft genome sequence of an ammonia-oxidizing archaeon, "Candidatus Nitrosopumilus koreensis" AR1, from marine sediment. J Bacteriol. 2012b Dec;194(24):6940–1.

Pitcher A, Wuchter C, Siedenberg K, Schouten S, Sinninghe Damsté JS. Crenarchaeol tracks winter blooms of ammonia-oxidizing Thaumarchaeota in the coastal North Sea. Limnol Oceanogr. 2011

Nov;56(6):2308–18.

Preston CM, Wu KY, Molinski TF, DeLong EF. A psychrophilic crenarchaeon inhabits a marine sponge: Cenarchaeum symbiosum gen. nov., sp. nov. Proc Natl Acad Sci U S A. 1996 Jun 25;93(13):6241–6.

Qin W, Amin SA, Martens-Habbena W, Walker CB, Urakawa H, Devol AH, et al. Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. Proc Natl Acad Sci U S A. 2014 Aug 26;111(34):12504–9.

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 ureiphilus sp. nov., four marine ammonia-oxidizing archaea of the phylum Thaumarchaeota. Int J Syst Evol Microbiol. 2017 Dec;67(12):5067–79.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013 Jan;41(Database issue):D590–6.

R Core Team (2022) R A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. - references - scientific research publishing. [cited 2024 Feb 11]. Available from: https://www.scirp.org/reference/referencespapers?referenceid=3456808

Ren M, Feng X, Huang Y, Wang H, Hu Z, Clingenpeel S, et al. Phylogenomics suggests oxygen availability as a driving force in Thaumarchaeota evolution. ISME J. 2019 Sep;13(9):2150–61.

Reusch TBH, Dierking J, Andersson HC, Bonsdorff E, Carstensen J, Casini M, et al. The Baltic Sea as a time machine for the future coastal ocean. Sci Adv. 2018 May;4(5):eaar8195.

Santoro AE, Buchwald C, McIlvin MR, Casciotti KL. Isotopic signature of N(2)O produced by marine ammonia-oxidizing archaea. Science. 2011 Sep 2;333(6047):1282–5.

Santoro AE, Casciotti KL. Enrichment and characterization of ammonia-oxidizing archaea from the open ocean: phylogeny, physiology and stable isotope fractionation. ISME J. 2011 Nov;5(11):1796–808.

Santoro AE, Dupont CL, Richter RA, Craig MT, Carini P, McIlvin MR, et al. Genomic and proteomic characterization of "*Candidatus* Nitrosopelagicus brevis": An ammonia-oxidizing archaeon from the open ocean. Proceedings of the National Academy of Sciences. 2015;112(4):1173–8.

Santoro AE, Kellom M, Laperriere SM. Contributions of single-cell genomics to our understanding of planktonic marine archaea. Philos Trans R Soc Lond B Biol Sci. 2019 Nov 25;374(1786):20190096.

Snoeijs-Leijonmalm P, Schubert H, Radziejewska T. Biological Oceanography of the Baltic Sea. Springer Science & Business Media; 2017.

Steinle L, Maltby J, Treude T, Kock A, Bange HW, Engbersen N, et al. Effects of low oxygen concentrations on aerobic methane oxidation in seasonally hypoxic coastal waters. Biogeosciences. 2017 Mar 29;14(6):1631–45.

Strickland JDH, Parsons TR. A Practical Handbook of Seawater Analysis. Fisheries Research Board of Canada; 1968.

Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000 Jan 1;28(1):33–6.

Thiele S, Vader A, Thomson S, Saubrekka K, Petelenz E, Müller O, et al. Seasonality of the bacterial and archaeal community composition of the Northern Barents Sea. Front Microbiol. 2023 Jul 7;14:1213718.

Tully BJ, Wheat CG, Glazer BT, Huber JA. A dynamic microbial community with high functional redundancy inhabits the cold, oxic subseafloor aquifer. ISME J. 2018 Jan;12(1):1–16.

Vajrala N, Martens-Habbena W, Sayavedra-Soto LA, Schauer A, Bottomley PJ, Stahl DA, et al. Hydroxylamine as an intermediate in ammonia oxidation by globally abundant marine archaea. Proc Natl Acad Sci U S A. 2013 Jan 15;110(3):1006–11.

Vetterli A, Hietanen S, Leskinen E. Spatial and temporal dynamics of ammonia oxidizers in the sediments of the Gulf of Finland, Baltic Sea. Mar Environ Res. 2016 Feb;113:153–63.

Vigneron A, Cruaud P, Lovejoy C, Vincent WF. Genomic insights into cryptic cycles of microbial hydrocarbon production and degradation in contiguous freshwater and marine microbiomes. Microbiome. 2023 May 12;11(1):104.

Vissers EW, Blaga CI, Bodelier PLE, Muyzer G, Schleper C, Sinninghe Damsté JS, et al. Seasonal and vertical distribution of putative ammonia-oxidizing thaumarchaeotal communities in an oligotrophic lake. FEMS Microbiol Ecol. 2013 Feb;83(2):515–26.

Voss M, Baker A, Bange HW, Conley D, Deutsch B, Engel A, et al. Nitrogen processes in coastal and marine ecosystems. In Cambridge University Press; 2011.

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 U S A. 2010 May 11;107(19):8818–23.

Wang Y, Huang JM, Cui GJ, Nunoura T, Takaki Y, Li WL, et al. Genomics insights into ecotype formation of ammonia-oxidizing archaea in the deep ocean. Environ Microbiol. 2019 Feb;21(2):716–29.

Webster NS, Watts JE, Hill RT. Detection and phylogenetic analysis of novel crenarchaeote and euryarchaeote 16S ribosomal RNA gene sequences from a Great Barrier Reef sponge. Mar Biotechnol . 2001 Nov;3(6):600–8.

Wickham H. Data Analysis. In: Wickham H, editor. ggplot2: Elegant Graphics for Data Analysis. Cham: Springer International Publishing; 2016. p. 189–201.

Wickham HV. D. & Girlich, M. _tidyr: Tidy Messy Data_. R package version 1.3. 0. 2023.

Wuchter C, Abbas B, Coolen MJL, Herfort L, van Bleijswijk J, Timmers P, et al. Archaeal nitrification in the ocean. Proc Natl Acad Sci U S A. 2006 Aug 15;103(33):12317–22.

Yang Y, Zhang C, Lenton TM, Yan X, Zhu M, Zhou M, et al. The Evolution Pathway of Ammonia-Oxidizing Archaea Shaped by Major Geological Events. Mol Biol Evol. 2021 Aug 23;38(9):3637–48.

Yeh YC, McNichol J, Needham DM, Fichot EB, Berdjeb L, Fuhrman JA. Comprehensive single-PCR 16S and 18S rRNA community analysis validated with mock communities, and estimation of sequencing bias against 18S. Environ Microbiol. 2021 Jun;23(6):3240–50.

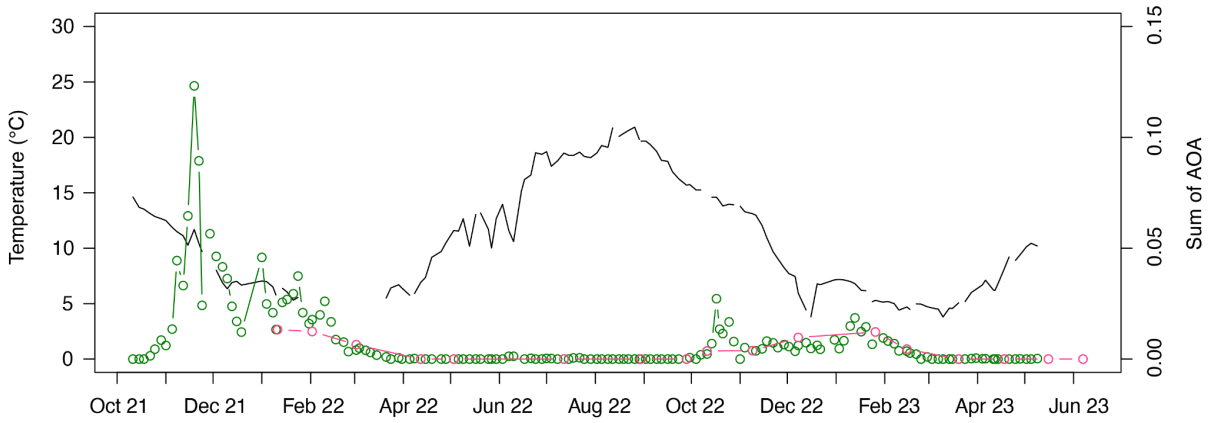
Yin Z, Bi X, Xu C. Ammonia-Oxidizing Archaea (AOA) Play with Ammonia-Oxidizing Bacteria (AOB) in Nitrogen Removal from Wastewater. Archaea. 2018 Sep 13;2018:8429145.

Yu G. Data Integration, Manipulation and Visualization of Phylogenetic Trees. CRC Press; 2022.

Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. Ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017 Jan;8(1):28–36.

Zhang S, Song W, Wemheuer B, Reveillaud J, Webster N, Thomas T. Comparative Genomics Reveals Ecological and Evolutionary Insights into Sponge-Associated Thaumarchaeota. mSystems [Internet]. 2019 Aug 13;4(4). Available from: http://dx.doi.org/10.1128/mSystems.00288-19

Supplementary Data



Supplementary Figure 1: Aligned dynamics of AOA in GPT (in green) and KBP (in pink) sites. The background line (black) represents the temperature dynamics at the GPT site for reference.

Supplementary Table 1: Average Nucleotide Identity (ANI) Metrics. This table presents the whole genome nucleotide similarity, assessed using FastANI (Jain et al. 2018), by comparing nine genomes obtained in this study against 64 genomes incorporated in the comparative analysis. The bolded cell in each column indicates the closest relative to the corresponding genome.

	BB04_55m	Bin_32a	Bin_54a	Bin_571-38	Bin_580-32	Bin_68KS	Bin_6a	Bin_7KS	metabat.KBP569_Feb_25m
BB04_55m	100	87.7407	78.8951	99.482	87.8742	82.0643	83.7819	86.7437	81.3597
Bin_32a	87.7405	100	79.0168	87.6409	98.4513	81.7974	83.8028	86.8649	80.6673
Bin_54a	79.1603	78.8232	100	78.7967	79.0014	79.2047	79.0317	79.2825	78.5852
Bin_571-38	99.5465	87.5195	78.7927	100	87.7569	82.1583	83.9366	86.4847	81.4247
Bin_580-32	87.8626	98.3264	79.178	87.7279	100	81.7291	83.7254	86.7215	80.7524
Bin_68KS	82.026	81.7734	79.3789	82.0877	81.7407	100	81.6078	81.8561	79.9683
Bin_6a	83.9416	83.7136	78.8968	83.9494	83.6473	81.4454	100	83.8226	80.6022
Bin_7KS	86.851	86.7543	79.3766	86.6642	87.0164	81.9229	84.1182	100	81.0123
metabat.KBP569_Feb_25m	81.3093	80.2749	78.5957	81.2387	80.3801	80.1406	80.4065	81.0309	100
Candidatus_Nitrosomarinus_catalina_SPOT01	80.886	80.607	79.1932	80.9249	80.889	80.5195	80.8057	80.7541	84.3756
Candidatus_Nitrosopelagicus_brevis_CN25	77.1432	76.8794	77.8033	77.1551	77.0747	77.2886	77.266	77.1888	77.3034
Candidatus_Nitrosopumilus_koreensis_AR1	80.6475	80.557	79.6058	80.5773	80.6111	80.9196	80.9075	80.5875	79.5896
Candidatus_Nitrosopumilus_salaria_BD31	82.0568	81.9758	79.0313	81.9551	82.0922	81.6838	81.8078	81.8629	80.1801
Candidatus_Nitrosopumilus_sediminis_AR2	84.6352	83.9905	78.994	84.6034	83.9783	81.7401	87.0239	84.1848	80.6335

Candidatus_Nitrosopumilus_sp._SW	80.991	80.7909	79.7296	81.1335	80.9526	81.0209	81.0626	81.326	79.8714
Nitrosopumilus_adriaticus_NF5	81.622	81.6215	79.5542	81.6762	81.5145	89.2605	81.6853	81.7291	80.2643
Nitrosopumilus_cobalaminigenes	84.8622	84.465	79.4272	84.925	84.5549	82.1444	84.263	85.2336	80.7182
Nitrosopumilus_limneticus	81.0753	80.5141	78.7312	80.8002	80.7399	80.0375	80.0992	80.5371	85.1704
Nitrosopumilus_maritimus_SCM1	80.6371	80.7429	79.7136	80.812	80.7891	80.9919	81.4342	81.2403	79.5168
Nitrosopumilus_oxyclinae_HCE1	86.5219	86.8702	79.3458	86.4364	86.6584	81.9348	84.0522	93.7431	80.8393
Nitrosopumilus_piranensis_D3C	81.1585	80.5678	79.7189	80.7246	80.51	80.7676	80.854	80.68	79.2356
Nitrosopumilus_sp._b3	81.3502	81.3798	79.4092	81.6167	81.4855	88.9729	81.4076	81.5228	79.6305
Nitrosopumilus_sp._Nsub	81.0122	80.5561	79.2407	81.1411	80.5487	80.4262	80.5468	80.6678	84.0305
Nitrosopumilus_sp000746765	81.1126	80.7413	79.9373	80.9386	80.7552	80.9305	81.2941	81.2149	79.4711
Nitrosopumilus_sp001437625	81.5923	80.8931	78.7714	81.3716	80.8021	80.4132	80.4775	81.0698	92.0694
Nitrosopumilus_sp001510275_casp-thauma3	82.5016	81.762	78.4402	82.5437	81.5746	82.1791	81.5278	81.763	80.1501
Nitrosopumilus_sp001543015	82.1059	81.2556	79.072	81.9945	81.3105	81.3891	81.2611	81.6146	80.6599
Nitrosopumilus_sp002317795	81.7297	80.8178	78.7837	81.7566	80.9028	81.0151	81.0934	81.3668	80.5147
Nitrosopumilus_sp002690535	81.1855	80.9111	79.682	81.2207	80.7408	80.5618	80.7803	81.0122	84.6713
Nitrosopumilus_sp002730325	81.349	80.6628	78.7146	81.1155	80.851	80.8982	80.9652	81.0395	80.0636
Nitrosopumilus_sp002788515	82.969	82.8316	78.7274	83.0049	82.6705	81.3172	82.3369	82.5623	79.8391
Nitrosopumilus_sp003702465	81.0458	80.7907	79.3593	80.9391	80.6995	80.3801	80.6091	80.8986	83.6091
Nitrosopumilus_sp003702495	80.379	80.2971	79.9808	80.399	80.4421	80.5106	80.7255	80.5415	79.3139
Nitrosopumilus_sp003702525	80.6072	80.3846	79.591	80.4009	80.5852	80.6378	81.1232	80.8235	79.2967
Nitrosopumilus_sp003702545	80.2767	80.1368	79.411	80.2942	79.9658	80.3993	80.5737	80.6414	79.0029
Nitrosopumilus_sp008080815	79.0815	78.6383	95.3316	78.6119	78.7823	78.7926	78.8041	79.0799	78.2763
Nitrosopumilus_sp013043295	83.5345	83.533	78.7407	83.3194	83.6523	81.9208	83.3077	83.8333	79.9912
Nitrosopumilus_sp013215285	82.4202	81.6625	78.6365	82.2397	81.7045	81.7044	81.2467	81.9238	80.7517
Nitrosopumilus_sp013390905_L19a	81.3614	80.9179	78.2895	81.5588	80.7809	81.2059	80.638	81.0046	79.7813
Nitrosopumilus_sp013867245	81.28	81.0588	79.5269	81.2457	80.8157	80.9189	81.2175	81.2539	83.2319
Nitrosopumilus_sp013911135	82.2098	81.3205	78.645	82.1541	81.3292	81.7268	81.2655	81.6465	80.1169
Nitrosopumilus_sp014075315	80.7429	80.6556	77.9398	80.838	80.6007	79.6253	80.7455	80.5889	79.4356
Nitrosopumilus_sp014384445	80.5229	79.6998	77.9869	80.3973	80.1421	79.4247	80.2957	80.0123	98.4637
Nitrosopumilus_sp014384555	85.7049	85.6432	79.1066	85.8473	85.4987	81.8966	85.6505	86.1123	80.8967

Nitrosopumilus_sp016125975	81.2189	80.7025	79.7133	80.934	80.8825	81.1591	81.3752	81.2777	79.5664
Nitrosopumilus_sp018263505	81.0537	80.856	79.0879	81.314	80.8356	80.6819	80.9856	81.0075	80.4888
Nitrosopumilus_sp018657375	80.4593	79.8492	77.9441	80.5929	80.0492	79.7492	79.8743	80.2134	92.6568
Nitrosopumilus_sp018668425	81.7855	81.3696	78.8211	81.874	81.2717	80.9444	81.5151	81.7221	80.2101
Nitrosopumilus_sp020431995	80.2586	80.0416	77.8258	80.0309	80.1931	80.0756	80.1878	80.0555	78.5712
Nitrosopumilus_sp021268355	80.3498	80.0381	79.8021	80.2211	80.0692	80.4332	80.3635	80.056	79.2232
Nitrosopumilus_sp022564385	81.1292	80.7397	78.2239	80.9558	80.7128	80.8816	80.3534	80.7552	79.677
Nitrosopumilus_sp022565935	81.2759	80.7668	78.0998	81.4812	80.6253	81.1534	80.7291	80.6941	79.5999
Nitrosopumilus_sp022573895	81.1652	80.6344	78.5775	81.1516	80.7186	80.7746	80.4709	80.6138	79.7006
Nitrosopumilus_sp024276695	85.2619	84.5899	79.0027	85.3129	84.4063	81.706	85.2934	84.8715	80.7355
Nitrosopumilus_sp024644105	84.0699	84.0244	79.2517	84.055	83.8369	81.9234	83.639	84.7133	80.3627
Nitrosopumilus_ureiphilus	81.7315	81.605	79.1962	81.8316	81.7068	87.1319	81.6025	81.5826	79.8856
Nitrosopumilus_zosterae	81.1725	81.0905	78.855	81.1737	81.1453	84.1497	81.3515	81.2091	79.8407

Supplementary Table 2: Metadata of reference genomes used in this study. This table lists 55 genomes, including 54 representative genomes of the genus *Nitrosopumilus*, along with *Nitrosopelagicus brevis* from the GTDB.

Species name	Accession	Source	Citation
Nitrosopumilus maritimus SCM1	GCA_000018465.1	Saltwater aquarium, Seattle, USA	(Walker et al. 2010)
Candidatus Nitrosopumilus salaria BD31	GCA_000242875.3	San Francisco Bay estuary, USA	(Mosier et al. 2012)
Candidatus Nitrosopumilus koreensis AR1	GCA_000299365.1	Arctic Ocean Sediments, Svalbard	(Park et al. 2012b)
Candidatus Nitrosopumilus sediminis AR2	GCA_000299395.1	Arctic Ocean Sediments, Svalbard	(Park et al. 2012a)
Nitrosopumilus sp000746765	GCA_000746765.1	Red Sea brine-seawater interface	
Nitrosopumilus piranensis D3C	GCA_000875775.1	Adriatic Sea, Coastal	(Bayer et al. 2016)
Nitrosopumilus adriaticus NF5	GCA_000956175.1	Adriatic Sea, Coastal	(Bayer et al. 2016)
Nitrosopumilus sp001437625	GCA_001437625.1	Baltic Sea surface water, near Bornholm island	
Nitrosopumilus sp001510275 casp-thauma3	GCA_001510275.1	Capian Sea	(Mehrshad et al. 2016)
Nitrosopumilus sp. Nsub	GCA_001541925.1	Deep Sea Sponge (*Suberites*)	
Nitrosopumilus sp001543015	GCA_001543015.1	Pacific Ocean, Sponge, South China Sea	
Candidatus Nitrosomarinus catalina SPOT01	GCA_002156965.1	Pacific Ocean: Coastal California	(Ahlgren et al. 2017)
Nitrosopumilus sp002317795	GCA_002317795.1	Atlantic Ocean: North Pond	(Tully et al. 2018)
Nitrosopumilus sp. UBA526	GCA_002506665.1	Deep Sea Sponge (*Neamphius huxleyi*)	
Nitrosopumilus sp002730325	GCA_002730325.1	Atlantic Ocean: North Atlantic Ocean	
Nitrosopumilus sp002788515	GCA_002788515.1	Crystal Geyser near Green River; Utah	

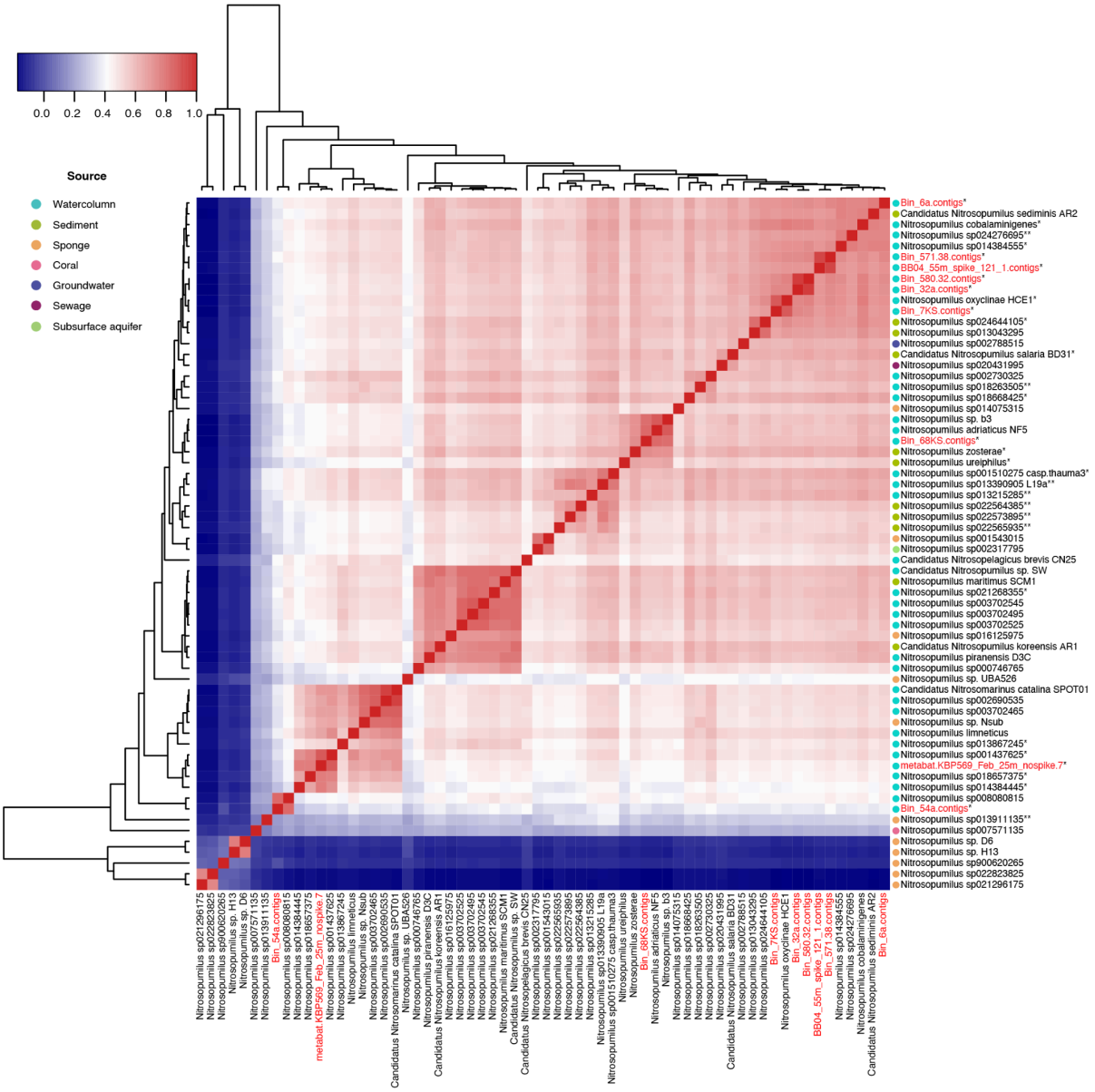
Nitrosopumilus zosterae	GCA_003175215.1	Coastal Japan; Eelgrass sediment	(Nakagawa et al. 2021)
Nitrosopumilus sp002690535	GCA_003331425.1	Mediterranean Sea: off coast of Spain	
Nitrosopumilus sp003702465	GCA_003702465.1	Gulf of Mexico, USA	
Nitrosopumilus sp003702495	GCA_003702495.1	Gulf of Mexico, USA	
Nitrosopumilus sp003702525	GCA_003702525.1	Gulf of Mexico, USA	
Nitrosopumilus sp003702545	GCA_003702545.1	Gulf of Mexico, USA	
Nitrosopumilus sp. H13	GCA_003724285.1	Atlantic Ocean: Ireland; Logachev Mounds, *Hexadella dedritifera*	(Zhang et al. 2019)
Nitrosopumilus sp. D6	GCA_003724325.1	Atlantic Ocean: Ireland; Logachev Mounds, *Hexadella dedritifera*	(Zhang et al. 2019)
Candidatus Nitrosopumilus sp. SW	GCA_006740685.1	Pacific Ocean, East Coast Korea	(Kim et al. 2016)
Nitrosopumilus sp007571135	GCA_007571135.1	Pacific Ocean, Papua New Guinea, coral host, *Coelocarteria singaporensis*	
Nitrosopumilus sp008080815	GCA_008080815.1	Atlantic Ocean, Sapelo Island, USA	
Nitrosopumilus sp013043295	GCA_013043295.1	Pacific Ocean, Port Phillip Bay sand	
Nitrosopumilus sp013215285	GCA_013215285.1	Pacific Ocean: North Pacific Gyre; Station ALOHA	
Nitrosopumilus sp013390905 L19a	GCA_013390905.1	Pacific Ocean south of Guam, 10000 deep	(Wang et al. 2019)
Nitrosopumilus cobalaminigenes	GCA_013407145.1	Pacific Ocean, Puget Sound, USA	(Qin et al. 2017)
Nitrosopumilus oxyclinae HCE1	GCA_013407165.1	Pacific Ocean, Puget Sound, USA	(Qin et al. 2017)
Nitrosopumilus ureiphilus	GCA_013407185.1	Pacific Ocean, Puget Sound sediments, USA	
Nitrosopumilus sp013867245	GCA_013867245.1	Jiulong River estuary, China	
Nitrosopumilus sp013911135	GCA_013911135.1	Gulf of Mexico; Chapopote Knoll, Asphalt Volcano, Mexico, *Hymedesmia methanophila*	
Nitrosopumilus sp014075315	GCA_014075315.1	Pacific Ocean, Marine Sponge *Mycale hentscheli*, New Zealand	
Nitrosopumilus sp. b3	GCA_014078525.1	Florida, USA, cell culture	
Nitrosopumilus sp014384445	GCA_014384445.1	Black Sea, Suspended Organic Matter	
Nitrosopumilus sp014384555	GCA_014384555.1	Black Sea, Suspended Organic Matter	
Nitrosopumilus sp016125975	GCA_016125975.1	Mediterranean Sea, coastal eastern boundary, Sponge *Petrosia ficiformis*	
Nitrosopumilus sp018263505	GCA_018263505.1	Red Sea; Suakin Deep, Brine Pool	
Nitrosopumilus sp018657375	GCA_018657375.1	Pacific Ocean, Saanich Inlet, Canada, hypoxic	
Nitrosopumilus sp018668425	GCA_018668425.1	Pacific Ocean, Saanich Inlet, Canada, hypoxic	
Nitrosopumilus sp020431995	GCA_020431995.1	Waste Water Treatment, Hong Kong	
Nitrosopumilus sp021268355	GCA_021268355.1	San Francisco Bay Estuary, USA	
Nitrosopumilus sp021296175	GCA_021296175.1	Mediterranean Sea, coastal eastern boundary, Sponge *Petrosia ficiformis*	
Nitrosopumilus limneticus	GCA_022117465.1	Lake Constance, Germany	(Klotz et al. 2022)
Nitrosopumilus sp022564385	GCA_022564385.1	Pacific Ocean: Mariana trench surficial sediment	
Nitrosopumilus sp022565935	GCA_022565935.1	Pacific Ocean: Mariana trench surficial sediment	
Nitrosopumilus sp022573895	GCA_022573895.1	Pacific Ocean: Mariana trench surficial sediment	
Nitrosopumilus sp022823825	GCA_022823825.1	Caribbean Sea, Marine Sponge tissue, Panama	
Nitrosopumilus sp024276695	GCA_024276695.1	Indian Ocean: Longqi hydrothermal vent field, black smoker	
Nitrosopumilus sp024644105	GCA_024644105.1	Waiwera River estuary sediment, New Zealand	

Nitrosopumilus sp900620265	GCA_900620265.1	Pacific Ocean, sponge Ianthella basta, Queensland, Australia	(Moeller et al. 2019)
Candidatus Nitrosopelagicus brevis CN25	GCF_000812185.1	Pacific Ocean, North Eastern, USA	(Santoro et al. 2015)

Supplementary Table 3: Metadata of the genes used in manual annotation.

Locus_tag	Gene name	Gene ID	NCBI Reference Sequence	NCBI description	Uniprot description
NMAR_RS08050	amoA	5774202	WP_012215883.1	ammonia monooxygenase [*Nitrosopumilus maritimus SCM1*]	
NMAR_RS08065	amoB	5774142	WP_012215886.1	methane monooxygenase/ammonia monooxygenase subunit B [*Nitrosopumilus maritimus SCM1*]	
NMAR_RS08060	amoC	5774105	WP_016939485.1	methane monooxygenase/ammonia monooxygenase subunit C [*Nitrosopumilus maritimus SCM1*]	
NMAR_RS05915	UvrA	5774147	WP_012215487.1	excinuclease ABC subunit UvrA [*Nitrosopumilus maritimus SCM1*]	
NMAR_RS05920	UvrB	5773173	WP_012215488.1	excinuclease ABC subunit UvrB [*Nitrosopumilus maritimus SCM1*]	
NMAR_RS05910	UvrC	5773952	WP_012215486.1	excinuclease ABC subunit UvrC [*Nitrosopumilus maritimus SCM1*]	
NMAR_RS07200	ectC	5773382	WP_012215727.1	ectoine synthase [*Nitrosopumilus maritimus SCM1*]	
NMAR_RS07195	thpD	5773792	WP_012215726.1	ectoine hydroxylase [*Nitrosopumilus maritimus SCM1*]	
NMAR_RS07210	ectA	5774098	WP_012215729.1	diaminobutyrate acetyltransferase [*Nitrosopumilus maritimus SCM1*]	
NMAR_RS07205	ectB	5773101	WP_012215728.1	diaminobutyrate--2-oxoglutarate transaminase [*Nitrosopumilus maritimus SCM1*]	
NPIRD3C_1392	ureA	41600519	WP_148703419.1	urease subunit gamma [*Nitrosopumilus piranensis*]	
NPIRD3C_1393	ureB	41600520	WP_148703420.1	urease subunit beta [*Nitrosopumilus piranensis*]	
NPIRD3C_1394	UreC	41600521	WP_148703421.1	urease subunit alpha [*Nitrosopumilus piranensis*]	
NPIRD3C_1388	UreD	41600515	WP_148703415.1	urease accessory protein UreD [*Nitrosopumilus piranensis*]	
NPIRD3C_1391	UreE	41600518	WP_148703418.1	Urease accessory protein [*Nitrosopumilus piranensis*]	
NPIRD3C_1390	UreF	41600517	WP_148703417.1	urease accessory UreF family protein [*Nitrosopumilus piranensis*]	
NPIRD3C_1389	UreG	41600516	WP_148703416.1	urease accessory protein UreG [*Nitrosopumilus piranensis*]	
NADRNF5_0800	CheY	24820030	WP_148313063.1	response regulator [*Nitrosopumilus adriaticus*]	
NADRNF5_0801	CheR	24820031	WP_048118956.1	protein-glutamate O-methyltransferase CheR [*Nitrosopumilus adriaticus*]	
NADRNF5_0802	CheC	24820032	WP_048115873.1	chemotaxis protein CheC [*Nitrosopumilus adriaticus*]	
NADRNF5_0803	CheB	24820033	WP_052661870.1	chemotaxis protein CheA [*Nitrosopumilus adriaticus*]	
NADRNF5_0804	CheB	24820034	WP_082051980.1	chemotaxis-specific protein-glutamate methyltransferase CheB [*Nitrosopumilus adriaticus*]	
NADRNF5_0806	CheW	24820036	WP_237089331.1	chemotaxis protein CheW [*Nitrosopumilus adriaticus*]	
NADRNF5_0807	CheR	24820037	WP_048115876.1	methyl-accepting chemotaxis protein [*Nitrosopumilus adriaticus*]	
NADRNF5_0808	FlaJ	24820038	WP_237089332.1	flagellar assembly protein FlaJ [Nitrosopumilus adriaticus]	
NADRNF5_0809	FlaI	24820039	WP_048115878.1	type II/IV secretion system ATPase subunit [*Nitrosopumilus adriaticus*]	
NADRNF5_0810	FlaH	24820040	WP_048115879.1	ATPase domain-containing protein [*Nitrosopumilus adriaticus*]	

NADRNF5_0811	FlaF	24820041	WP_048115880.1	hypothetical protein [*Nitrosopumilus adriaticus*]	Putative flagella assembly protein FlaF
NADRNF5_0812	FlaG	24820042	WP_048115881.1	flagellin [*Nitrosopumilus adriaticus*]	
NADRNF5_0814	FlaB1	24820044	WP_048115882.1	flagellin [*Nitrosopumilus adriaticus*]	
NADRNF5_0815	FlaB2	24820045	WP_048115883.1	flagellin [*Nitrosopumilus adriaticus*]	
NADRNF5_0816	FlaB3	24820046	WP_048115884.1	flagellin [*Nitrosopumilus adriaticus*]	
NADRNF5_0817	FlaB4	24820047	WP_048115885.1	flagellin [*Nitrosopumilus adriaticus*]	
NADRNF5_0819	FlaB5	24820049	WP_048115886.1	flagellin [*Nitrosopumilus adriaticus*]	
NMAR_RS02585	pstA	5774691	WP_012214865.1	phosphate ABC transporter permease PstA [*Nitrosopumilus maritimus SCM1*]	
NMAR_RS02590	pstB	5774694	WP_012214866.1	phosphate ABC transporter ATP-binding protein PstB [*Nitrosopumilus maritimus SCM1*]	
NMAR_RS02580	pstC	5774687	WP_012214864.1	phosphate ABC transporter permease subunit PstC [*Nitrosopumilus maritimus SCM1*]	
NMSP_RS06040	dndA	32901714	WP_086907923.1	cysteine desulfurase family protein [*Candidatus Nitrosomarinus catalina*]	
NMSP_RS06040	dndB	32901718	WP_086907927.1	DNA sulfur modification protein DndB [*Candidatus Nitrosomarinus catalina*]	
NMSP_RS06055	dndC	32901717	WP_086907926.1	DNA phosphorothioation system sulfurtransferase DndC [*Candidatus Nitrosomarinus catalina*]	
NMSP_RS06050	dndD	32901716	WP_192866156.1	DNA sulfur modification protein DndD [*Candidatus Nitrosomarinus catalina*]	
NMSP_RS06045	dndE	32901715	WP_192866155.1	DndE family protein [*Candidatus Nitrosomarinus catalina*]	



Supplementary Figure 2: Heatmap of genome correlation based on gene counts per ortholog. This symmetrical heatmap displays the Pearson correlation coefficients between 64 *Nitrosopumilus* genomes. These coefficients are computed from the number of genes contributing to each ortholog, as determined by OrthoFinder. The color gradient reflects the correlation coefficient, with red indicating higher similarity and blue indicating lower similarity.

Academic Thesis: Declaration Of Authorship

I, Sookyoung Kim declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Ammonia-oxidizing archaea in the Baltic Sea: dynamics and enrichment-enabled comparative genomics

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Either none of this work has been published before submission.

Signed:

Date: