Probe capture enrichment sequencing of *amoA* genes discloses diverse ammonia-oxidizing archaeal and bacterial populations

Satoshi Hiraoka[1†]*, Minoru Ijichi[2†], Hirohiko Takeshima[2], Yohei Kumagai[2], Ching-Chia Yang[2], Yoko Makabe-Kobayashi[2], Hideki Fukuda[2], Susumu Yoshizawa[2], Wataru Iwasaki[2,3], Kazuhiro Kogure[2], Takuhei Shiozaki[2]*

1. Research Center for Bioscience and Nanoscience (CeBN), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2–15 Natsushima-cho, Yokosuka, Kanagawa 237–0061, Japan
2. Atmosphere and Ocean Research Institute, the University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8564, Japan
3. Department of Integrated Biosciences, Graduate School of Frontier Sciences, the University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-0882, Japan.

Present address

- Minoru Ijichi: Department of Biological Sciences, Graduate School of Science, Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji, Tokyo 192-0397, Japan
- Minoru Ijichi: Field Application LS, Life Science Sales, QIAGEN, 3-13-1 Kachidoki, Chuo-ku, Tokyo 104-0054, Japan
- Hirohiko Takeshima: Research Center for Marine Bioresources, Faculty of Marine Bioscience, Fukui Prefectural University, 49-8-2 Katsumi, Obama, Fukui, 917-0003, Japan
- Yohei Kumagai: tayo, 1-7-101 Noukendai-higashi, Kanazawa-ku, Yokohama, Kanagawa 236-0058, Japan
- Ching-Chia Yang: Institute of Biomedical Informatics, National Yang Ming Chiao Tung University, No. 155, Sec. 2, Linong St. Beitou Dist., Taipei City 112304, Taiwan
- Kazuhiro Kogure: University of the Ryukyus, Senbaru, Nishihara, Okinawa 903-0213, Japan

Running headline: Probe capture enrichment sequencing of *amoA* genes

† *Contributed equally*

* *Corresponding author*

Email: hiraokas@jamstec.go.jp, shiozaki@g.ecc.u-tokyo.ac.jp

Abstract

The ammonia monooxygenase subunit A (*amoA*) gene has been used to investigate the phylogenetic diversity, spatial distribution, and activity of ammonia-oxidizing archaeal (AOA) and bacterial (AOB) populations that contribute greatly to the nitrogen cycle in various ecosystems. Amplicon sequencing of *amoA* genes is a widely used method; however, it generally produces inaccurate results owing to the lack of a 'universal' primer set. Moreover, currently available primer sets suffer from amplification biases, which can lead to severe misinterpretation. Although shotgun metagenomic and metatranscriptomic analyses are alternative approaches without amplification bias, the low concentration of target genes in heterogeneous environmental DNA restricts comprehensive analysis to a realizable sequencing depth. Here, we developed a method for *amoA* gene enrichment sequencing using a probe capture hybridization technique. Using mock community samples, our approach effectively enriched *amoA* genes with low compositional changes, outperforming amplification and meta-omics sequencing analyses. Following the analysis of marine metatranscriptome samples, we predicted 79 AOA and AOB operational taxonomic units (OTUs), of which 33 (42%) and 60 (76%) were unidentified by simple metatranscriptomic and amplicon sequencing, respectively. Mapped read ratios were significantly higher for the capture samples (50.4 ± 27.6%) than for non-capture samples (0.05 ± 0.02%), demonstrating the high enrichment efficiency of the method. The analysis also revealed the spatial diversity of AOA ecotypes with high sensitivity and phylogenetic resolution, which are difficult to examine using conventional approaches.

Importance

The nitrogen cycle is largely driven by prokaryotes via nitrogen metabolism, including ammonia oxidation. To understand the phylogenetic diversity, spatial distribution, and potential activity of these ammonia-oxidizing prokaryotes, the ammonia monooxygenase subunit A (*amoA*) gene has been used. However, current conventional approaches, such as *amoA* amplicon sequencing and meta-omics (metagenomics and metatranscriptomics), generally produce inaccurate results owing to PCR biases and low concentrations of target genes in the samples. Here, as an alternative, we developed a novel enrichment sequencing method that condenses the *amoA* gene in metagenomic/metatranscriptomic samples via probe capture hybridization for high-throughput sequencing. We demonstrated that this method is applicable to both artificial mock and oceanic samples and generates reliable results with high phylogenetic resolution and sensitivity.

Keywords

Ammonia-oxidizing prokaryotes / *amoA* gene / Target gene enrichment / Metagenome / Metatranscriptome / Marine prokaryotes

Introduction

Nitrification is a two-step process involving ammonia oxidation followed by nitrite oxidation. Ammonia oxidation is a rate-limiting process driven by ammonia-oxidizing archaea (AOA) and bacteria (AOB) that utilize the ammonia monooxygenase subunit A (*amoA*) gene. This process is fundamental to the nitrogen cycle, including the emission of nitrous oxide, a primary greenhouse gas (Fowler et al., 2013; Niu et al., 2016), and occurs in various ecosystems, such as soil, hot springs, freshwater, estuaries, sediments, and marine ecosystems (Isobe & Ohte, 2014; Lehtovirta-Morley, 2018; Monteiro et al., 2014; Takai, 2019). A substantial proportion of the ammonia oxidation process on Earth is believed to occur in the ocean via AOAs (Hutchins & Capone, 2022; Voss et al., 2013). Oceanic AOAs are largely classified into three ecotypes: water column cluster A (WCA), water column cluster B (WCB), and *Nitrosopumilus maritimus*–like cluster (NMC, known as NPM) (Santoro et al., 2019). WCA and NMC are reported to be abundant just below the euphotic zone (Beman et al., 2008; Santoro et al., 2019; Shiozaki et al., 2016), whereas WCB inhabits deeper waters (Beman et al., 2008; Santoro et al., 2019; Shiozaki et al., 2016). According to the recently updated taxonomy based on the *amoA* gene phylogeny, WCA, WCB, and NMC correspond to NP-ε-2, NP-α-2.2.2.1, and NP-γ-2.1, respectively (Alves et al., 2018; Beman et al., 2008; Francis et al., 2005). While AOA contribute greatly to ammonia oxidization in general, AOBs are compatible with AOAs in environments with high substrate availability (e.g., agricultural soil), regardless of their lower relative abundance (Jia & Conrad, 2009; Prosser et al., 2020; Sterngren et al., 2015). Among the AOBs, complete ammonia oxidizers (comammox) can perform the entire two-step nitrification process alone, are found in both terrestrial and coastal areas, and are considered important contributors to nitrification on a global scale (Daims et al., 2015; Fei et al., 2018; van Kessel et al., 2015).

amoA amplicon sequencing is a commonly used approach to characterize AOAs and AOBs in prokaryotic communities. This methodology relies on polymerase chain reaction (PCR) for selective amplification of *amoA* gene without cultivation. Currently, *amoA* is the second most frequently sequenced marker gene in microbiology, preceded by the 16S ribosomal RNA (rRNA) gene (Alves et al., 2018), allowing us to investigate the phylogenetic diversity and spatial distribution of diverse AOA and AOB members in various environments, including the ocean (Table S1) (Beman et al., 2008; Francis et al., 2005; O'Mullan & Ward, 2005). However, PCR amplification is not an all-encompassing technique. First, it is difficult to design a universal primer set for all AOA and AOB owing to the lack of a conserved region in the cording sequence of *amoA* gene. Although a number of primer sets with different targeted clades have been proposed (Beman et al., 2008; Coolen et al., 2007; Francis et al., 2005; Hallam et al., 2006; Hornek et al., 2006; Leininger et al., 2006; Löscher et al., 2012; Mincer et al., 2007; Mosier & Francis, 2011; Pester et al., 2012; Pjevac et al., 2017; Purkhold et al., 2000; Rotthauwe et al., 1997; Sintes et al., 2013; Wuchter et al., 2006; Yakimov et al., 2011), a true 'universal' primer set is yet to be developed. Second, even among the targeted clades of the used primer set, PCR amplification frequently causes unexpected results owing to mismatches between the primer and targeted gene sequences, differences in the efficiency of denaturation and annealing steps, and chimeric amplification, known as PCR biases (Jolinda et al., 2018). For example, the most frequently used *amoA* primer set, Arch-amoAF/R (Francis et al., 2005), results in mismatches against many AOA *amoA* sequences that further leads to the underestimation of total AOAs (Tolar et al., 2013) and is unable to amplify any AOB *amoA* gene. Taken together,

the current *amoA* amplicon sequencing approach produces distorted results, leading to misinterpretation in downstream analyses (Dechesne et al., 2016; Ijichi & Hamasaki, 2017; Shimomura et al., 2012).

Alternatively, shotgun metagenomics and metatranscriptomics employ high-throughput sequencing technology to generate community-wide sequencing reads without PCR amplification. These strategies allowed us to analyze diverse gene sequences to estimate the original lineages and expression levels of each gene of interest, providing a complete picture of taxonomic diversity and activities, respectively (Hiraoka et al., 2016). However, target genes typically exist at low concentrations in heterogeneous DNA samples. The concentration of *amoA* gene is estimated to be up to 10^8 copies per ng DNA in soil (Bahram et al., 2022; Bannert et al., 2011; Ding et al., 2020; Han et al., 2013; Marusenko et al., 2013), wastewater (Spasov et al., 2020; Wang et al., 2021; Zhang et al., 2015), estuary (Hollibaugh et al., 2011), and seawater (Christman et al., 2011) using quantitative real-time PCR (qPCR), and less than 0.001% and 0.05% of metagenomic and metatranscriptomic reads, respectively, were assigned to the *amoA* gene from soil samples (Bahram et al., 2022; Clark et al., 2021; Paungfoo-Lonhienne et al., 2017). This low concentration limits the community-wide analysis of genes of interest with realizable sequencing depth.

Targeted gene enrichment is an innovative approach that overcomes the limitations of PCR amplification and meta-omics. This method relies on solution hybrid selection (SHS) to recover the genes of interest from heterogeneous DNAs for sequencing. Hybridization authorizes mismatches between the designed probes and distant targeted sequences, allowing us to investigate large pools of divergent genes, including previously unknown ones (Hedtke et al., 2013). The hybridization capture experiments demonstrated high reproducibility (Chilamakuri et al., 2014; Gnirke et al., 2009; Hernandez-Rodriguez et al., 2018; Mamanova et al., 2010). In field research, this method has been applied to various genes as biomarkers, including methyl coenzyme M reductase subunit A (Denonfoux et al., 2013), endo-xylanase (Bragalini et al., 2014), carbohydrate-active enzyme and protease (Manoharan et al., 2015), reductive dehalogenase, insertion element, hypothetical protein from *Dehalococcoidia* (Biderre-Petit et al., 2016), benzylsuccinate synthase alpha subunit (Ranchou-Peyruse et al., 2017), lindane degradation enzyme (Gasc & Peyret, 2018), CH_3Cl dehalogenase (Jaeger et al., 2018), hexachlorocyclohexane degradation enzyme (Gasc & Peyret, 2017), 16S rRNA (Beaudry et al., 2021; Biderre-Petit et al., 2016; Cariou et al., 2018; Gasc & Peyret, 2018; Günther et al., 2022; Sara et al., 2020), and 18S rRNA (Günther et al., 2022). A pioneering study also attempted the enrichment sequencing of *amoA* gene recently (Siljanen et al., 2022). However, to the best of our knowledge, no study has focused on *amoA* gene with full-length sequence reconstruction and *de novo* identification of operational taxonomic units (OTUs) for the precise analysis of phylogenetic diversity despite their great importance in the nitrogen cycle.

Here, we developed a probe capture enrichment sequencing method for the effective analysis of *amoA* genes in metagenomic and metatranscriptomic samples. Capture probes were designed using *amoA* gene sequences from a wide range of prokaryotes retrieved from public databases to cover diverse *amoA* genes, including novel ones that would hybridize with the probes. Technical validation using an artificial mock community showed the high efficiency of the approach without severe compositional disorder. The analysis of metatranscriptomic marine samples collected in the euphotic to mesopelagic layers from coastal to open ocean areas (Figure S1, Table S2) showed outstanding performance compared with conventional shotgun and amplicon sequencing approaches. The analysis also showed that the spatial distribution of AOA ecotypes could be explained by niche separation through the vertical

water column and distance from the land area, providing great insights into nitrogen cycling in the oceanic environment and highlighting the advantages of our *amoA* enrichment sequencing approach.

Materials and methods

Capture probe design

The *amoA* gene belongs to the copper membrane monooxygenase (CuMMO) superfamily, which is composed of several sequentially close but functionally different genes (e.g., *amoA*, *pmoA*, *pxmA*, *bmoA*, *peoA*, *etnA*, *pamA*, *AlkB*, *and bmoX*) that accept different substrates (e.g., ammonia, methane, propene, ethene, and butane). CuMMOs show complicated phylogenies, and thus, it is challenging to distinguish them based on sequence similarity (Alves et al., 2018; Coleman et al., 2012; Diamond et al., 2022; Orellana et al., 2016; Tavormina et al., 2011). To design capture probes that cover a wide range of *amoA* genes, we designed probes to capture diverse CuMMO genes, allowing us to manually select potential *amoA* genes for downstream sequence analysis. We manually collected 20 well-annotated representative CuMMO subunit A genes spanning a wide range of phylogenetic clades retrieved from the NCBI database as an initial database (Alves et al., 2018; Coleman et al., 2012; Diamond et al., 2022; Orellana et al., 2016; Tavormina et al., 2011) (Supplementary Data S1). The sequences include four *amoA* genes from all major AOA and AOB groups (Thaumarchaeota, Betaproteobacteria, Gammaproteobacteria, and Nitrospirota). Using the initial database, 29683 sequences were retrieved from NCBI nt and env_nt (March 2016) using blastn search with a cut-off e-value ≤ 1E-5, and 195 *amoA* gene sequences were manually collected from GenBank (Supplementary Data S2). A total of 29878 CuMMO sequences (referred to as the CuMMO sequence database) were then used to design 95584 100-mer probes using NimbleDesign (Roche NimbleGen). To increase the recruitment power of metatranscriptome samples, sequences that showed similarity to the 16S rRNA gene from the Ribosomal Database Project (RDP) database (Cole et al., 2014) were removed. Finally, capture beads were constructed using a SeqCap RNA Developer Enrichment Kit (Roche NimbleGen, Madison, WI, USA).

Mock sample construction and DNA sample preparation

We prepared an artificial mixture of plasmids encoding six *amoA* genes as a mock sample to examine the condition and efficiency of our probe capture enrichment method. Six *amoA* sequences were selected from three AOA and three AOB lineages: *Nitrosopumilus maritimus* SCM1^{T} (CP000866), *Candidatus* (*Ca.*) Nitrosopelagicus brevis CN25 (CP007026), and Thaumarchaeota archaeon SCGC AAA007-O23 (ARWO00000000), and *Nitrosomonas europaea* ATCC 19718^{T} (AL954747), *Nitrosococcus oceani* ATCC 19707^{T} (NC_007484), and *Ca.* Nitrospira inopinata ENR4 (LN885086), respectively. Six sequences were artificially synthesized by Eurofins Genomics (Ota, Tokyo, Japan) and cloned into the pTAKN-2 cloning vector (BioDynamics Laboratory, Bunkyo, Tokyo, Japan). Each plasmid was transformed into *Escherichia coli* DH5α competent cells (Takara Bio, Kusatsu, Shiga, Japan) according

to the manufacturer's instructions. *E. coli* was cultured in LB liquid medium supplemented with kanamycin, and plasmids were extracted using the PureLink HiPure Plasmid FP (Filter and Precipitator) Maxiprep Kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The quality and quantity of plasmid DNA were measured using NanoDrop 2000 (Thermo Fisher Scientific, Wilmington, DE, USA) and Microplate Reader SH-9000Lab (Corona Electric, Hitachinaka, Ibaraki, Japan) with a Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen) according to the manufacturer's instructions.

Seawater sampling and RNA sample preparation

Seawater samples were collected at five and six stations along the Otsuchi (OT) and Onagawa (ON) transect lines (from the coastal area inside each bay to the pelagic area), respectively, during cruise KS16-01 of the R/V Shinsei-maru in March 2016 (Table S1, Figure S1). The sampling stations are located in the temperate western North Pacific Ocean. Sampling permits for expeditions in Japan's exclusive economic zone were not required because our work was centered on domestic areas and did not involve endangered or protected species. Seawater was collected using a bucket from the surface (0 m below sea level [mbsl]) or using Niskin bottle samplers (General Oceanic, Miami, Florida, USA) in deeper layers ranging from 33 to 500 mbsl (epipelagic to mesopelagic zones). Water temperature and salinity in the sampling layer were measured using an SBE9plus CTD system (Sea-Bird Scientific, Bellevue, WA, USA). For the molecular experiments, approximately 8 L of seawater was filtered onto 0.22-µm Sterivex-GP pressure filter units (Millipore, Billerica, MA, USA). The filters were immediately soaked in 2 mL of RNAlater (Ambion) and maintained at 4 °C for 12 h of immersion. The filters were then stored at −30 °C until subsequent molecular experiments. We also collected samples for nutrient analyses in duplicate 10-mL acrylic tubes.

RNA was extracted from the filters onshore using the *mir*Vana miRNA Isolation Kit (Ambion), purified using the TURBO DNA-free Kit (Ambion), and concentrated using RNeasy MinElute Cleanup (QIAGEN, Hilden, Germany). The quality and quantity of RNA were measured using a Nanodrop 2000 and Quantus fluorometer with the QiantiFluor RNA System (Promega, Madison, WI, USA) according to the manufacturer's instructions. Concentrations of nitrate (NO_3^-), nitrite (NO_2^-), ammonium (NH_4^+), and phosphate (PO_4^{3-}) were measured colorimetrically using an autoanalyzer as described previously (Fukuda et al., 2016). The detection limits of the used system were 0.05, 0.02, 0.1, and 0.02 µM for NO_3^-, NO_2^-, NH_4^+, and PO_4^{3-} measurements, respectively.

Library construction, probe capture, and sequencing

Sequencing libraries of metagenomic mock samples and metatranscriptomic seawater samples were prepared using the KAPA Library Preparation Kit (Roche NimbleGen) and the KAPA Stranded RNA-Seq Library Preparation Kit (Roche NimbleGen), respectively, with ERCC RNA Spike-In Mix (Ambion) according to the manufacturer's instructions. Before library construction, the quality of all RNA samples was checked using a 2100 Bioanalyzer (Agilent Technologies) with an RNA Pico Kit (Agilent Technologies), which confirmed that no obvious RNA was observed. DNA and RNA were shared to a size of 200 bp by sonication using Covaris (Woburn, MA, USA)

and exposure to divalent cations in accordance with the SeqCap EZ Developer Kit and SeqCap RNA Developer Enrichment Kit, respectively.

To evaluate potential biases in probe capture enrichment and optimize the methodology, different settings were tested using the mock sample and a seawater sample collected from 0 mbsl at St2 station (Table 1). In this experiment, times of the probe capture hybridization were varied with zero, once, and twice, referred to as 'non-capture,' 'single-capture,' and 'double-capture' settings, respectively. The non-capture setting thus indicates an unenriched shotgun sequencing that can produce a bias-free community composition. Before probe capture hybridization, libraries were prepared as described above and mixed. Hybridization was performed using the SHS method according to the SeqCap EZ Reagent Kit Plus (Roche NimbleGen) with modifications to the number of captures, number of ligation-mediated PCR (LM-PCR) cycles, and template volume for LM-PCR amplification. For the non-capture setting, libraries were prepared without the hybridization or LM-PCR amplification steps. For single-capture settings, LM-PCR amplification after the hybridization step was performed for 20 cycles according to the manufacturer's instructions. For the double-capture settings, hybridization was performed twice with different combinations of the first and second LM-PCR amplification (pairs of 5 and 20 (5/20), 7 and 14 (7/14), and 7 and 20 (7/20) cycle settings for the mock sample, and 5/20 and 7/14 cycle settings for the seawater sample). Template volume for the LM-PCR amplification was 5 μL in all cases. For library size selection, DNA fragments <150 bp in length were removed using the GeneRead Size Selection Kit (QIAGEN) according to the manufacturer's instructions.

To evaluate the effect of dilution of the targeted gene, mock DNA was pooled with pTAC-2 (BioDynamics Laboratory) to vary the *amoA* gene concentration. The molar concentration ratios of mock DNA to pTAC-2 DNA were adjusted to 2:1000 (named Mock-copy1e6), 2:10000 (Mock-copy1e5), and 2:100000 (Mock-copy1e4). The final concentrations were approximately equivalent to 10^6, 10^5, and 10^4 *amoA* gene copies per ng of DNA, where the lengths of the *amoA* genes, pTAKN-2, and pTAC-2 were 653–846, 2739, and 2786 bp, respectively. The sequencing libraries for the three samples were prepared as described above. Hybridization was then performed with the single-capture settings to obtain capture samples, whereas without the step for non-capture samples. LM-PCR amplifications for the Mock-copy1e6, Mock-copy1e5, and Mock-copy1e4 samples were performed for 16, 18, and 20 cycles, respectively, to obtain a sufficient amount of DNA for sequencing. The subsequent steps were performed as described above.

For metatranscriptomic seawater samples, libraries from each transect were mixed and size-selected as described above, followed by probe capture hybridization. Evaluation using mock samples suggested that the single-capture setting was optimal for obtaining sequencing reads that reflected the original community structure (see the Results section). Thus, the pooled cDNA libraries were hybridized using the single-capture settings. The libraries proceeded with the hybridization and amplification steps as capture samples but not as non-capture samples. The sample name represents the library preparation type, sampling station, and water depth; for example, 'capture_OT6-200m' represents the sample collected from OT6 station at 200 mbsl layer and prepared with probe capture enrichment method, and 'noncapture_ON1-B-5m' represents the sample collected from ON1 station at the layer 5 m above the seafloor and without probe capturing. The remaining libraries were then mixed for sequencing.

The final libraries of each pool were sequenced (2 × 300 bp paired-end reads) using the Illumina MiSeq sequencing platform (Illumina, San Diego, California, USA) at the Bioengineering Lab (Atsugi, Kanagawa, Japan).

Amplicon sequencing

To compare the results with those of the probe capture enrichment and shotgun sequencing methods, near-full-length *amoA* amplicon sequencing was performed. The most frequently used *amoA* primer set (Arch-amoAF:5′-STAATGGTCTGGCTTAGACG-3′, and Arch-amoAR: 5′-GCGGCCATCCATCTGTATGT-3′) (Francis et al., 2005), hereafter referred to as 'Francis primers,' was used for the amplification. For this analysis, mock metagenomic sample and three metatranscriptomic seawater samples (OT3-0m, OT3-50m, and OT3-B-5m) were selected. For metatranscriptome samples, cDNA was prepared using the PrimeScript II 1st Strand cDNA Synthesis Kit (TaKaRa Bio). The *amoA* amplicon was prepared using two-step PCR method. For the first targeted PCR step, Francis primers were used to amplify *amoA* sequence region using KAPA HiFi HotStart ReadyMix PCR Kit (Roche). Thermal cycles were as follows: 95 °C for 2 min, (98 °C for 20 s, 65 °C for 15 s, 72 °C for 30 s) cycle for 35 times, and 72 °C for 5 min. The products were then subjected to a second barcoding PCR to add index sequences using the same PCR kit. Next, the amplicons were subjected to single-molecule real-time (SMRT) sequencing. SMRT libraries were prepared according to the 'Amplification of bacterial full-length 16S gene with barcoded primers' protocol (Pacific Biosciences of California, Menlo Park, California, USA). The final SMRT libraries were sequenced using a PacBio Sequel IIe system (Pacific Biosciences of California) in the circular consensus sequencing (CCS) mode to generate accurate high-fidelity (HiFi) reads at the Bioengineering Lab (Sagamihara, Kanagawa, Japan).

Bioinformatics

For Illumina short-read sequencing reads, both ends of the reads containing low-quality bases (Phred quality score < 20) and adapter sequences were trimmed using TrimGalore (https://github.com/FelixKrueger/TrimGalore) with default settings. Sequencing reads derived from the PhiX genome were removed using Bowtie2 (Langmead & Salzberg, 2012) with default settings. Sequences with low complexity or shorter than 100 bp were discarded using PRINSEQ++ (Cantu et al., 2019) with default settings. The remaining high-quality (HQ) paired-end reads were *de novo* assembled using metaSPAdes (Nurk et al., 2017) or rnaSPAdes (Bushmanova et al., 2019) with default options for metagenome and metatranscriptome samples, respectively. CDSs with >33 aa length in the assembled contigs were predicted using Prodigal (Hyatt et al., 2010) with '-meta' setting. Next, *amoA* genes were searched using DIAMOND (Buchfink et al., 2021) with '—max-target-seqs 1 —evalue 1e-30 —subject-cover 50 —query-cover 50 —id 60' settings against the CuMMO sequence database. CDSs <400 or >1000 bp were excluded from further analysis. Chimeric sequences were also removed using VSEARCH (Rognes et al., 2016) with '—dn 0.5' setting. Remained sequences were clustered into 97% sequence similarity with >50% length coverage using MMseq2 (Steinegger & Söding, 2017) with '—cov-mode 0' setting. After discarding singletons, each cluster was designated as an operational taxonomic unit (OTU). As similar to the typical 16S rRNA amplificon sequencing analysis, we used the 'sample

concatenating' approach: all the retrieved sequences from multiple samples were pooled for clustering to define OTU. A phylogenetic tree of the OTUs was constructed using MAFFT (Katoh & Standley, 2013) for multiple alignments and FastTree2 (Price et al., 2010) using the maximum-likelihood method (JTT + CAT model) for tree estimation. The taxonomy of the AOA OTUs was assigned using DIAMOND (Buchfink et al., 2021) against the AOA *amoA* sequence database defined by Alves et al. (Alves et al., 2018) after manual curation. Ecotypes of AOA OTUs were estimated based on taxonomy and phylogeny according to previous definitions (Beman et al., 2008; Francis et al., 2005). The functions of the encoding genes and the original taxonomy of each OTU were estimated via annotation using DIAMOND (Buchfink et al., 2021) against the CuMMO sequence database and a phylogenetic tree topology. For abundance estimation, the HQ paired-end reads were merged using FLASH (Magoč & Salzberg, 2011) with default settings (HQ merged reads) and subject to the read mapping onto the OTUs using Bowtie2 (Langmead & Salzberg, 2012) with '-N 1' setting. The mapped reads were used to calculate the reads per kilobase of genes per million reads sequenced (RPKMS) for each OTU. The abundance of RPKMS in each ecotype, along with the coordinates and water depths, was visualized using Ocean Data View (https://odv.awi.de/).

For HiFi amplicon sequencing reads, CCS reads containing at least three full-pass subreads on each polymerase read with >99% average base-call accuracy were retained as HiFi reads using the standard PacBio SMRT software package with the default settings. Reads <400 or >1000 bp were removed for further analysis. To remove the primer sequences, 20 bp in both terminals of the HiFi reads was trimmed using SeqKit (Shen et al., 2016). Chimeric sequences were removed using VSEARCH (Rognes et al., 2016) with '—dn 0.5' setting. The remaining reads (HQ HiFi reads) were mapped to OTUs using Bowtie2 (Langmead & Salzberg, 2012) to calculate the RPKMS, as described above.

Data availability

All Illumina short-read and PacBio HiFi read sequences were deposited in the DDBJ Sequence Read Archive (DRA) (Supplementary Table S2). All data are registered under the BioProject ID PRJDB15057 [https://ddbj.nig.ac.jp/resource/bioproject/ PRJDB15057]. The designed capture probes are available in HyperDesign under the Custom Design Name 16TKYKT05.

Results

Validation using metagenomic mock community samples

To evaluate the efficiency of the probe capture enrichment approach, we first prepared an artificial mock DNA sample as the ground truth. The mock sample was composed of six plasmids into which different *amoA* genes were transformed at equal molar concentrations. The *amoA* genes were selected from three AOAs and three AOBs belonging to phylogenetically distant lineages (see Materials and Methods). Using the mock sample, we sought good

experimental settings for a number of hybridization and LM-PCR conditions, which could affect the efficiency of probe capture enrichment. OTUs at a 97% sequence identity threshold were *de novo* identified from the sequence data produced using five different experimental settings. In addition, we performed *amoA* amplicon sequencing using a common primer set (Francis primers) (Francis et al., 2005) for comparison with a probe capture enrichment approach.

Using our methodology (Figure 1), seven OTUs were predicted from a pool of four capture and one non-capture samples (Figure 2A, Table S3). Among the seven OTUs, all six *amoA* sequences induced in the plasmids were successfully retrieved with complete sequence matches. Only one of the other OTU (OTU_mock6) showed high sequence similarity to those from marine uncultured AOA (Thaumarchaeota archaeon casp-thauma1), possibly indicating unexpected mixed contamination in the sample before probe hybridization. The OTU represented only up to 0.09% of reads through the capture samples (single_Mock and double_Mock), suggesting little impact on the analysis and rather the high sensitivity of the approach. The mapped ratio of sequence reads to the OTUs was 15.8% in the non-capture sample (noncapture_Mock). This value was roughly as expected; the ratio of nucleotide bases belonging to *amoA* genes in the construct was close to 21%, where the lengths of *amoA* genes and the cloning vector were 651–846 bp and 2739 bp, respectively. In contrast, the ratios from the capture samples were higher: the ratio from single-capture (single_Mock-cycle20) was 72.2%, whereas those from double-capture (double_Mock-cycle5/20, double_Mock-cycle7/14, double_Mock-cycle7/20) ranged from 75.4 to 79.8%. Almost all reads from the amplicon sequencing were mapped to the OTUs (99.9%, amplicon_Mock).

The relative RPKMS abundance of the six OTUs corresponding to the induced *amoA* genes was generally consistent even under non-capture settings, whereas it fluctuated slightly in the capture samples. This trend was accounted for by the decrease in OTU_mock7 (*Ca*. Nitrospira inopinata ENR4), with abundances of 5.7% and 0.6% in the single-capture and double-capture settings, respectively. Similarly, OTU_mock1 (*Nitrosococcus oceani* ATCC 19707) showed a slight decreasing tendency with an increase in the hybridization step. This bias may be explained by the lower recruitment power of the probes against AOB, owing to the small size of AOB *amoA* sequences in our sequence database used for probe design. Only one OTU (OTU_mock3, *Ca.* Nitrosopelagicus brevis CN25, which belongs to the WCA ecotype (Santoro et al., 2015)) occupied almost all parts of the amplicon sequencing sample (99.0%, amplicon_Mock). This was likely due to primer bias; the primer set used for amplification (Francis primers) perfectly matched the *amoA* sequences from *Ca.* Nitrosopelagicus brevis CN25 possessed three and two mismatches to the other two AOA *amoA* sequences (*Nitrosopumilus maritimus* SCM1 and Thaumarchaeota archaeon SCGC AAA007-O23, respectively) and showed no significant match to any of the three AOB *amoA* sequences. Overall, despite the potential slight bias on some AOBs, the probe capture enrichment strategy effectively enriched *amoA* genes with low compositional changes and high sensitivity that overwhelm metagenomic and amplicon sequencing approaches.

Study of targeted gene concentration sensitivity

Next, we evaluated the effect of target gene concentration on probe capture enrichment efficiency. Diluted mock samples were prepared, in which the total concentrations of the six *amoA* genes were adjusted to 10^6, 10^5, and 10^4 copies/ng of DNA using commercially available plasmid DNA, referred to as Mock-copy1e6, Mock-

copy1e5, and Mock-copy1e4, respectively. The adjusted concentrations were compatible with those in a wide range of environments such as soil (Bahram et al., 2022; Bannert et al., 2011; Ding et al., 2020; Han et al., 2013; Marusenko et al., 2013), wastewater (Spasov et al., 2020; Wang et al., 2021; Zhang et al., 2015), estuary (Hollibaugh et al., 2011), and seawater (Christman et al., 2011). OTUs were *de novo* predicted using pooled reads from three capture and three non-capture samples. Using probe capture enrichment and our bioinformatic methodology with 'single-capture' settings against diluted samples, all six induced *amoA* genes were successfully retrieved (Figure 2B). Mapped read ratios on the OTUs decreased with the dilution rate; 63.1%, 53.7%, and 22.3% were mapped from samples with 10^6, 10^5, and 10^4 copies per ng DNA (capture_Mock-copy1e6, capture_Mock-copy1e5, and capture_Mock-copy1e4, respectively). In contrast, only 0.0–0.1% of the reads were mapped from the non-capture samples (noncapture_Mock-copy1e6, noncapture_Mock-copy1e5, and noncapture_Mock-copy1e4). When comparing the ratios between samples with the same *amoA* concentration, the capture samples showed 990–8100 times higher values than the non-capture samples. The relative RPKMS abundances of the OTUs were generally even within the capture samples, whereas they were disturbed in the non-capture samples owing to the severely low numbers of mapped reads. Overall, the results demonstrated the high efficiency and specificity of our probe capture enrichment approach under conditions of low target gene concentrations, analogous to the natural environment.

Study of enrichment settings for metatranscriptomic seawater samples

To elucidate the efficiency of the probe capture enrichment approach for metatranscriptomic data and the potential impact of LM-PCR settings on performance, we analyzed cDNA libraries prepared using a coastal marine water sample collected at station St2. Using pooled reads from four samples with different LM-PCR settings, eight OTUs were identified (Supplementary Figure S2). The mapped read ratios were low in the single-capture sample (0.5%, single_St2-cycle20) and the original metatranscriptomic sample (0.1%, non-capture_St2). The numbers of sequenced reads from non-capture and single-capture samples were 2.3M and 2.2M, respectively. The relative abundances were concordant between the two samples, suggesting a low compositional effect in the single-capture enrichment settings. Notably, we detected a comammox member (OTU_st2-8; assigned to *Ca.* Nitrospira inopinata) that is hardly amplified using the Francis primers (Daims et al., 2015; van Kessel et al., 2015). In contrast to the non-capture and single-capture samples, the double-capture samples showed higher mapped read ratio and lower numbers of sequenced reads (86.0% and 92.6%, and 36.4 K and 5.6 K reads from double_St2-cycle5/20 and double_St2-cycle7/14, respectively) (Supplementary Table S3). This indicates a highly selective but weak recruitment efficiency in the second probe-hybridization step. In addition, likely because of the low number of sequenced reads, RPKMS relative abundances in the double_St2-cycle7/14 were disturbed compared to the other samples. Taken together, the analysis demonstrated that the single-capture settings are applicable to marine metatranscriptomic samples.

Probe capture enrichment sequencing of metatranscriptomic seawater samples

Using the single-capture settings described above, we conducted probe capture enrichment sequencing of *amoA* genes against 39 transcriptomic seawater samples collected at 11 stations spanning the coastal to open ocean at different water depths (Figure S1, Table S2). In addition to capture and non-capture samples, we performed near-full-length *amoA* amplicon sequencing of three transcriptomic seawater samples using the Francis primers for comparison. Despite of the equal mol concentrations of libraries in a pool before the hybridization step, the numbers of sequenced reads from capture samples ranged from 0.3 to 3.3 M with an increasing tendency along with water depth and distance from land (Table S3). This likely reflects the general tendency of the abundance of ammonia oxidizers that are more abundant at the surface than in deeper layers in coastal areas and outside than inside the bay, as widely observed previously (Lu et al., 2020; Molina et al., 2020; Newell et al., 2013; Shafiee et al., 2021; Tolar et al., 2013; Zou et al., 2020). This tendency was also concordant with the geochemical analysis, which showed that nitrate (NO_3^-) and ammonium (NH_4^+) concentrations were lower in the samples collected at the surface than near the bottom and in the coastal areas than in the pelagic areas (Table S2). In contrast, as expected, the read numbers were generally even (0.96–1.89 M) for the non-capture samples. Notably, among the amplicon sequencing samples, 34.2% (amplicon_OT3-50m) and 34.2% (amplicon_OT3-B-5m) of the data were lost, mainly because of chimeric read filtering (Table S3), indicating drastic chimeric production during the amplification step.

The probe capture enrichment approach revealed the expression levels of *amoA* genes by ammonia oxidizers in the community with high resolution (Figure 3). We identified 87 OTUs, of which at least 77 and 2 OTUs were assigned to *amoA* genes from AOA and AOB, respectively. Among the other OTUs, two and four were estimated to be particulate methane monooxygenase (*pmoA*) and Cytochrome P450 genes from methane-oxidizing bacteria (MOB) and Actinomycetota, respectively, both of which are included in CuMMOs and are concordant with the designed probe (see Materials and Methods). The other two OTUs showed moderate (94.2%) and low (84.8%) sequence similarity to known *amoA* genes as described by Alves et al. (Alves et al., 2018), and considering the phylogenetic topology, it was difficult to conclude whether unprecedented AOA *amoA* genes or functionally different. In contrast, the possible 77 AOA OTUs showed 97.4–100% sequence similarity to *amoA* genes in the sequence database (Alves et al., 2018). Mapped read ratios were significantly higher for the capture samples (0.8%–82.0%, 50.4 ± 27.6% on average) than for non-capture samples (0.02%–0.08%, 0.05 ± 0.02%) ($p < 0.05$, U-test), and almost all the reads (>99.3%) from the amplicon sequencing corresponded to the OTUs. In addition, the number of detected OTUs in capture samples (ranging from 35 to 85, 60.9 on average) was significantly higher than that in non-capture samples (ranging from 4 to 27, 11.3 on average) ($p < 0.05$, U-test), demonstrating the sensitivity of the approach. Those from the amplicon sequencing ranged from 10 to 19, and no AOBs were detected. Among the AOA and AOB OTUs, 33 (42%) and 60 (76%) were unidentified in the non-capture and amplicon sequencing samples, respectively.

All the AOA OTUs were assigned to order Nitrosopumilales and spanned to four order-level clades (NP-α, NP-δ, NP-ε, NP-γ). NP-α and NP-ε were reported to be observed in pelagic marine water, whereas NP-δ was both coastal water and marine sediment, and NP-γ was composed of members inhabiting all the three oceanic environments (Alves et al., 2018), as consistent with the samples (Figure S1, Table S2). The NMDS analysis showed significantly association of the composition with water depth ($p<0.05$, permutation tests for regression models implemented by

'envfit' function, vegan package) (Figure S3). Especially, in the capture samples, expression levels of WCB (OTUs assigned to NP-α-2.2.2.1) were increased along with water depth, whereas NMC (NP-γ-2.1) were decreased and WCA (NP-ε-2) were evenly distributed (Figure 4). Overall, NMC was the dominant group, followed by WCB and WCA. These findings were scarcely observed in non-capture samples. In contrast to capture and non-capture sequencing, WCA was the most dominant ecotype in the expression profiles from amplicon sequencing, which is likely due to the PCR bias caused by the Francis primers, as described above. These results are consistent with those of previous studies that investigated the AOA community structures in the ocean using *amoA* amplicon sequencing (Lu et al., 2019; Santoro et al., 2017; Shiozaki et al., 2016; Wuchter et al., 2006).

Discussions

We developed a probe capture enrichment sequencing method for the *amoA* gene for effective investigation of the community structure of ammonia oxidizers and their potential activities (Figure 1). By applying this approach to metagenomic mock community samples, our method successfully reconstructed the true community structure, demonstrating that the approach can provide a qualitatively reliable estimation of AOAs and AOBs composition in the community sample. In addition, this approach successfully identified *amoA* expression levels of diverse ammonia oxidizers in oceanic samples with high phylogenetic resolution. This performance was contrary to that of the shotgun metatranscriptome and common amplicon sequencing, both of which resulted in highly divergent profiles with lower resolution. Thus, probe capture enrichment sequencing should provide detailed and comprehensive information on *amoA* genes in the prokaryotic community, thereby contributing to a substantial understanding of the prokaryotic nitrogen cycle in the environment.

In contrast to 16S rRNA genes, which contain nine hypervariable region (V1-V9) flanked by highly conserved sequences through the bacterial and archaeal domains, there are no well-conserved regions in most functional genes, including *amoA*. This severely restricts the design of a 'universal' primer set that perfectly matches diverse *amoA* genes and covers many AOAs and AOBs. Thus, common *amoA* amplicon sequencing can underestimate many lineages of ammonia oxidizers because of primer mismatches, and in contrast, it will overestimate specific members who possess genes with higher binding affinity to the primer. Indeed, our experiments revealed significant PCR biases that *amoA* gene from *Ca.* Nitrosopelagicus brevis CN25, a member of the WCA family, was specifically amplified by amplicon sequencing (Figure 2). The same trend was also observed in the field metatranscriptomic samples: WCA was predominantly detected by amplicon sequencing, whereas NMC was more abundant by probe capture enrichment and simple shotgun sequencing (Figure 3). Current knowledge of the ecotypic distribution in the ocean is thus likely affected by primer biases, and further reconsideration of the potential contribution of NMC as well as WCA to the nitrogen cycle in the ocean is needed.

Meta-omics (i.e., metagenomics and metatranscriptomics) relies on shotgun sequencing and is now widely used to investigate the phylogenetic and genomic diversity of AOAs and AOBs, as well as the potential metabolisms involved in the nitrogen cycle. One advantage of this approach is that it avoids PCR bias, allowing us to recover sequences that would escape conventional PCR amplification (Cheung et al., 2019; Delmont et al., 2018). In this study, OTUs that were undetected by amplicon sequencing were identified by probe capture enrichment and simple

shotgun sequencing (Figure 3). Using a shotgun sequencing approach, it was recently reported that ammonia oxidizers are more diverse in the ocean than previously thought (Cheung et al., 2019). However, as we demonstrated, the low concentration of the target genes/transcripts in the heterogeneous community severely limits the efficiency of the approach. Here, the probe capture enrichment approach showed a greater number and more diverse expression levels of *amoA* gene than simple shotgun sequencing. This trend is similar to that observed in a previous study focusing on 16S rRNA (Gasc & Peyret, 2018). Consequently, the probe capture enrichment approach provides more informative results for *amoA* sequences than the typical omics-based methods.

The *amoA* probe capture enrichment sequencing methodology is not limited to the oceanic environment but is applicable to any community sample, including soil, fresh water, sediment, and wastewater treatment plants. Further applications of this method will lead to innovative findings regarding the nitrogen cycle and microbial ecology. We should note that, however, the method still needs to take care of the potential biases and weak recruitment power. Multiple hybridizations (i.e., double-capture settings) generated disturbed results, whereas single-capture settings generated low numbers of sequencing reads in some cases (Figure 2A, Table S3). These issues warrant further improvements in the method to obtain more accurate results through economic sequencing efforts. In addition, likely due to our probe design, we found a possible underestimation of AOBs, including comammox (Figure 2A). Further collection of *amoA* gene sequences for probe design would increase the recruitment power, thereby improving the accuracy of enrichment, especially in the case of AOBs.

Conclusion

This study demonstrated that the probe capture enrichment sequencing approach can effectively recover the community structure of ammonia oxidizers and the expression levels of *amoA* genes in the community, outperforming conventional methods. From the application of this method to oceanic transcriptome samples, we detected a higher diversity of AOAs than ever before, even if their expression levels were likely low in total microbial gene expression. Importantly, from the analysis of oceanic samples, we found that amplicon sequencing incorporating the most common primer set caused a significant PCR bias (i.e., most ammonia oxidizers were missed, and selective WCA members were amplified), resulting in low phylogenetic resolution and inconsistent relative abundance profiles. This indicates that *amoA* amplicon sequencing generally produces results far from the true composition, although almost all current knowledge of the phylogenetic diversity and spatial distribution of ammonia oxidizers has been based on conventional techniques. Thus, this approach could be an alternative tool to examine the community structure and biological responses of ammonia oxidizers at high resolution to understand the geochemical contributions of nitrogen-cycling prokaryotes in vast environments on Earth.

Acknowledgements

We would like to thank the captain, crew, and onboard scientists and technicians of the R/V Shinsei-maru (JAMSTEC) during the KS16-01 cruise. Bioinformatic analysis was performed on the supercomputing system at National Institute of Genetics (NIG), Research Organization of Information and Systems (ROIS), Japan.

Contributions

SH designed the study, performed the bioinformatics analyses, and wrote the manuscript. MI conceived and designed the study, performed the sampling and molecular experiments, and wrote the manuscript. HT designed and performs the molecular experiments. YK contributed to the probe design. CCY contributed to the bioinformatics analyses. YMK contributed to the molecular experiments. HF performed nutrient measurements. SY, WI, and KK supervised the project. TS designed the study, wrote the manuscript, and supervised the project. All authors read and approved the final manuscript.

Funding

This work was supported by The Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (JPMXD1521474594), Japan Society for the Promotion of Science (JP19H04263, JP20K15444, JP21H03592, JP22H03716, JP22H04925), and Japan Science and Technology Agency, CREST (JPMJCR11A3, JPMJCR19S2)

Ethics declarations

Minoru Ijichi is an employee of QIAGEN (Chuo-ku, Tokyo, JAPAN) and a former employee of the Bioengineering Lab (Sagamihara, Kanagawa, Japan). The remaining authors declare no conflicts of interest.

References

Alves, R. J. E., Minh, B. Q., Urich, T., von Haeseler, A., & Schleper, C. (2018). Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on *amoA* genes. *Nature Communications*, *9*(1), 1517. https://doi.org/10.1038/s41467-018-03861-1

Bahram, M., Espenberg, M., Pärn, J., Lehtovirta-Morley, L., Anslan, S., Kasak, K., Kõljalg, U., Liira, J., Maddison, M., Moora, M., Niinemets, Ü., Öpik, M., Pärtel, M., Soosaar, K., Zobel, M., Hildebrand, F., Tedersoo, L., & Mander, Ü. (2022). Structure and function of the soil microbiome underlying N2O emissions from global wetlands. *Nature Communications*, *13*(1), 1430. https://doi.org/10.1038/s41467-022-29161-3

Bannert, A., Kleineidam, K., Wissing, L., Mueller-Niggemann, C., Vogelsang, V., Welzl, G., Cao, Z., & Schloter, M. (2011). Changes in diversity and functional gene abundances of microbial communities involved in nitrogen fixation, nitrification, and denitrification in a tidal wetland versus paddy soils cultivated for different time periods. *Applied and Environmental Microbiology*, *77*(17), 6109–6116. https://doi.org/10.1128/AEM.01751-10

Beaudry, M. S., Wang, J., Kieran, T. J., Thomas, J., Bayona-Vásquez, N. J., Gao, B., Devault, A., Brunelle, B., Lu, K., Wang, J.-S., Rhodes, O. E., & Glenn, T. C. (2021). Improved microbial community characterization of 16S rRNA via metagenome hybridization capture enrichment. *Frontiers in Microbiology*, *12*, 644662. https://doi.org/10.3389/fmicb.2021.644662

Beman, J. M., Popp, B. N., & Francis, C. A. (2008). Molecular and biogeochemical evidence for ammonia oxidation by marine Crenarchaeota in the Gulf of California. *The ISME Journal*, *2*(4), 429–441. https://doi.org/10.1038/ismej.2007.118

Biderre-Petit, C., Dugat-Bony, E., Mege, M., Parisot, N., Adrian, L., Moné, A., Denonfoux, J., Peyretaillade, E., Debroas, D., Boucher, D., & Peyret, P. (2016). Distribution of *Dehalococcoidia* in the anaerobic deep water of a remote meromictic crater lake and detection of *Dehalococcoidia*-derived reductive dehalogenase homologous genes. *PLOS ONE*, *11*(1), e0145558. https://doi.org/10.1371/journal.pone.0145558

Bragalini, C., Ribière, C., Parisot, N., Vallon, L., Prudent, E., Peyretaillade, E., Girlanda, M., Peyret, P., Marmeisse, R., & Luis, P. (2014). Solution hybrid selection capture for the recovery of functional full-length eukaryotic cDNAs from complex environmental samples. *DNA Research : An International Journal for Rapid Publication of Reports on Genes and Genomes*, *21*(6), 685–694. https://doi.org/10.1093/dnares/dsu030

Buchfink, B., Reuter, K., & Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, *18*(4), 366–368. https://doi.org/10.1038/s41592-021-01101-x

Bushmanova, E., Antipov, D., Lapidus, A., & Prjibelski, A. D. (2019). rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *GigaScience*, *8*(9), giz100. https://doi.org/10.1093/gigascience/giz100

Cantu, V. A., Sadural, J., & Edwards, R. (2019). PRINSEQ++, a multi-threaded tool for fast and efficient quality control and preprocessing of sequencing datasets. *PeerJ Preprints*, *7*, e27553v1. https://doi.org/10.7287/peerj.preprints.27553v1

Cariou, M., Ribière, C., Morlière, S., Gauthier, J.-P., Simon, J.-C., Peyret, P., & Charlat, S. (2018). Comparing 16S rDNA amplicon sequencing and hybridization capture for pea aphid microbiota diversity analysis. *BMC Research Notes*, *11*(1), 461. https://doi.org/10.1186/s13104-018-3559-3

Cheung, S., Mak, W., Xia, X., Lu, Y., Cheung, Y., & Liu, H. (2019). Overlooked genetic diversity of ammonia oxidizing archaea lineages in the global oceans. *Journal of Geophysical Research: Biogeosciences*, *124*(7), 1799–1811. https://doi.org/10.1029/2018JG004636

Chilamakuri, C. S. R., Lorenz, S., Madoui, M.-A., Vodák, D., Sun, J., Hovig, E., Myklebost, O., & Meza-Zepeda, L. A. (2014). Performance comparison of four exome capture systems for deep sequencing. *BMC Genomics*, *15*(1), 449. https://doi.org/10.1186/1471-2164-15-449

Christman, G. D., Cottrell, M. T., Popp, B. N., Gier, E., & Kirchman, D. L. (2011). Abundance, diversity, and activity of ammonia-oxidizing prokaryotes in the coastal arctic ocean in summer and winter. *Applied and Environmental Microbiology*, *77*(6), 2026–2034. https://doi.org/10.1128/AEM.01907-10

Clark, I. M., Hughes, D. J., Fu, Q., Abadie, M., & Hirsch, P. R. (2021). Metagenomic approaches reveal differences in genetic diversity and relative abundance of nitrifying bacteria and archaea in contrasting soils. *Scientific Reports*, *11*(1), 15905. https://doi.org/10.1038/s41598-021-95100-9

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., Brown, C. T., Porras-Alfaro, A., Kuske, C. R., & Tiedje, J. M. (2014). Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Research*, *42*(D1), D633–D642. https://doi.org/10.1093/nar/gkt1244

Coleman, N. v, Le, N. B., Ly, M. A., Ogawa, H. E., McCarl, V., Wilson, N. L., & Holmes, A. J. (2012). Hydrocarbon monooxygenase in Mycobacterium: recombinant expression of a member of the ammonia monooxygenase superfamily. *The ISME Journal*, *6*(1), 171–182. https://doi.org/10.1038/ismej.2011.98

Coolen, M. J. L., Abbas, B., van Bleijswijk, J., Hopmans, E. C., Kuypers, M. M. M., Wakeham, S. G., & Sinninghe Damsté, J. S. (2007). Putative ammonia-oxidizing Crenarchaeota in suboxic waters of the Black Sea: a basin-wide ecological study using 16S ribosomal and functional genes and membrane lipids. *Environmental Microbiology*, *9*(4), 1001–1016. https://doi.org/https://doi.org/10.1111/j.1462-2920.2006.01227.x

Daims, H., Lebedeva, E. v, Pjevac, P., Han, P., Herbold, C., Albertsen, M., Jehmlich, N., Palatinszky, M., Vierheilig, J., Bulaev, A., Kirkegaard, R. H., von Bergen, M., Rattei, T., Bendinger, B., Nielsen, P. H., & Wagner, M. (2015). Complete nitrification by Nitrospira bacteria. *Nature*, *528*(7583), 504–509. https://doi.org/10.1038/nature16461

Dechesne, A., Musovic, S., Palomo, A., Diwan, V., & Smets, B. F. (2016). Underestimation of ammonia-oxidizing bacteria abundance by amplification bias in *amoA*-targeted qPCR. *Microbial Biotechnology*, *9*(4), 519–524. https://doi.org/https://doi.org/10.1111/1751-7915.12366

Delmont, T. O., Quince, C., Shaiber, A., Esen, Ö. C., Lee, S. T. M., Rappé, M. S., McLellan, S. L., Lücker, S., & Eren, A. M. (2018). Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nature Microbiology*, *3*(7), 804–813. https://doi.org/10.1038/s41564-018-0176-9

Denonfoux, J., Parisot, N., Dugat-Bony, E., Biderre-Petit, C., Boucher, D., Morgavi, D. P., Le Paslier, D., Peyretaillade, E., & Peyret, P. (2013). Gene capture coupled to high-throughput sequencing as a strategy for targeted metagenome exploration. *DNA Research*, *20*(2), 185–196. https://doi.org/10.1093/dnares/dst001

Diamond, S., Lavy, A., Crits-Christoph, A., Matheus Carnevali, P. B., Sharrar, A., Williams, K. H., & Banfield, J. F. (2022). Soils and sediments host Thermoplasmata archaea encoding novel copper membrane monooxygenases (CuMMOs). *The ISME Journal*, *16*(5), 1348–1362. https://doi.org/10.1038/s41396-021-01177-5

Ding, J., Ma, M., Jiang, X., Liu, Y., Zhang, J., Suo, L., Wang, L., Wei, D., & Li, J. (2020). Effects of applying inorganic fertilizer and organic manure for 35 years on the structure and diversity of ammonia-oxidizing archaea communities in a Chinese Mollisols field. *MicrobiologyOpen*, *9*(1), e00942. https://doi.org/https://doi.org/10.1002/mbo3.942

Fei, X., Jian-Gong, W., Ting, Z., Bin, Z., Sung-Keun, R., & Zhe-Xue, Q. (2018). Ubiquity and diversity of complete ammonia oxidizers (comammox). *Applied and Environmental Microbiology*, *84*(24), e01390-18. https://doi.org/10.1128/AEM.01390-18

Fowler, D., Coyle, M., Skiba, U., Sutton, M. A., Cape, J. N., Reis, S., Sheppard, L. J., Jenkins, A., Grizzetti, B., Galloway, J. N., Vitousek, P., Leach, A., Bouwman, A. F., Butterbach-Bahl, K., Dentener, F., Stevenson, D., Amann, M., & Voss, M. (2013). The global nitrogen cycle in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1621), 20130164. https://doi.org/10.1098/rstb.2013.0164

Francis, C. A., Roberts, K. J., Beman, J. M., Santoro, A. E., & Oakley, B. B. (2005). Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(41), 14683–14688. https://doi.org/10.1073/pnas.0506625102

Fukuda, H., Katayama, R., Yang, Y., Takasu, H., Nishibe, Y., Tsuda, A., & Nagata, T. (2016). Nutrient status of Otsuchi Bay (northeastern Japan) following the 2011 off the Pacific coast of Tohoku Earthquake. *Journal of Oceanography*, *72*(1), 39–52. https://doi.org/10.1007/s10872-015-0296-2

Gasc, C., & Peyret, P. (2017). Revealing large metagenomic regions through long DNA fragment hybridization capture. *Microbiome*, *5*(1), 33. https://doi.org/10.1186/s40168-017-0251-0

Gasc, C., & Peyret, P. (2018). Hybridization capture reveals microbial diversity missed using current profiling methods. *Microbiome*, *6*(1), 61. https://doi.org/10.1186/s40168-018-0442-3

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S., & Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, *27*(2), 182–189. https://doi.org/10.1038/nbt.1523

Günther, B., Marre, S., Defois, C., Merzi, T., Blanc, P., Peyret, P., & Arnaud-Haond, S. (2022). Capture by hybridization for full-length barcode-based eukaryotic and prokaryotic biodiversity inventories of deep sea ecosystems. *Molecular Ecology Resources*, *22*(2), 623–637. https://doi.org/https://doi.org/10.1111/1755-0998.13500

Hallam, S. J., Mincer, T. J., Schleper, C., Preston, C. M., Roberts, K., Richardson, P. M., & DeLong, E. F. (2006). Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine *Crenarchaeota*. *PLoS Biology*, *4*(4), e95. https://doi.org/10.1371/journal.pbio.0040095

Han, J., Jung, J., Park, M., Hyun, S., & Park, W. (2013). Short-term effect of elevated temperature on the abundance and diversity of bacterial and archaeal *amoA* genes in Antarctic soils. *Journal of Microbiology and Biotechnology*, *23*(9), 1187–1196. https://doi.org/10.4014/jmb.1305.05017

Hedtke, S. M., Morgan, M. J., Cannatella, D. C., & Hillis, D. M. (2013). Targeted enrichment: Maximizing orthologous gene comparisons across deep evolutionary time. *PLoS ONE*, *8*(7), e67908. https://doi.org/10.1371/journal.pone.0067908

Hernandez-Rodriguez, J., Arandjelovic, M., Lester, J., de Filippo, C., Weihmann, A., Meyer, M., Angedakin, S., Casals, F., Navarro, A., Vigilant, L., Kühl, H. S., Langergraber, K., Boesch, C., Hughes, D., & Marques-Bonet, T. (2018). The impact of endogenous content, replicates and pooling on genome capture from faecal samples. *Molecular Ecology Resources*, *18*(2), 319–333. https://doi.org/https://doi.org/10.1111/1755-0998.12728

Hiraoka, S., Yang, C., & Iwasaki, W. (2016). Metagenomics and bioinformatics in microbial ecology: Current status and beyond. *Microbes and Environments*, *31*(3), 204–212. https://doi.org/10.1264/jsme2.ME16024

Hollibaugh, J. T., Gifford, S., Sharma, S., Bano, N., & Moran, M. A. (2011). Metatranscriptomic analysis of ammonia-oxidizing organisms in an estuarine bacterioplankton assemblage. *The ISME Journal*, *5*(5), 866–878. https://doi.org/10.1038/ismej.2010.172

Hornek, R., Pommerening-Röser, A., Koops, H.-P., Farnleitner, A. H., Kreuzinger, N., Kirschner, A., & Mach, R. L. (2006). Primers containing universal bases reduce multiple *amoA* gene specific DGGE band patterns when analysing the diversity of beta-ammonia oxidizers in the environment. *Journal of Microbiological Methods*, *66*(1), 147–155. https://doi.org/https://doi.org/10.1016/j.mimet.2005.11.001

Hutchins, D. A., & Capone, D. G. (2022). The marine nitrogen cycle: new developments and global change. *Nature Reviews Microbiology*, *20*(7), 401–414. https://doi.org/10.1038/s41579-022-00687-z

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*(1), 119. https://doi.org/10.1186/1471-2105-11-119

Ijichi, M., & Hamasaki, K. (2017). Distinctive physiological response of shallow and deep ecotypes of ammonia-oxidizing marine archaea in seawater cultures. *Plankton and Benthos Research*, *12*(4), 259–265. https://doi.org/10.3800/pbr.12.259

Isobe, K., & Ohte, N. (2014). Ecological perspectives on microbes involved in N-cycling. *Microbes and Environments*, *29*(1), 4–16. https://doi.org/10.1264/jsme2.ME13159

Jaeger, N., Besaury, L., Röhling, A. N., Koch, F., Delort, A.-M., Gasc, C., Greule, M., Kolb, S., Nadalig, T., Peyret, P., Vuilleumier, S., Amato, P., Bringel, F., & Keppler, F. (2018). Chloromethane formation and degradation in the fern phyllosphere. *Science of The Total Environment*, *634*, 1278–1287. https://doi.org/https://doi.org/10.1016/j.scitotenv.2018.03.316

Jia, Z., & Conrad, R. (2009). Bacteria rather than Archaea dominate microbial ammonia oxidation in an agricultural soil. *Environmental Microbiology*, *11*(7), 1658–1671. https://doi.org/https://doi.org/10.1111/j.1462-2920.2009.01891.x

Jolinda, P., Laura, G., Trong, W., & Mick, W. (2018). The madness of microbiome: Attempting to find consensus "best practice" for 16S microbiome studies. *Applied and Environmental Microbiology*, *84*(7), e02627-17. https://doi.org/10.1128/AEM.02627-17

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lehtovirta-Morley, L. E. (2018). Ammonia oxidation: Ecology, physiology, biochemistry and why they must all come together. *FEMS Microbiology Letters*, *365*(9), fny058. https://doi.org/10.1093/femsle/fny058

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G. W., Prosser, J. I., Schuster, S. C., & Schleper, C. (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature*, *442*(7104), 806–809. https://doi.org/10.1038/nature04983

Löscher, C. R., Kock, A., Könneke, M., LaRoche, J., Bange, H. W., & Schmitz, R. A. (2012). Production of oceanic nitrous oxide by ammonia-oxidizing archaea. *Biogeosciences*, *9*(7), 2419–2429. https://doi.org/10.5194/bg-9-2419-2012

Lu, Y., Cheung, S., Chen, L., Kao, S.-J., Xia, X., Gan, J., Dai, M., & Liu, H. (2020). New insight to niche partitioning and ecological function of ammonia oxidizing archaea in subtropical estuarine ecosystem. *Biogeosciences*, *17*(23), 6017–6032. https://doi.org/10.5194/bg-17-6017-2020

Lu, Y., Xia, X., Cheung, S., Jing, H., & Liu, H. (2019). Differential distribution and determinants of ammonia oxidizing archaea sublineages in the oxygen minimum zone off Costa Rica. *Microorganisms*, *7*(10), 453. https://doi.org/10.3390/microorganisms7100453

Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957–2963. https://doi.org/10.1093/bioinformatics/btr507

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J., & Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nature Methods*, *7*(2), 111–118. https://doi.org/10.1038/nmeth.1419

Manoharan, L., Kushwaha, S. K., Hedlund, K., & Ahrén, D. (2015). Captured metagenomics: large-scale targeting of genes based on 'sequence capture' reveals functional diversity in soils. *DNA Research*, *22*(6), 451–460. https://doi.org/10.1093/dnares/dsv026

Marusenko, Y., Bates, S. T., Anderson, I., Johnson, S. L., Soule, T., & Garcia-Pichel, F. (2013). Ammonia-oxidizing archaea and bacteria are structured by geography in biological soil crusts across North American arid lands. *Ecological Processes*, *2*(1), 9. https://doi.org/10.1186/2192-1709-2-9

Mincer, T. J., Church, M. J., Taylor, L. T., Preston, C., Karl, D. M., & DeLong, E. F. (2007). Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. *Environmental Microbiology*, *9*(5), 1162–1175. https://doi.org/https://doi.org/10.1111/j.1462-2920.2007.01239.x

Molina, V., Belmar, L., Levipan, H. A., Ramírez-Flandes, S., Anguita, C., Galán, A., Montes, I., & Ulloa, O. (2020). Spatiotemporal distribution of key pelagic microbes in a seasonal oxygen-deficient coastal upwelling system of the eastern south pacific ocean. *Frontiers in Marine Science*, *7*, 561597. https://doi.org/10.3389/fmars.2020.561597

Monteiro, M., Séneca, J., & Magalhães, C. (2014). The history of aerobic ammonia oxidizers: from the first discoveries to today. *Journal of Microbiology*, *52*(7), 537–547. https://doi.org/10.1007/s12275-014-4114-0

Mosier, A. C., & Francis, C. A. (2011). Determining the distribution of marine and coastal ammonia-oxidizing archaea and bacteria using a quantitative approach. In M. G. Klotz (Ed.), *Methods in Enzymology* (Vol. 486, pp. 205–221). Academic Press. https://doi.org/10.1016/B978-0-12-381294-0.00009-2

Newell, S. E., Fawcett, S. E., & Ward, B. B. (2013). Depth distribution of ammonia oxidation rates and ammonia-oxidizer community composition in the Sargasso Sea. *Limnology and Oceanography*, *58*(4), 1491–1500. https://doi.org/https://doi.org/10.4319/lo.2013.58.4.1491

Niu, S., Classen, A. T., Dukes, J. S., Kardol, P., Liu, L., Luo, Y., Rustad, L., Sun, J., Tang, J., Templer, P. H., Thomas, R. Q., Tian, D., Vicca, S., Wang, Y.-P., Xia, J., & Zaehle, S. (2016). Global patterns and substrate-based mechanisms of the terrestrial nitrogen cycle. *Ecology Letters*, *19*(6), 697–709. https://doi.org/https://doi.org/10.1111/ele.12591

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Research*, *27*(5), 824–834. https://doi.org/10.1101/gr.213959.116

O'Mullan, G. D., & Ward, B. B. (2005). Relationship of temporal and spatial variabilities of ammonia-oxidizing bacteria to nitrification rates in Monterey Bay, California. *Applied and Environmental Microbiology*, *71*(2), 697–705. https://doi.org/10.1128/AEM.71.2.697-705.2005

Orellana, L. H., Rodriguez-R, L. M., & Konstantinidis, K. T. (2016). ROCker: accurate detection and quantification of target genes in short-read metagenomic data sets by modeling sliding-window bitscores. *Nucleic Acids Research*, *45*(3), gkw900. https://doi.org/10.1093/nar/gkw900

Paungfoo-Lonhienne, C., Wang, W., Yeoh, Y. K., & Halpin, N. (2017). Legume crop rotation suppressed nitrifying microbial community in a sugarcane cropping soil. *Scientific Reports*, *7*(1), 16707. https://doi.org/10.1038/s41598-017-17080-z

Pester, M., Rattei, T., Flechl, S., Gröngröft, A., Richter, A., Overmann, J., Reinhold-Hurek, B., Loy, A., & Wagner, M. (2012). *amoA*-based consensus phylogeny of ammonia-oxidizing archaea and deep sequencing of *amoA* genes from soils of four different geographic regions. *Environmental Microbiology*, *14*(2), 525–539. https://doi.org/https://doi.org/10.1111/j.1462-2920.2011.02666.x

Pjevac, P., Schauberger, C., Poghosyan, L., Herbold, C. W., van Kessel, M. A. H. J., Daebeler, A., Steinberger, M., Jetten, M. S. M., Lücker, S., Wagner, M., & Daims, H. (2017). *AmoA*-targeted polymerase chain reaction primers for the specific detection and quantification of comammox *Nitrospira* in the environment. *Frontiers in Microbiology*, *8*, 1508. https://doi.org/10.3389/fmicb.2017.01508

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLoS ONE*, *5*(3), e9490. https://doi.org/10.1371/journal.pone.0009490

Prosser, J. I., Hink, L., Gubry-Rangin, C., & Nicol, G. W. (2020). Nitrous oxide production by ammonia oxidizers: Physiological diversity, niche differentiation and potential mitigation strategies. *Global Change Biology*, *26*(1), 103–118. https://doi.org/https://doi.org/10.1111/gcb.14877

Purkhold, U., Pommerening-Röser, A., Juretschko, S., Schmid, M. C., Koops, H.-P., & Wagner, M. (2000). Phylogeny of all recognized species of ammonia oxidizers based on comparative 16S rRNA and *amoA* sequence analysis: Implications for molecular diversity surveys. *Applied and Environmental Microbiology*, *66*(12), 5368–5382. https://doi.org/10.1128/AEM.66.12.5368-5382.2000

Ranchou-Peyruse, M., Gasc, C., Guignard, M., Aüllo, T., Dequidt, D., Peyret, P., & Ranchou-Peyruse, A. (2017). The sequence capture by hybridization: a new approach for revealing the potential of mono-aromatic hydrocarbons

bioattenuation in a deep oligotrophic aquifer. *Microbial Biotechnology*, *10*(2), 469–479. https://doi.org/10.1111/1751-7915.12426

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, *4*, e2584. https://doi.org/10.7717/peerj.2584

Rotthauwe, J. H., Witzel, K. P., & Liesack, W. (1997). The ammonia monooxygenase structural gene *amoA* as a functional marker: Molecular fine-scale analysis of natural ammonia-oxidizing populations. *Applied and Environmental Microbiology*, *63*(12), 4704–4712. https://doi.org/10.1128/aem.63.12.4704-4712.1997

Santoro, A. E., Dupont, C. L., Richter, R. A., Craig, M. T., Carini, P., McIlvin, M. R., Yang, Y., Orsi, W. D., Moran, D. M., & Saito, M. A. (2015). Genomic and proteomic characterization of " *Candidatus* Nitrosopelagicus brevis": An ammonia-oxidizing archaeon from the open ocean. *Proceedings of the National Academy of Sciences*, *112*(4), 1173–1178. https://doi.org/10.1073/pnas.1416223112

Santoro, A. E., Richter, R. A., & Dupont, C. L. (2019). Planktonic Marine Archaea. *Annual Review of Marine Science*, *11*(1), 131–158. https://doi.org/10.1146/annurev-marine-121916-063141

Santoro, A. E., Saito, M. A., Goepfert, T. J., Lamborg, C. H., Dupont, C. L., & DiTullio, G. R. (2017). Thaumarchaeal ecotype distributions across the equatorial Pacific Ocean and their potential roles in nitrification and sinking flux attenuation. *Limnology and Oceanography*, *62*(5), 1984–2003. https://doi.org/https://doi.org/10.1002/lno.10547

Sara, R. B., G, H. N., Christopher, R., Andrew, B., A, M. D., Joshua, L., L, G. A., Vikas, P., T, C. B., & J, S. S. (2020). Sensitive identification of bacterial DNA in clinical specimens by broad-range 16S rRNA gene enrichment. *Journal of Clinical Microbiology*, *58*(12), e01605-20. https://doi.org/10.1128/JCM.01605-20

Shafiee, R. T., Diver, P. J., Snow, J. T., Zhang, Q., & Rickaby, R. E. M. (2021). Marine ammonia-oxidising archaea and bacteria occupy distinct iron and copper niches. *ISME Communications*, *1*(1), 1. https://doi.org/10.1038/s43705-021-00001-7

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLOS ONE*, *11*(10), e0163962-. https://doi.org/10.1371/journal.pone.0163962

Shimomura, Y., Morimoto, S., Takada Hoshino, Y., Uchida, Y., Akiyama, H., & Hayatsu, M. (2012). Comparison among *amoA* primers suited for quantification and diversity analyses of ammonia-oxidizing bacteria in soil. *Microbes and Environments*, *27*(1), 94–98. https://doi.org/10.1264/jsme2.ME11230

Shiozaki, T., Ijichi, M., Isobe, K., Hashihama, F., Nakamura, K., Ehama, M., Hayashizaki, K., Takahashi, K., Hamasaki, K., & Furuya, K. (2016). Nitrification and its influence on biogeochemical cycles from the equatorial Pacific to the Arctic Ocean. *The ISME Journal*, *10*(9), 2184–2197. https://doi.org/10.1038/ismej.2016.18

Siljanen, H. M. P., Manoharan, L., Hilts, A. S., Bagnoud, A., Alves, R. J. E., Jones, C. M., Sousa, F., Hallin, S., Biasi, C., & Schleper, C. (2022). Targeted metagenomics using probe capture detects a larger diversity of nitrogen and methane cycling genes in complex microbial communities than traditional metagenomics. *BioRxiv*. https://doi.org/10.1101/2022.11.04.515048

Sintes, E., Bergauer, K., De Corte, D., Yokokawa, T., & Herndl, G. J. (2013). Archaeal *amo*A gene diversity points to distinct biogeography of ammonia-oxidizing *Crenarchaeota* in the ocean. *Environmental Microbiology*, *15*(5), 1647–1658. https://doi.org/10.1111/j.1462-2920.2012.02801.x

Spasov, E., Tsuji, J. M., Hug, L. A., Doxey, A. C., Sauder, L. A., Parker, W. J., & Neufeld, J. D. (2020). High functional diversity among Nitrospira populations that dominate rotating biological contactor microbial communities in a municipal wastewater treatment plant. *The ISME Journal*, *14*(7), 1857–1872. https://doi.org/10.1038/s41396-020-0650-2

Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, *35*(11), 1026–1028. https://doi.org/10.1038/nbt.3988

Sterngren, A. E., Hallin, S., & Bengtson, P. (2015). Archaeal ammonia oxidizers dominate in numbers, but Bacteria drive gross nitrification in N-amended grassland soil. *Frontiers in Microbiology*, *6*, 1350. https://doi.org/10.3389/fmicb.2015.01350

Takai, K. (2019). The nitrogen cycle: A large, fast, and mystifying cycle. *Microbes and Environments*, *34*(3), 223–225. https://doi.org/10.1264/jsme2.ME3403rh

Tavormina, P. L., Orphan, V. J., Kalyuzhnaya, M. G., Jetten, M. S. M., & Klotz, M. G. (2011). A novel family of functional operons encoding methane/ammonia monooxygenase-related proteins in gammaproteobacterial methanotrophs. *Environmental Microbiology Reports*, *3*(1), 91–100. https://doi.org/https://doi.org/10.1111/j.1758-2229.2010.00192.x

Tolar, B. B., King, G. M., & Hollibaugh, J. T. (2013). An analysis of Thaumarchaeota populations from the northern Gulf of Mexico. *Frontiers in Microbiology*, *4*, 72. https://doi.org/10.3389/fmicb.2013.00072

van Kessel, M. A. H. J., Speth, D. R., Albertsen, M., Nielsen, P. H., Op den Camp, H. J. M., Kartal, B., Jetten, M. S. M., & Lücker, S. (2015). Complete nitrification by a single microorganism. *Nature*, *528*(7583), 555–559. https://doi.org/10.1038/nature16459

Voss, M., Bange, H. W., Dippner, J. W., Middelburg, J. J., Montoya, J. P., & Ward, B. (2013). The marine nitrogen cycle: recent discoveries, uncertainties and the potential relevance of climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1621), 20130121. https://doi.org/10.1098/rstb.2013.0121

Wang, Y., Qin, W., Jiang, X., Ju, F., Mao, Y., Zhang, A., Stahl, D. A., & Zhang, T. (2021). Seasonal prevalence of ammonia-oxidizing archaea in a full-scale municipal wastewater treatment plant treating saline wastewater revealed by a 6-year time-series analysis. *Environmental Science & Technology*, *55*(4), 2662–2673. https://doi.org/10.1021/acs.est.0c07703

Wuchter, C., Abbas, B., Coolen, M. J. L., Herfort, L., van Bleijswijk, J., Timmers, P., Strous, M., Teira, E., Herndl, G. J., Middelburg, J. J., Schouten, S., & Sinninghe Damsté, J. S. (2006). Archaeal nitrification in the ocean. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(33), 12317–12322. https://doi.org/10.1073/pnas.0600756103

Yakimov, M. M., Cono, V. la, Smedile, F., DeLuca, T. H., Juárez, S., Ciordia, S., Fernández, M., Albar, J. P., Ferrer, M., Golyshin, P. N., & Giuliano, L. (2011). Contribution of crenarchaeal autotrophic ammonia oxidizers to the dark primary production in Tyrrhenian deep waters (Central Mediterranean Sea). *The ISME Journal*, *5*(6), 945–961. https://doi.org/10.1038/ismej.2010.197

Zhang, Y., Tian, Z., Liu, M., Shi, Z. J., Hale, L., Zhou, J., & Yang, M. (2015). High concentrations of the antibiotic spiramycin in wastewater lead to high abundance of ammonia-oxidizing archaea in nitrifying populations. *Environmental Science & Technology*, *49*(15), 9124–9132. https://doi.org/10.1021/acs.est.5b01293

Zou, D., Wan, R., Han, L., Xu, M. N., Liu, Y., Liu, H., Kao, S.-J., & Li, M. (2020). Genomic characteristics of a novel species of ammonia-oxidizing archaea from the Jiulong River estuary. *Applied and Environmental Microbiology*, *86*(18), e00736-20. https://doi.org/10.1128/AEM.00736-20

Figures

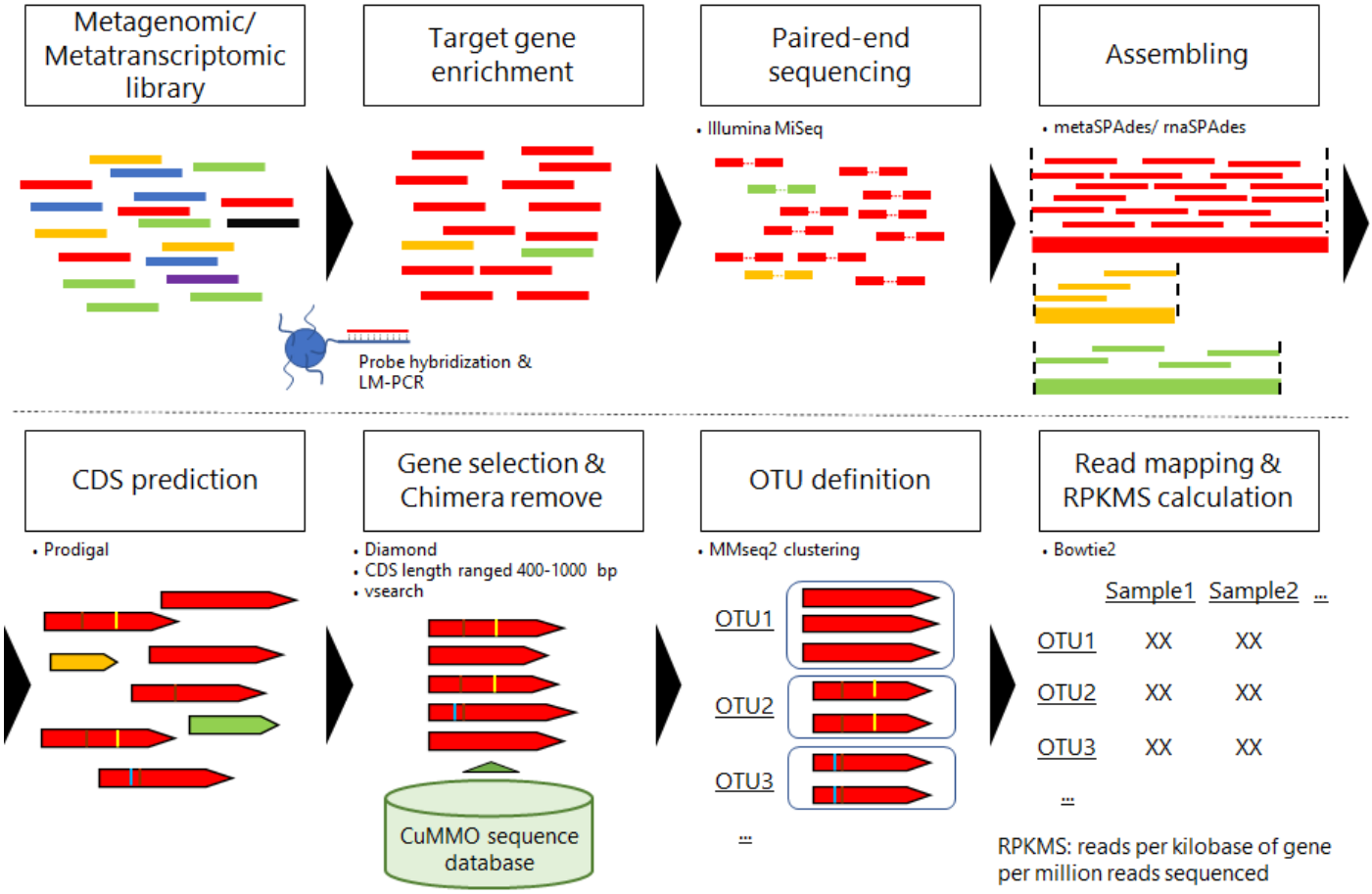

Figure 1. A schematic overview of probe capture enrichment sequencing analysis.

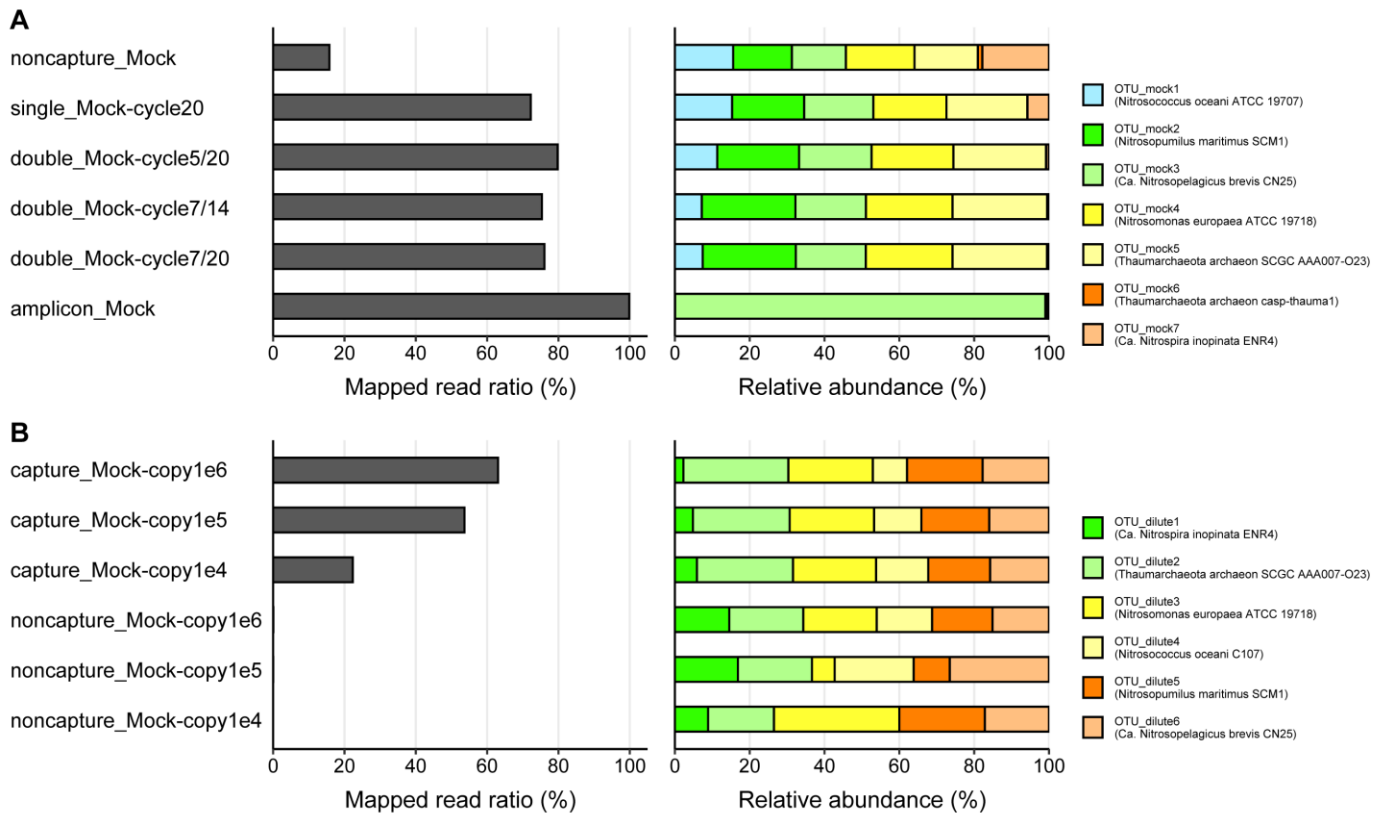

Figure 2. The *amoA* probe capture enrichment sequencing of metagenomic mock samples.

The mock sample comprised six plasmids containing artificially synthesized *amoA* genes from three AOA and three AOB lineages at the same molar concentration. The left bar represents the ratio of mapped HQ reads onto the OTU sequences. The right percentage-stacked bar plot represents the relative abundance of each OTU based on the RPKMS. The assigned lineages for each OTU are shown in the legends. (**A**) Probe capture enrichment was conducted using different repeat times of hybridization repetitions and LM-PCR cycle settings. Amplicon sequencing was performed using Francis primers. OTU were defined using pooled reads from all five capture samples, that is, those from amplicon sequencing were not used. (**B**) Diluted samples were prepared using the original mock sample rarefied with control DNAs (cloning vector pTAC-2) and adjusted to 10^6, 10^5, and 10^4 *amoA* gene copies per ng DNA. Data generated with/without probe capture enrichment are represented as ‘capture’ and ‘noncapture’ in the sample name, respectively. OTU were defined using the pooled reads from all six samples.

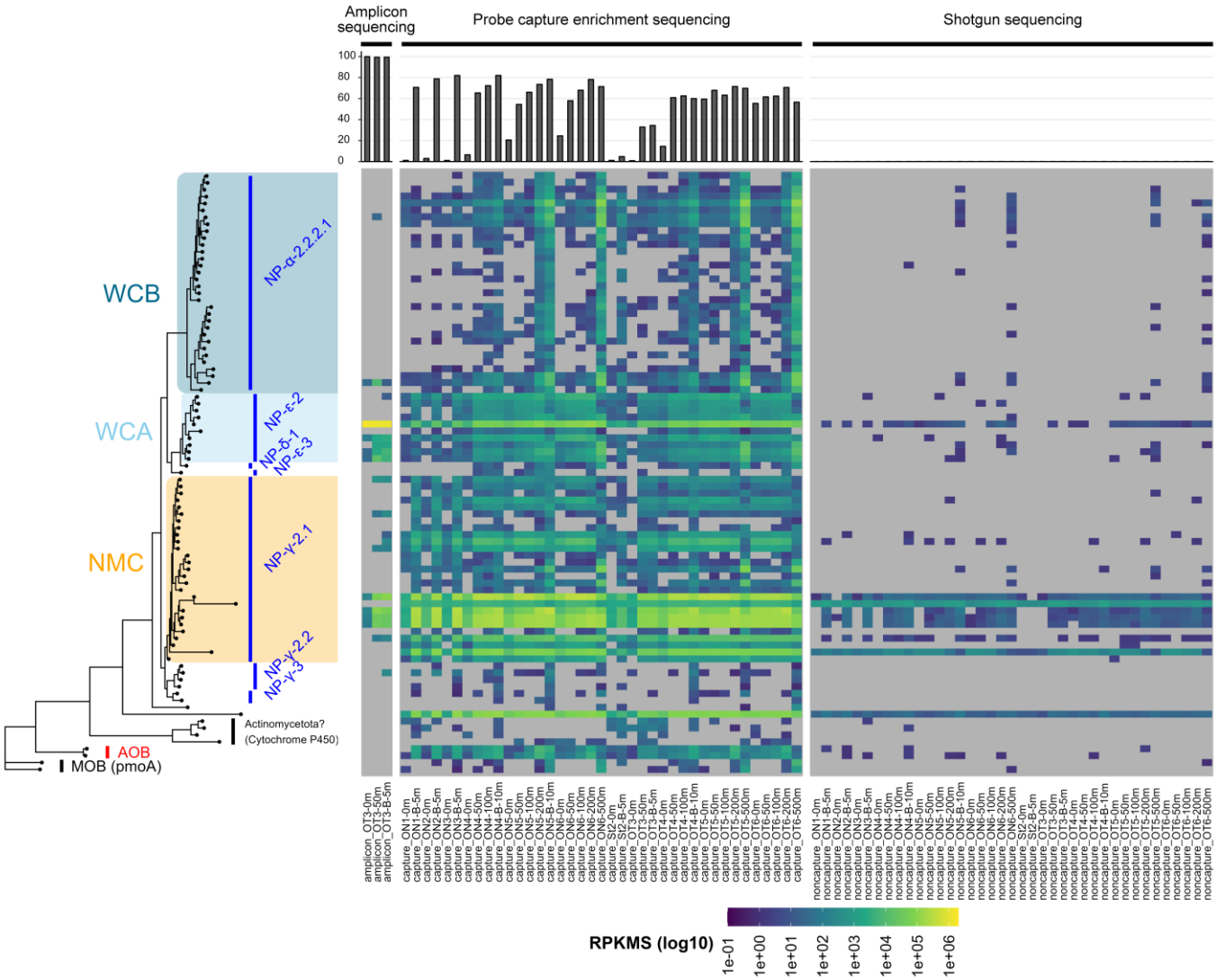

Figure 3. The *amoA* amplicon, *amoA* probe capture enrichment, and shotgun sequencings of metatranscriptomic seawater samples.

OTUs were predicted using pooled reads from all 39 capture and 39 non-capture samples, that is, excluding those from the three amplicon sequencing samples. For amplicon sequencing, *amoA* was amplified from transcriptomic cDNA using Francis primers. The left tree represents a phylogenetic tree of the OTUs obtained using the maximum-likelihood method. Nodes were grouped into clades defined by Alves et al. for AOA (blue bars and text), the domain level for AOB (red), MOB, and possible Actinomycetota with genes other than *amoA* (black). The background color indicates the three major AOA ecotypes in the ocean. The top bar plot represents the ratios of mapped reads to OTUs by the total HQ merged/HQ HiFi reads. Heat maps represent the RPKMS abundance of each OTU in each seawater sample. Heat maps were separated into amplicon, capture, and non-capture samples. Gray cells indicate OTUs to which no reads were mapped.

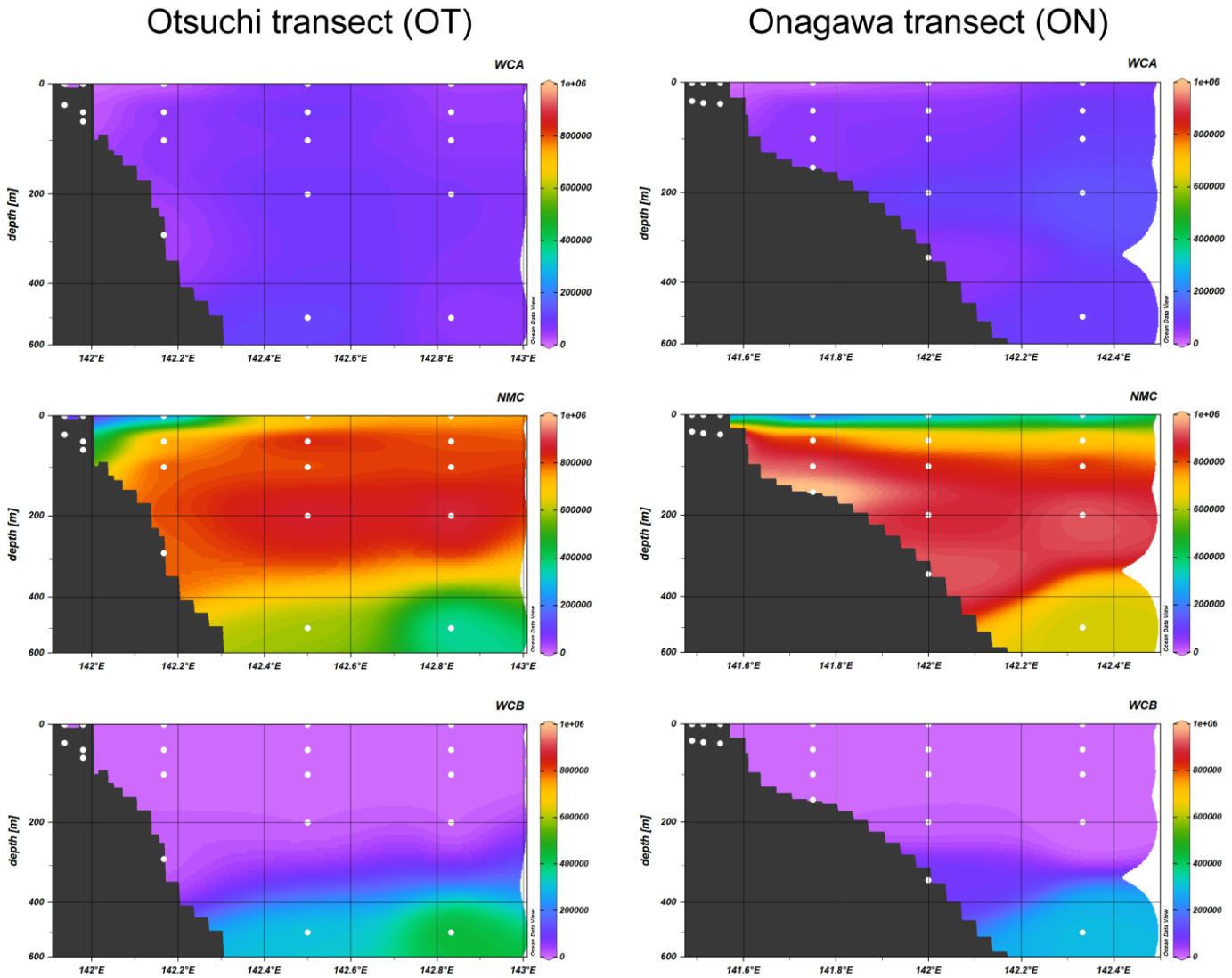

Figure 4. RPKMF abundances of the three major ecotypes in the ocean along the east-west section of Onagawa (ON) and Otsuchi (OT) transects.

White dots represent the sampling depth at each station. WCA, NMC, and WCB are represented by OTUs assigned to NP-ε-2, NP-γ-2.1, and NP-α-2.2.2.1 (see Figure 3).