



Composition and distribution of diazotrophs in the Baltic Sea

Ellen R. Salamon Slater^a, Kendra A. Turk-Kubo^b, Søren Hallstrøm^a, Katharina Kesý^c,
Peeter Laas^{d,e}, Jonathan Magasin^b, Jonathan P. Zehr^b, Matthias Labrenz^c, Lasse Riemann^{a,*}

^a Marine Biological Section, Department of Biology, University of Copenhagen, Denmark

^b Ocean Sciences Department, University of California Santa Cruz, Santa Cruz, CA, USA

^c Biological Oceanography, Leibniz Institute for Baltic Sea Research Warnemünde (IOW), Rostock, Germany

^d Institute of Technology, Faculty of Science and Technology, University of Tartu, 50411, Tartu, Estonia

^e Department of Marine Systems, School of Science, Tallinn University of Technology, 12618, Tallinn, Estonia

ARTICLE INFO

Keywords:

Baltic Sea

Non-cyanobacterial diazotrophs

nifH

Nitrogen fixation

Cyanobacteria

ABSTRACT

Nitrogen (N₂) fixation rates in the brackish Baltic Sea are among the highest per unit of area in the world. However, beyond the filamentous heterocyst-forming cyanobacteria, knowledge about the composition and distribution of N₂-fixing microbes (diazotrophs) is limited. To address this, we investigated nitrogenase gene (*nifH*) composition and expression at coastal (<10 km offshore) and offshore (>10 km offshore) stations, at surface (avg. 1.8 m) and at depth (avg. 24 m) and in free-living (0.2–3.0 μm) and particle-associated size fractions (>3 μm). Surprisingly, *nifH* genes affiliated with *Pseudanabaena* and non-cyanobacterial diazotrophs (NCDs) dominated the composition whereas filamentous heterocyst-forming cyanobacteria accounted for almost 80% of the *nifH* transcripts. Salinity had a minor influence on the composition, but *Aphanizomenon* and *Nodularia* showed increased relative *nifH* gene expression at low and higher salinity, respectively. *Pseudanabaena* only accounted for up to 5% of the *nifH* transcripts and *nifH* gene expression by *Candidatus Atelocyanobacterium thalassa* (sublineage UCYN-A2) was mainly observed in the most saline western part of the Baltic. The only notable expression by NCDs (up to 15% of *nifH* transcripts at a given station) coincided with an upwelling event at the southern coast and was largely accounted for by a *Pseudomonas*-like *nifH* phylotype, recurrently found in the Baltic Sea. NCD relative abundances were dominant in coastal stations, presumably driven by sediment resuspension as evidenced by higher turbidity and DOC levels and the recovery of sediment diazotrophs in the pelagic zone. This study reveals the heterogeneity of the composition and activity of diazotrophs in the Baltic Sea, and underscores the need for future N₂ fixation studies that include coastal and offshore Baltic waters.

1. Introduction

The Baltic Sea is among the largest brackish bodies of water on Earth, covering 370,000 km². It spans a salinity gradient from 1 to 2 in the Bothnian Bay in the north to around 9 in the Arkona basin in the south, and is semi-enclosed with a mean depth of 56 m (Kullenberg and Jacobsen, 1981; Stal et al., 2003). In the Baltic Sea, dinitrogen (N₂)-fixing (or “diazotrophic”) cyanobacteria are important for N cycling (Larsson et al., 2001). N₂ fixation rates in late summer are so high that relative to its size the central Baltic Sea Proper is the body of water with the most N₂ fixation per unit of area in the world (Niemistö et al., 1989; Wasmund et al., 2001, 2005; Klawonn et al., 2016). N₂ fixation can sustain up to 90% of the net summer community production with an estimated fixation of 180–430 Gg N yr⁻¹ (Larsson et al., 2001).

The filamentous heterocyst-forming cyanobacteria *Aphanizomenon* sp., *Dolichospermum* sp. (previously *Anabaena* sp; Wacklin et al., 2009) and *Nodularia spumigena* are considered the main diazotrophs in the Baltic Sea (Stal et al., 2003), and are monitored by microscopy (Kownacka et al., 2020; STATECON, 2021). Molecular analysis of the *nifH* gene, a common marker for N₂-fixing organisms (Zehr and McReynolds, 1989; Raymond et al., 2004), is a widely used approach to characterize the composition of diazotrophic communities (e.g., Zehr et al., 1998; Farnelid et al., 2011) not readily identified using microscopy. Few studies have applied *nifH* amplicon sequencing in the Baltic Sea, and they have limited geographical coverage. The available analyses of *nifH* gene composition mainly cover the central Baltic Sea, representing a few discrete sites mainly offshore (Bostrom et al., 2007a, 2007b; Farnelid et al., 2009, 2013, 2014; Bentzon-Tilia et al., 2014; Reeder et al., 2022).

* Corresponding author. Marine Biological Section, Department of Biology, University of Copenhagen, Strandpromenaden 5, 3000, Helsingør, Denmark.
E-mail address: riemann@bio.ku.dk (L. Riemann).

<https://doi.org/10.1016/j.ecss.2023.108527>

Received 12 July 2023; Received in revised form 29 September 2023; Accepted 13 October 2023

Available online 16 October 2023

0272-7714/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Nonetheless, this approach revealed the presence of the small symbiotic unicellular cyanobacteria UCYN-A (*Candidatus Atelocyanobacterium thalassa*; Bentzon-Tilia et al., 2015b; Reeder et al., 2022). This organism plays an important role in the oceanic nitrogen cycle (Martínez-Pérez et al., 2016; Zehr and Capone, 2020), but its distribution in the Baltic Sea is unknown. Additionally, *nifH* gene sequencing at specific sampling sites in the Baltic Sea revealed diverse assemblages of non-cyanobacterial diazotrophs (NCDs; Farnelid et al., 2009; 2013). Cultivation of NCDs obtained from the Baltic Sea water column (Boström et al., 2007a; Farnelid et al., 2014; Bentzon-Tilia et al., 2014, 2015a) and associated N_2 fixation rate measurements (Bentzon-Tilia et al., 2015a) indicate that they may contribute to N_2 fixation *in situ*. Still, these insights are based on few localized studies. Hence, present knowledge about the diazotrophs in the Baltic Sea is mainly based on microscopy. A molecular analysis of the composition and biogeography of diazotrophs could, therefore, provide important new information about the present and active diazotrophs and their environmental regulation in the Baltic Sea.

Since bacterial and cyanobacterial biomass (Legrand et al., 2015) and N_2 fixation rates (Klawonn et al., 2016) differ between coastal and offshore stations, and coastal and offshore stations are kept separate in monitoring (Kownacka et al., 2020), it is essential to also include coastal environments when analyzing the overall composition of diazotrophs in the Baltic Sea. Shallow coastal and estuarine areas generally have a high suspended matter load and are rich in particulate organic matter (Simon et al., 2002), to a large extent originating from sediment resuspension (Turner and Millward, 2002). Particles have been proposed as loci for N_2 fixation by NCDs as they may represent high-substrate and low O_2 conditions (Paerl et al., 1988; Bombar et al., 2016; Chakraborty et al., 2021), which are considered essential for N_2 fixation by NCDs in the oxic water column (Riemann et al., 2022). Interestingly, a major part of the N_2 fixation in a Danish estuary influenced by the Baltic Sea was attributed to NCDs, and resuspension of sediment bacteria was considered a contributing factor (Bentzon-Tilia et al., 2015b). A later experiment in the estuary indicated that pelagic N_2 fixation could be enhanced almost 100-fold by sediment resuspension (Pedersen et al., 2018). Indeed, resuspension of sediments as a source of diazotrophs in the water column has been proposed for diverse coastal or estuarine regions; e.g. Spencer Gulf (Australia; Messer et al., 2021), Chesapeake Bay (USA; Short et al., 2004; Moisaner et al., 2007), Monterey Bay (USA; Cabello et al., 2020) and Narragansett Bay (USA; Hallström et al., 2022). Hence, it is likely that the combination of higher productivity (Sigman and Hain, 2012) and concentration of particles might facilitate particle-associated lifestyles in coastal areas. Thus, the composition of diazotrophs, and the relative importance of different size-fractions, is expected to differ between coastal and offshore regions.

In the present study, we aimed to obtain the first comprehensive overview of present and active diazotrophs in the Baltic Sea along the salinity gradient from south to north, including both coastal and offshore areas, as well as surface and deep samples. We analyzed the diazotroph community composition using *nifH* gene amplicons, characterized which taxa were actively transcribing the *nifH* gene, as a proxy for N_2 fixation activity, and compared this to a range of environmental parameters. To capture diazotrophs with potentially different diel patterns in *nifH* expression, sampling encompassed different time points, but also stations sampled at the same time of day for comparison. Finally, all samples were size fractionated (0.2–3.0 μm and $\geq 3.0 \mu\text{m}$) to gain insight into potential particle associations.

2. Materials and methods

Samples were collected from ‘surface’ (average 1.8 m) and ‘deep’ (average 24 m, see Table S1 for max bottom depths) waters from 18 stations as part of the POS488 cruise on the R/V Poseidon from August 17th to September 4th, 2015. Stations were between 2 and 118 km from shore and maximum bottom depths were 15.5–365.8 m (Tables S1 and

S2). The stations were chosen to sample the salinity gradient from the Arkona Basin in the south to the Bothnian Sea in the north and were operationally divided into coastal (<10 km offshore, European Environment Agency, 2006) and offshore (>10 km offshore) stations (Fig. 1). The two categories differed in their distance from shore (Wilcoxon rank sum test with continuity correction; Wilcoxon CC), $p < 0.001$) and the coastal stations had shallower maximum bottom depths ($p = 0.0057$, Wilcoxon rank sum exact; Wilcoxon ET).

2.1. Environmental parameters

Temperature, turbidity, salinity and O_2 were measured using a Sea-Bird SBE 9 CTD. Dissolved organic carbon (DOC) and chlorophyll *a* (Chl *a*) samples were filtered on board, frozen and analyzed according to standard protocols at Leibniz Institute for Baltic Sea Research Warnemünde (IOC, 1994; Grasshoff et al., 1999; Lysiak-Pastuszek and Andersens, 2004). Inorganic nutrients (PO_4^{3-} , NO_3^- , NO_2^- , NH_4^+ and SiO_2) were measured by standard colorimetric methods (Grasshoff et al., 1999).

2.2. Bacterial enumeration

Samples for flow cytometry were fixed in triplicate with para-formaldehyde and glutaraldehyde (final concentrations 1% and 0.05%, respectively; pH = 7.4) and incubated at 4 °C for 30 min. Thereafter, each triplicate was stained with SYBR™ Green I Nucleic Acid Stain (Thermo Fisher Scientific, USA) at room temperature for 15 min following Marie et al. (1999). Cell counts were obtained via on-board flow cytometry (BD Accuri™ C6 cytometer, BD Biosciences, USA) at ‘slow speed’ mode (14 $\mu\text{L min}^{-1}$, 10- μm core). Low nucleic acid content (LNA) and high nucleic acid content (HNA) cells were discriminated using BD Accuri C6 software and final counts were averaged from triplicate subsamples.

2.3. Sampling and extraction of DNA and RNA

Size-fractionated DNA and RNA samples were obtained from surface and deep samples from all 18 stations. DNA samples were collected using Niskin bottles while RNA samples were collected and immediately fixed using an autonomous *in situ* fixation sampler (Charvet et al., 2019). Samples of 500–1000 ml were filtered serially through 3.0 μm (Dura-pore, Merck Millipore, Germany) and 0.2 μm (Isopore, Merck Millipore, Germany) filters to separate free-living (FL; 0.2–3.0 μm) and particle-associated bacteria (PA; >3.0 μm); i.e. associated with phytoplankton, filamentous cyanobacteria or other biotic and abiotic particles. The filters were immediately flash frozen in liquid nitrogen and stored at –80 °C.

DNA was extracted from half-filters from each sample using the DNeasy® PowerSoil® Kit (Qiagen) with the following modifications: Filters were freeze-thawed for three cycles, and then cut aseptically in small pieces into bead beater tubes prior to extraction. DNA was quantified using Quant-iT™ PicoGreen™ dsDNA Assay Kit (Invitrogen, Oregon, USA). RNA was extracted using the RNeasy Mini kit (Qiagen) with modifications including an on-column DNA digestion (RNase-Free DNase Set, Qiagen), initial bead beating (200 μm Low Binding Zirconium Beads, OPS Diagnostics, LLC), an additional DNA digestion (Invitrogen, TM TURBO DNA-free™ Kit), and additional clean-up using the RNA clean and Concentrator™ - 5 (Zymo Research, USA). RiboLock RNase Inhibitor (Thermo Scientific, Vilnius, Lithuania) was added to the final extract.

RNA was quantified (Quant-iT™ RiboGreen™ RNA Assay kit, Invitrogen) and stored at –80 °C or immediately reverse transcribed to complementary DNA (cDNA) using SuperScript™ IV (Invitrogen) and the *nifH3* primer (Zani et al., 2000) according to the manufacturer’s protocol. Reverse transcriptase-free controls were run for each RNA sample by substituting the reverse transcriptase with PCR grade water.

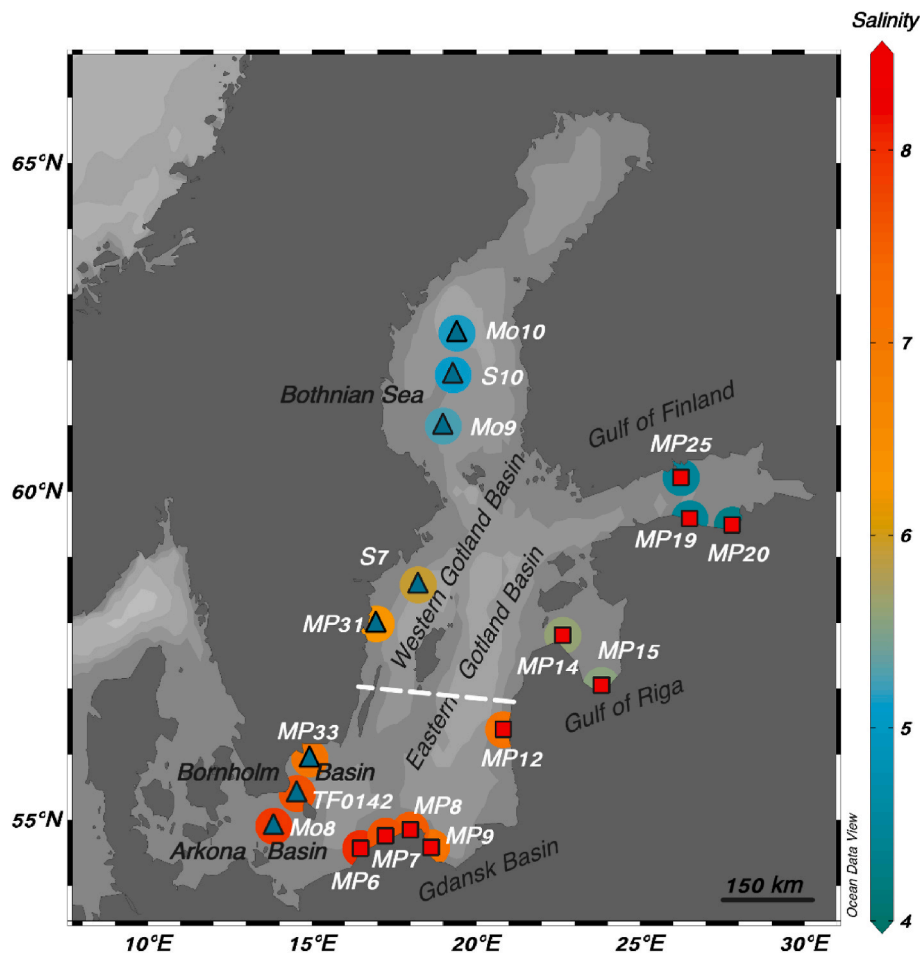


Fig. 1. Map of the Baltic Sea depicting surface salinity at the 18 sampled stations. The white, dashed line represents the division of salinities >7 and <7 . Red squares and blue triangles define coastal (<10 km from shore) and offshore (>10 km from shore) stations, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

The cDNA product was stored at -20°C .

2.4. PCR amplification of nitrogenase (*nifH*) genes

Amplification of *nifH* was done with a nested PCR approach (Zehr et al., 1998). In the first reaction, *nifH3* and *nifH4* primers were added (Zani et al., 2000) while *nifH1* and *nifH2* primers (with added Illumina adapter overhangs) were used for the second reaction (Zehr and McReynolds, 1989). Each 25 μl reaction contained MyTaqTM Red DNA Polymerase (Bio-line, Meridian Bioscience, United Kingdom), 0.2 or 0.4 μM primers, 10 mM MgCl_2 and 0.8 $\mu\text{g } \mu\text{l}^{-1}$ or 1.2 $\mu\text{g } \mu\text{l}^{-1}$ Bovine Serum Albumin with either 1–10 ng DNA template or 1–2.5 μl cDNA template. The thermocycling conditions for both reactions were 2 min at 94°C , followed by 30 cycles of: 1 min at 94°C , 1 min at 54°C and 1 min at 72°C ending with 7 min at 72°C . Triplicate PCR products were pooled and purified using the GeneClean[®] Turbo Kit (MP Biomedicals, LLC). A negative PCR control with PCR grade water instead of template was included in all runs and showed no amplification. The reverse transcriptase-free controls were checked for complete DNA digestion by PCR as described above and showed no amplification. The PCR products were indexed using custom designed Illumina indexes with an 8 cycle PCR (20 s at 94°C , 30 s at 62°C and 30 s at 72°C). Finally, samples were purified using magnetic beads (Agencourt, Beckman Coulter, Ca, USA), pooled in equimolar ratios and sequenced using an Illumina MiSeq Platform (GeoGenetics Sequencing Core, Denmark).

2.5. Processing of amplicon sequences

Of the 144 total samples, 122 (55 DNA and 67 RNA) were successfully amplified and sequenced for *nifH*. Most of the samples that failed to amplify, had very low concentrations of DNA/RNA, possibly due to low extraction efficiency. Reads were processed using DADA2 (version 1.20.0; Callahan et al., 2016) and R (version 4.1.3; R Core Team, 2021). Quality profiles for all paired FASTQs were prepared using FastQC (version 0.11.9; Andrews, 2010). Then primers were trimmed using cutadapt (version 3.4; Martin, 2011) with tolerance for 1 mismatch ($-m$ 1) and no indels ($-no-indels$). Reads from both direction were required to have primers or the pair was rejected ($-discard-untrimmed$). Each combination of sample type (DNA or RNA) and size fraction was processed separately with DADA2 and R using mainly standard steps: filter and trim, denoise, and merge the reads into amplicon sequence variants (ASVs). An exception was that the sequencing error models used by DADA2 for the function *filterAndTrim()* were created in advance (with the function *learnErrors()*) using only the paired reads that appeared to represent *nifH* amplicons to remove off-target PCR artefacts. These were identified by predicting open reading frames (ORFs) on the forward reads with FragGeneScan (version 1.31; Rho et al., 2010) and then scanning the ORFs for the *NifH*/*frxC* family (Fer4 *NifH*; PF00142) using *hmmsearch* in HMMER software (version 3.3.2; Finn et al., 2011). Only the paired reads with ORFs that had ≥ 80 residues and met the trusted cut-offs encoded in PF00142 were used to create the error models that were passed to the *filterAndTrim()* function. Additional parameters to the *filterAndTrim()* function specified trim lengths of 220 nt and 180 nt

for the forward and reverse reads, respectively, selected based on the FastQC quality reports, and maximum tolerated base call errors (maxEE) of 4 nt and 6 nt, respectively. The parameters of the DADA2 function *mergePairs()* required sequences to overlap by > 12 nt with at most 1 mismatch. Chimera ASVs were detected and removed first using the DADA2 function *isBimeraDenovo()* with *minFoldParentOverAbundance* = 3.5, and then using UCHIME2 (Edgar, 2016) as implemented in *vsearch* (version 2.18.0; Rognes et al., 2016) through option *-uchime3_denovo*. Non-*nifH* ASVs were identified and removed using a similar approach to that used for creating the error models. Finally, ASVs that did not meet the criteria of having >10 reads in ≥ 3 samples or >40 reads in one sample were removed.

The total number of reads after quality control was 15,206,287, yielding 1139–244,590 reads per sample (mean; 129,968 reads). For the downstream analysis, 2656 ASV and 117 samples were analyzed due to removal of samples with less than 1000 reads. Sequence data were not rarefied before analysis as this may lead to loss of information (McMurdie and Holmes, 2014) and most samples exceeded the depth considered acceptable for comparing microbiome composition between samples (e.g. in gut microbiomes; 10,000 reads per sample; Falony et al., 2016). Of the 2656 total ASVs, 866 were represented in the RNA and 1843 in the DNA. BLASTX against a curated database containing phylogenetically characterized *nifH* sequences (genome879; jzehlrlab.com/nifh) was used to categorize the *nifH* sequences into NCD and cyanobacterial *nifH*. The ASV data was used to create relative abundances for the respective ASVs in each sample. Relative abundances, as used in the present work, are often used in marine microbiology (e.g. Herlemann et al., 2011; Farnelid et al., 2019), but should be interpreted cautiously as they are compositional in nature and not strictly quantitative.

2.6. Clustering of the main cyanobacterial ASVs

To gain insight into the composition of cyanobacterial diazotrophs, the 36 most dominant cyanobacterial ASVs, accounting for 90% of the cyanobacterial relative abundances, were translated into 15 unique amino acid sequences and clustered into seven groups using CD-hit at 97% similarity (Li and Godzik, 2006; Huang et al., 2010; Fu et al., 2012). Reference sequences were selected among known Baltic cyanobacteria (Stal et al., 2003) and nearest relatives using BLASTP (National Center for Biotechnology Information, NCBI). The RefSeq non-redundant proteins with the best match were chosen, then further culled to one reference sequence per cluster. Two reference sequences from isolated strains were chosen as they were specific to the Baltic Sea: *Aphanizomenon* KAC15 and *Nodularia spumigena* AV1 (Janson and Granéli, 2002; Vintila et al., 2011). After ASVs were assigned a cluster, the read count for each of the seven clusters were summed for each sample. The groups were named after their reference sequence. The remaining 10% of cyanobacterial relative abundances were represented by 1052 ASVs that were grouped and labeled 'Other cyanobacteria'.

2.7. Statistics

Statistics were performed with R-4.1.1 in R studio (R Core Team, 2021; RStudio Team, 2021). The reads for each ASV were transformed into relative abundances per sample with the "microbiome" package (Lahti and Shetty, 2019). The function "*knrImputation*" from the "DMwR2" package (Torgo, 2016) was used to impute a missing Chl *a* value (a lost filter).

To check for significant differences in environmental parameters between coastal and offshore stations, Wilcoxon rank tests were applied using the "stats" package. Wilcoxon signed-rank tests were furthermore applied to compare size-fractions. The same package was used for performing T-tests on log-transformed bacterial abundances to check for significant differences between the averages for coastal and offshore stations.

Linear Mixed-Effects Models (LME) and Generalized Linear Mixed-Effects Models (GLMM) were made with the "lme4" package (Bates et al., 2015) to evaluate how much effect certain environmental parameters had on the variation of the relative abundances of cyanobacterial CD-hit clusters and NCD classes. The kind of transformation needed to make the data follow a normal distribution was based on Box Cox transformation from the package "MASS" (Venables and Ripley, 2002). To achieve normal distribution both log and square root transformations were applied, according to the Box Cox lambda value. *Deltaproteobacteria* and *Pseudanabaena*-like *nifH* gene abundance were log transformed and NCD *nifH* genes were square root transformed. Each model was validated with residual plots and Quantile-Quantile plots. *Dolichospermum*-like, *Aphanizomenon*-like and *Nodularia* AV1-like *nifH* were transformed into presence/absence since they did not meet the validation criteria. GLMM was used on presence/absence data as an alternative to LME. Fixed effects included: Size-fraction $\times$ depth, off-shore/coastal and salinity. 'Station' was set as a Random effect. The influence of salinity on the variance of total *nifH* composition was examined by redundancy analysis (RDA) on centered log ratio transformed *nifH* classes for surface and deep samples separately, with salinity and temperature as the constraint variables, using "phyloseq", and "microViz" (McMurdie and Holmes, 2013; Wickham, 2016; Barnett et al., 2021).

To determine the phylogeny of the 36 dominant cyanobacterial ASVs, amino acid sequences were aligned with Mafft v7.450 (Katoh et al., 2002; Katoh and Standley, 2013) in Geneious. RaxML 8.2.11 was used to build a Maximum-likelihood tree (GTR GAMMA) with 1000 bootstraps (Stamatakis, 2014). To build a tree representing the most relative abundant ASVs, amino acid sequences for 36 cyanobacterial ASVs, and 32 NCD ASVs were aligned with Muscle 3.8.425 (Edgar, 2021) in Geneious. RaxML 8.2.11 was used to build a Maximum-likelihood tree (GTR GAMMA) with 1000 bootstraps (Stamatakis, 2014). UCYN-A ASVs were aligned with UCYN-A nucleotide reference sequences from Farnelid et al. (2016) and Turk-Kubo et al. (2017) and processed as described above to create a Maximum-likelihood tree and determine affiliation with known UCYN-A sublineages. Graphs and tables were made with iTOL v6 (Letunic and Bork, 2021), Ocean Data View (Schlitzer, 2021) and the R packages "ggplot2", "mapplots", "knitr" and "kableExtra" (Wickham, 2016; Geritsen, 2018; Xie, 2021; Zhu, 2021).

3. Results

3.1. Environmental conditions

Environmental data for all stations are summarized in Tables S1 and S2. Salinity ranged from 4.3 to 8.6 with samples north of station MP12 being <7 (except S7 at 46.1 m) while samples south of MP12 were >7 (Fig. 1). Temperatures at the surface and deep sample points ranged 11.7–20.7 °C (Fig. S1a) and 4.1–17.5 °C, respectively. All stations were oxic with O₂ concentrations between 174.9 and 375.5 $\mu\text{mol L}^{-1}$. Chl *a* varied from 1.6 to 9.3 $\mu\text{g L}^{-1}$ at the surface and 0.7–1.5 $\mu\text{g L}^{-1}$ for the deeper samples (Fig. S1b). Turbidity at the surface was 0.19–1.5 Nephelometric Turbidity Units (NTU) and 0.1–1.8 NTU at depth (Fig. S1c). DOC was 320.6–514.3 $\mu\text{mol L}^{-1}$ at the surface (data from deep samples not available). Total dissolved inorganic nitrogen (DIN) was low at the surface (<0.5 $\mu\text{mol L}^{-1}$) and higher at depth (0.1–7.5 $\mu\text{mol L}^{-1}$). Dissolved inorganic phosphorus (DIP) was 0–0.8 $\mu\text{mol L}^{-1}$ at the surface and 0.1–1.2 $\mu\text{mol L}^{-1}$ at depth. At both depths at nearly all stations (except for stations Mo10 and S10 in the Bothnian Sea), the DIN/DIP ratios were less than the Redfield ratio of 16:1 (Fig. S1d; Redfield et al., 1963). The lowest surface temperatures were at the most southern stations (MP6 – 8; Fig. S1a), where also surface DIP (0.4, 0.6 and 0.6 $\mu\text{mol L}^{-1}$) was higher than average, indicating a local upwelling event.

Salinity, temperature, Chl *a*, DIP and DIN/DIP ratios did not differ

between coastal and offshore stations (Wilcoxon); however, deep O₂ concentrations were lower at coastal stations ($p < 0.001$; Wilcoxon CC). Furthermore, coastal samples had higher turbidity, both at the surface and at depth ($p = 0.003$ and $p = 0.001$; Wilcoxon ET). Additionally, DOC was highest in the coastal surface waters ($p = 0.005$; Wilcoxon CC) and DIN was highest in the coastal deep ($p = 0.009$; Wilcoxon CC).

3.2. Bacterial abundances

Bacterial abundance ranged 1.45×10^6 – 5.6×10^6 cells ml⁻¹ and 3.8×10^5 – 3.7×10^6 cells ml⁻¹, for surface and deep samples, respectively (Tables S1 and S2). When excluding stations MP6–8 that showed signs of upwelling (see above), bacterial abundance in surface waters was higher for the coastal stations than offshore ($p < 0.001$; T-test). The upwelling stations (MP6–8) showed the lowest abundances at the surface and the lowest ratio between HNA and LNA of all stations (including station S7 at 46 m).

3.3. Overall composition of present and active cyanobacterial diazotrophs

Cyanobacteria accounted overall for 57% and 99% of the *nifH* genes and transcripts, respectively. Of the overall cyanobacterial ASVs, a total of 36 ASVs accounted for 90% of the cyanobacterial *nifH* sequences from DNA (13 ASVs) and RNA (32 ASVs). These ASVs clustered mainly in 7 groups containing cyanobacteria previously found in the Baltic Sea (Figs. S2 and S3; Stal et al., 2003; Sivonen et al., 2007; Reeder et al., 2022). The remaining 10% of the cyanobacteria reads (1052 ASVs) were labeled ‘Other cyanobacteria’ and consisted of 12 different taxa: *Nostoc azollae*, *Nostoc* sp., *Oscillatoria* sp., cyanobacterium UCYN A, *Synechococcus* sp., *Nodularia spumigena*, cyanobacterium endosymbiont *Rhopalodia gibba*, cyanobacterium endosymbiont *Epithemia turgida*, *Anabaena variabilis*, *Cyanothece* sp., *Lyngbya* sp., and *Cylindrospermopsis raciborskii*. More than half of these ASVs were associated with *Nostoc azollae*.

Genes affiliated with *Pseudanabaena* dominated the cyanobacterial *nifH* sequences, and were more prevalent in the FL size-fraction (surface: 56.7% and deep: 37.6%) than in the PA size-fraction (surface: 36.6% and deep: 24.4%; Fig. 2a; $p = 0.011$; Wilcoxon CC). The filamentous heterocyst-forming cyanobacteria *Nodularia* AV1-like *nifH* and *Aphanizomenon*-like *nifH* showed highest relative abundances in the PA size-fractions (Fig. 2a; $p = 0.005$ and $p = 0.003$, respectively; Wilcoxon CC). Some of the dominant *Aphanizomenon*-like, *Pseudanabaena*-like and *Nodularia* AV1-like ASVs (ASV 1, 5 and 3, respectively) were identical at nucleotide level to *nifH* sequences previously found in the Baltic Sea (Farnelid et al., 2009, 2014). *nifH* genes from the unicellular

cyanobacterial symbiont UCYN-A (*Candidatus Atelocyanobacterium thalassa*) was only detected at station MP25 (<0.1% relative abundance) but showed expression at stations Mo8 and MP9 (see below). All 15 UCYN-A ASVs were affiliated with the UCYN-A2 sublineage (Fig. S4). The most abundant of the ASVs was identical to a UCYN-A2 *nifH* gene from the Californian coast (KF806612.1).

Aphanizomenon-like, *Nodularia* AV1-like and *Dolichospermum*-like *nifH*, accounted for >78% of expression in all sample categories, with *Aphanizomenon* and *Nodularia* dominating (Fig. 2b). These were identical to *nifH* transcripts from a previous Baltic Sea study (Boström et al., 2007b). *Pseudanabaena*-like *nifH* accounted for only ca. 5% of the transcripts in the FL size-fractions, and <0.1% in the PA size-fractions (Fig. 2b). UCYN-A2 contributed overall with only ca. 1% of the expression (Fig. 2b). UCYN-A2 showed high expression at station Mo8, where it accounted for 19% and 25% of the transcripts in the surface and deep FL size-fraction, respectively. UCYN-A2 was also found in the PA-size fraction, but to a smaller extent (<1%). In addition, at station MP9 UCYN-A2 was responsible for <0.1% of total *nifH* transcripts.

3.4. Overall composition of present and active NCDs

NCDs accounted for 43% and 1% of the total *nifH* genes and transcripts, respectively. A total of 412 ASVs accounted for 90% of the NCD *nifH* sequences, and 32 dominant ASVs accounted for 50% of the NCD *nifH* sequences (Fig. S3). The total NCD ASVs were dominated by the classes Deltaproteobacteria (20.6% of total relative abundance), Gammaproteobacteria (11.2%) and Verrucomicrobia (3.1%), whereas of the remaining classes only Opitutae, Chlorobia and Betaproteobacteria accounted for >1% (Fig. 3). Whereas Deltaproteobacteria showed no preference for either size-fraction (Fig. S5; Wilcoxon CC), Gammaproteobacteria showed highest relative abundances in the FL size-fraction (Fig. 3; $p = 0.003$; Wilcoxon CC).

Deltaproteobacteria are known affiliates of *nifH* Cluster III, containing mostly anaerobic bacteria (Zehr et al., 2003), which could therefore show preference for the PA size-fraction. Among the 32 dominant NCD ASVs, only three Cluster III ASVs were found in a single size-fraction. A *Desulfovibrio* (ASV 3530) and a *Desulfomicrobium* (ASV 3531) were associated with the PA size-fraction while another *Desulfovibrio* (ASV 1449) was only associated with the FL size-fraction.

Highest *nifH* transcript relative abundances by NCDs occurred in the deep samples (Fig. 2b), where the relative abundance of NCD *nifH* genes also was the highest (Fig. 2a). NCD expression accounted for >1% of the relative *nifH* gene expression in only the coastal stations MP6, MP7, MP19 and the offshore station TF0142 (Fig. 1). None of these samples

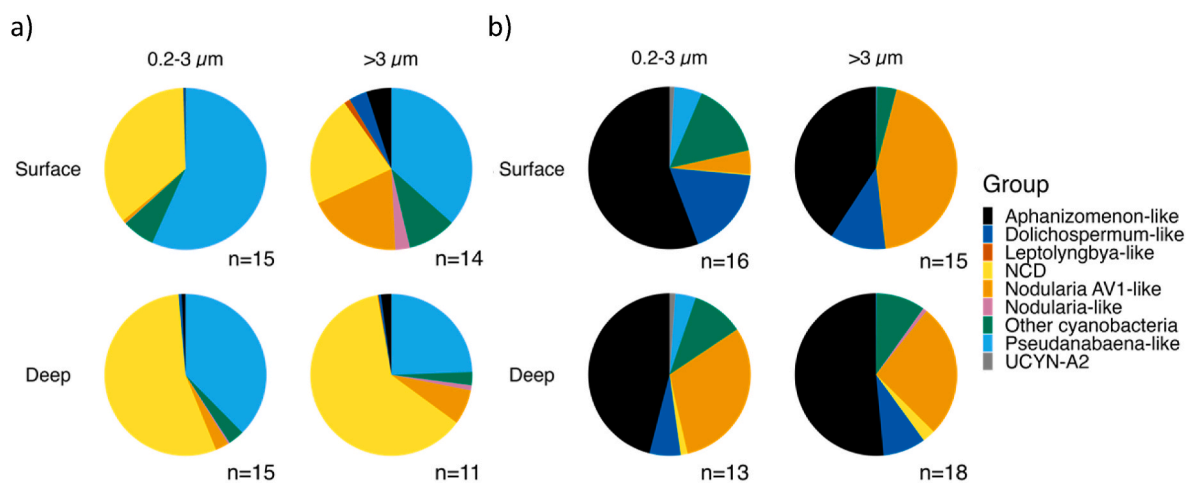


Fig. 2. Composition of present and active diazotrophs in the Baltic Sea represented by average relative abundance of (a) the *nifH* gene and (b) *nifH* gene expression at ‘Surface’ (avg. 1.8 m) and ‘Deep’ (avg. 24 m), for the free-living (0.2–3.0 µm) and the particle associated size-fraction (>3.0 µm). n = number of samples. Non-cyanobacterial diazotrophs (NCD) and *Candidatus Atelocyanobacterium thalassa* (UCYN-A2).

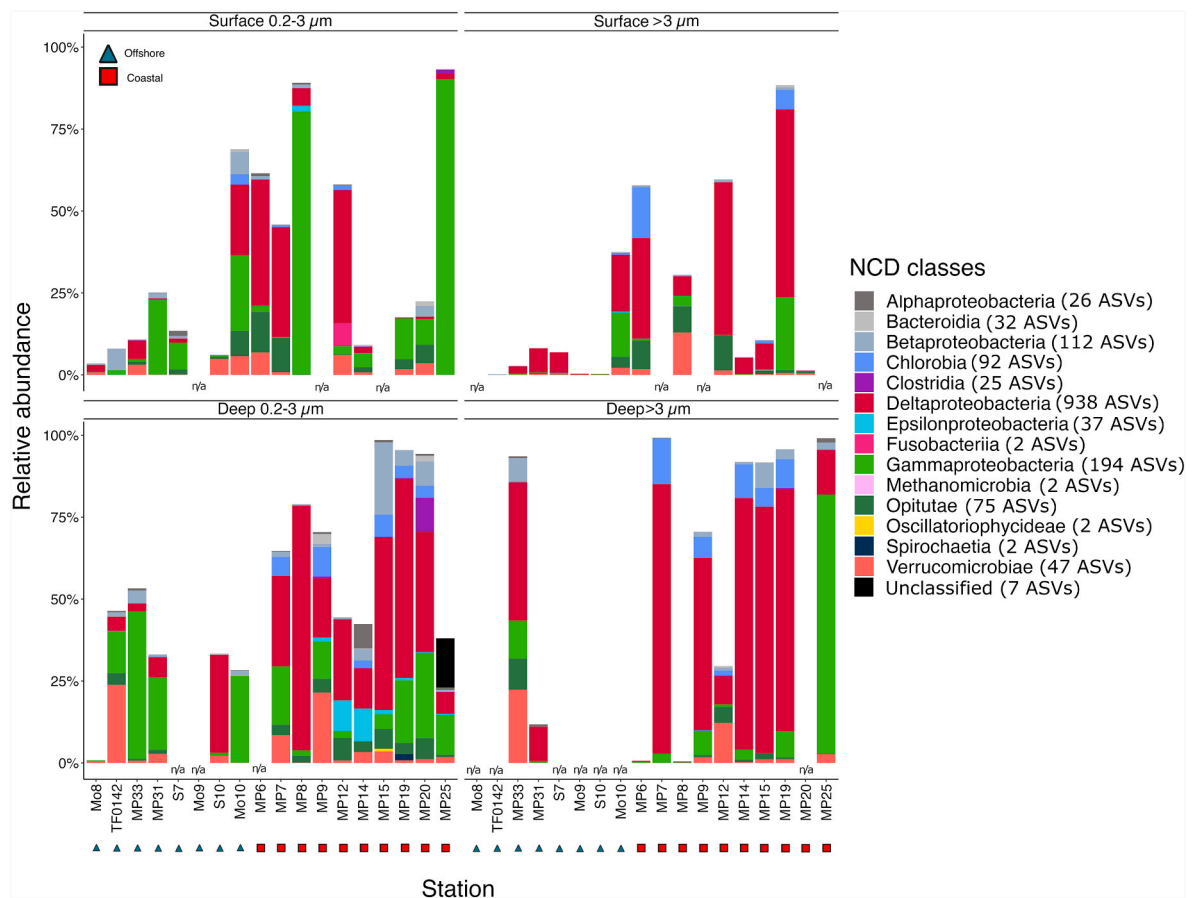


Fig. 3. Relative abundance of non-cyanobacterial diazotrophs (NCD) in all samples. Surface' (avg. 1.8 m) and 'Deep' (avg. 24 m), for the free-living (0.2–3.0 µm) and the particle associated (>3 µm) size-fractions. Stations Mo8 – Mo10 are offshore (>10 km from shore, blue triangles) and stations MP6 - MP25 are coastal (<10 km from shore, red squares). Offshore (blue triangles) and coastal (red squares) stations are ordered from south west going north and down to south east (see Fig. 1). Stations with no available data (n/a). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

were from the surface PA size-fraction. More than half of the NCD *nifH* gene expression in each sample was represented by Gammaproteobacteria, Deltaproteobacteria and Clostridia, with Gammaproteobacteria as especially prominent (section 3.6).

3.5. Environmental influence on *nifH* gene composition

We used Linear Mixed-Effects Models (LME; relative abundance data) and Generalized Linear Mixed Models (GLMM: absence/presence data) to examine the affiliation of diazotrophs with the type of environment (Tables S3 and S4). Offshore areas had a positive effect on cyanobacterial *nifH* relative abundances ($p < 0.001$; LME) and a negative effect on the presence of *Dolichospermum*-like *nifH* ($p = 0.043$; GLMM). Area (offshore/coastal), salinity or size-fraction $\times$ depth interaction did not have significant effects on the variation in *Pseudanabaena*-like *nifH* relative abundances (LME). For *Nodularia* AV1-like and *Aphanizomenon*-like *nifH* presence/absence data, the interaction between size-fraction and depth had a significant effect on the variance with the FL size-fraction and 'surface' having a combined negative effect ($p = 0.043$ and $p = 0.026$, respectively; GLMM).

The relative abundances of NCDs were negatively affected by offshore environments ($p < 0.001$; LME). Similarly, the relative abundance of Deltaproteobacteria were negatively affected ($p = 0.001$; LME). Salinity explained only ca. <8% of the variance in composition of *nifH* classes, at the surface and at depth (Fig. S6, RDA).

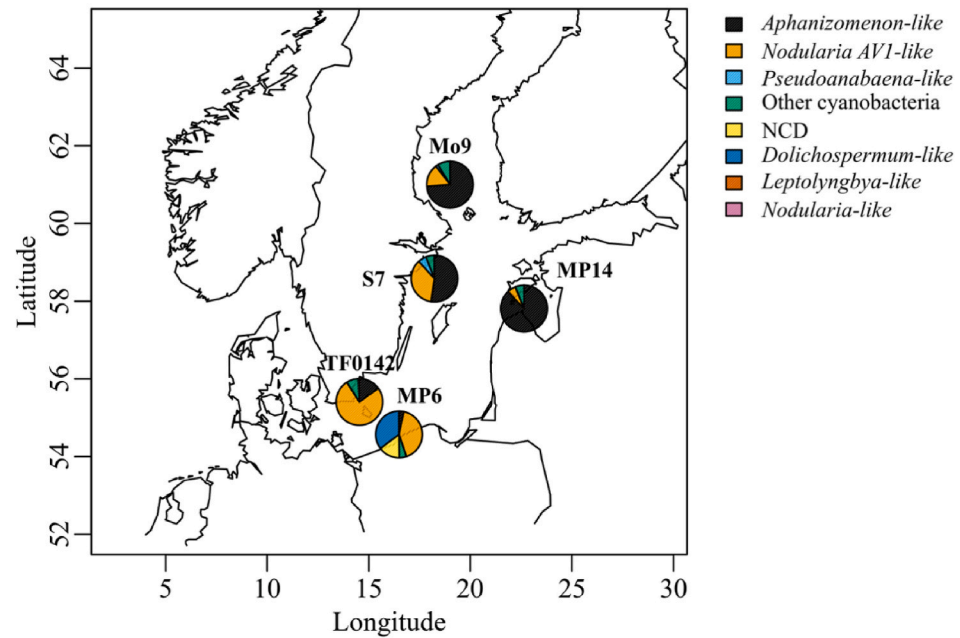
We interpreted the higher turbidity and DOC levels at the coastal stations (see above), as indication of sediment resuspension, and examined differences in diazotroph composition between coastal and

offshore waters to investigate whether distinct diazotroph communities could be detected. Deltaproteobacteria showed ca. 4 times higher relative abundance in coastal surface samples (18.4%) than offshore (4.6%; Fig. S5), and the benthic cyanobacterium *Leptolyngbya* (Sivonen et al., 2007), seemed also to have higher relative abundances at the coastal surface stations (1.4%) compared to offshore (0.08%; Fig. S7). Furthermore, 118 NCD ASVs from our study were identical to ASVs previously found in coastal Baltic Sea sediments (Liesirova et al., 2023). The NCD ASVs accounted for 1.7% and 15.6% of the average relative abundances at offshore and coastal surface samples, respectively. This indicated on average 11 times higher relative abundance in the coastal surface samples compared to offshore.

3.6. Expression of diazotrophs represented by 5 stations

The expression of *nifH* genes shows species-specific and diurnal variation (e.g. Evans et al., 2000; Church et al., 2005; Boström et al., 2007a). Therefore, to compare *nifH* gene expression across the Baltic Sea, stations sampled at roughly the same time of day were selected (afternoon; $15:45 \pm 2.15$ h; Fig. 4). The samples at each station were pooled to get an overall picture of the expression at the station. Unfortunately, the dataset was too small to statistically establish putative linkages to environmental conditions, however, there was a trend of increased relative abundance of *Aphanizomenon*-like *nifH* expression towards the north of the Baltic Sea (Fig. 4a). More than half of this expression was represented by ASV 1, which is identical to a sequence previously recovered from the Baltic (EU916460; Farnelid et al., 2009). Expression by *Nodularia* AV1 appeared higher towards the south. The

a)



b)

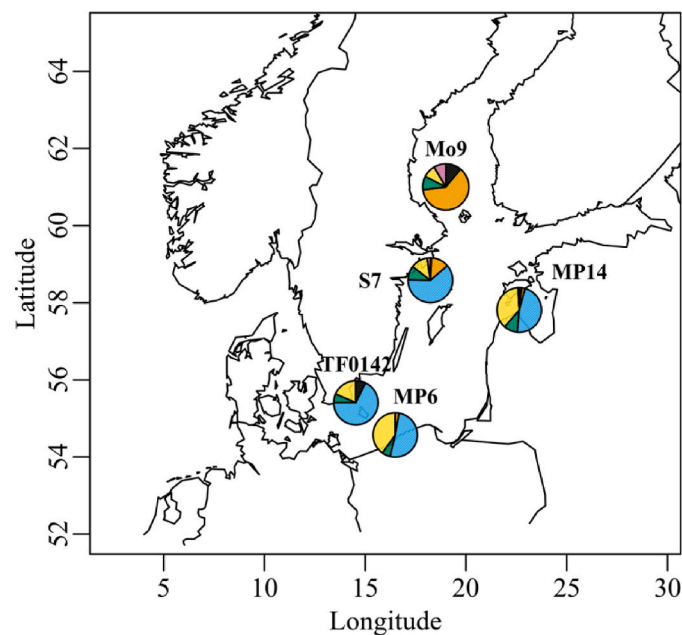


Fig. 4. Composition of present and active diazotrophs from stations sampled at the same time of day (afternoon). Average relative abundance for *nifH* gene expression (a) and *nifH* genes (b; for reference). *nifH* genes from surface, deep, and both size-fractions were pooled for each station. Note that at station Mo9 composition is only represented by the surface particle-associated size-fraction.

highest NCD expression (14.9% of the transcripts relative abundances) was found at station MP6 (Fig. 4a), and largely accounted (>90%) for by the gammaproteobacterial *Pseudomonas*-like ASV 325, which was identical to *nifH* gene sequences recovered from Baltic surface water (Farnelid et al., 2014) and sediment (Liesirova et al., 2023). This ASV also showed 99.7% nucleotide similarity to a diazotroph (clade GD11.0) cultivated from the central Baltic Sea (Bentzon-Tilia et al., 2014). Station MP7 was not sampled in the afternoon, but was the only other

station than MP6, which showed substantial NCD *nifH* expression (4.8% of transcript relative abundances). This station had the highest expression by Deltaproteobacteria of any station sampled (1.9% of transcript relative abundances). Deltaproteobacteria accounted for 0.1% of expression on average across all samples.

4. Discussion

Our study yields insights of critical importance for the understanding of N_2 fixation in the Baltic Sea. In particular, the finding that *Pseudanabaena* and NCDs dominate the *nifH* gene pool contrasts with the prevailing view of filamentous heterocyst-forming cyanobacteria as the only relevant N_2 -fixers (e.g. Klawonn et al., 2016). Despite dominating the gene pool, these organisms showed highly variable *nifH* expression illustrating a need for clarifying their significance for local N cycling. The sampling of offshore and coastal environments revealed a striking difference in *nifH* gene composition and expression, conceivably linked to the high particle load of coastal waters, influenced by sediment resuspension.

4.1. *Pseudanabaena*-like ASVs dominated the overall composition of cyanobacterial *nifH*

The dominance of *Pseudanabaena* was surprising, as most studies using microscopy to investigate cyanobacteria report dominance by *Aphanizomenon* sp., sometimes followed by *Nodularia* sp. (e.g., Berner et al., 2018; Eigemann et al., 2019; Klawonn et al., 2016; Kownacka et al., 2020). *Pseudanabaena* spp. has, however, been reported as a dominant cyanobacterium during summer when counted by microscopy (Riemann et al., 2008), but is reported in other studies as a rather minor constituent of phytoplankton (Eigemann et al., 2019) or cyanobacterial biomass (Berner et al., 2018). *Pseudanabaena*-like *nifH* gene sequences have been recovered in Baltic Sea *nifH* gene libraries (Stal et al., 2003; Acinas et al., 2009; Farnelid et al., 2009, 2013), and accounted for 8% of the relative *nifH* gene abundance in the southern Baltic Sea (Reeder et al., 2022). Our finding of *Pseudanabaena* occasionally accounting for more than half of the cyanobacterial *nifH* genes is, nevertheless, striking. The year of our study, 2015, was the second warmest year in the Baltic Sea since 1990 (Siegel and Gerth, 2016). Since warm temperature increases *Pseudanabaena* spp. biomass (Berner et al., 2018), we speculate that high temperature caused the exceptionally high prevalence of *Pseudanabaena*-like *nifH*. Importantly, the recurrent findings of *Pseudanabaena* spp. in Baltic Sea plankton highlights a need for mapping its importance for N_2 fixation, in space and over time. In particular, considering the future elevated temperatures predicted for the Baltic Sea (Meier, 2006).

4.2. Resuspension impacts the coastal diazotrophic community

Consistent with an earlier study suggesting complex environmental regulation of diazotrophs in the Baltic Sea (Farnelid et al., 2009), no clear effect of discrete environmental parameters on the *nifH* gene composition was found. The negligible impact of salinity on composition was particularly surprising, given its reported effects on *nifH* gene distribution (Reeder et al., 2022), cyanobacterial community composition (Dupont et al., 2014; Olofsson et al., 2020), and bacterial composition in the Baltic Sea (Herlemann et al., 2011; Dupont et al., 2014). We, therefore, speculate that the salinity gradient sampled (4.3–8.6) is not a strong driver of diazotroph community structure, like it has been concluded for general bacterioplankton community structure (Herlemann et al., 2011).

We hypothesized that the anticipated higher productivity and load of particles in coastal areas would cause differences in composition of diazotrophs between coastal and offshore areas, and between size-fractions in the plankton. Our sampling confirmed that the coastal stations differed from offshore stations by having higher DOC content at the surface and higher turbidity at both sampled depths. Turbidity can be considered a proxy for suspended sediment in the water column (Davies-Colley and Smith, 2001), and the elevated levels of turbidity and DOC may, therefore, suggest sediment resuspension and/or terrestrial inputs at the coastal stations. Such input may promote diazotrophic diversity and N_2 fixation in estuarine and coastal waters (Mulholland

et al., 2012; Bentzon-Tilia et al., 2015b; Pedersen et al., 2018; Liesirova et al., 2023), and diazotrophs, particularly putative anaerobic NCDs, likely of sediment origin have previously been observed in estuarine waters (Moisander et al., 2007; Cabello et al., 2020; Hallström et al., 2022). Indeed, we found higher relative abundances of NCDs at the coastal stations, whereas highest relative abundance of cyanobacterial *nifH* was found offshore. Many diazotrophs previously found in sediment samples of coastal origin in the Baltic Sea were found at our coastal stations. For instance, *nifH* genes affiliated with *Leptolyngbya*, a cyanobacterium known to be associated with Baltic Sea sediments and mats (Sivonen et al., 2007; Włodarska-Kowalczyk et al., 2014), were found at most of the coastal stations, accounting for on average 2.4% of the *nifH* reads from the PA surface size-fraction. In contrast, *Leptolyngbya* accounted for <0.1% of the *nifH* reads from all sample categories from the offshore stations, and was detected at <40% of the offshore stations. We also recovered 118 NCD ASVs identical to *nifH* phylotypes from coastal Baltic sediments (Liesirova et al., 2023). These ASVs showed on average 11 times higher relative abundance at the coast compared to offshore. These observations point to sediment resuspension as a strong driver of diazotroph composition at the coastal stations.

Furthermore, the coastal stations in our study showed a roughly 4-fold higher relative abundance of Deltaproteobacteria than the offshore stations. They have previously been found in Baltic sediments (Bertics et al., 2013; Liesirova et al., 2023) and belong mainly to the *nifH* Cluster III, which contains primarily strict anaerobes (Zehr et al., 2003). This makes their presence in the water-column puzzling and has led to discussions about their possible association with pelagic particles (Braun et al., 1999; Church et al., 2005; Debeljak et al., 2021). In our study there was, however, no significant difference in the prevalence of Deltaproteobacteria in the free-living and particulate size-fractions, and many ASVs were found associated with both size-fractions. This highlights the need for investigating adaptations by these putative anaerobic diazotrophs, pertaining to their survival in the oxygenated water-column (Farnelid et al., 2013), but also to the fact that their status as strict anaerobes potentially needs re-evaluation (Man-Aharonovich et al., 2007).

Our present survey of *nifH* genes reveals large differences in the composition of putative diazotrophs between coastal and offshore stations in the Baltic Sea. This difference conceivably pertains partially to coastal upwelling and sediment resuspension, and preponderance of groups of diazotrophs (e.g. *Pseudanabaena* spp. and NCDs) for which the ecology is poorly understood. Strikingly, the composition of the putative diazotrophs (DNA) were only to some extent mirrored in the composition of the active *nifH* expressing taxa (RNA; see below).

4.3. Heterocyst-forming cyanobacteria account for a majority of *nifH* expression

The heterocyst-forming *Aphanizomenon*-like, *Nodularia* AV1-like and *Dolichospermum*-like ASVs accounted for almost 80% of our *nifH* transcripts across all stations. This may even be an underestimate since *nifH* gene expression in *Nodularia* has been found to peak at noon (Boström et al., 2007b), at which time only few stations were sampled. They were found in both size-fractions, as observed earlier (Wasmund et al., 2001; Farnelid et al., 2009), likely due to variable cell sizes (Wojtasiewicz and Stoń-Egert, 2016) or break-up of filaments during filtration. The prominence of these taxa among *nifH* gene transcripts is consistent with earlier studies (Boström et al., 2007b; Klawonn et al., 2016) and likely also reflect a main contribution to overall N_2 fixation (Klawonn et al., 2016).

While *Pseudanabaena* spp. dominated the cyanobacterial *nifH* genes, it accounted for a relatively minor proportion of the *nifH* transcripts. The reason for this apparent discrepancy is unknown. The *nifH* gene expression by *Pseudanabaena* falls in line with earlier data from the Baltic Sea Proper (Farnelid et al., 2013) and documented low cell-specific N_2 fixation by *Pseudanabaena* from the same station and

year/month, as sampled by us (station S7; Eigemann et al., 2019). This apparently contrasts with labeling it as a non-N₂-fixer (Berner et al., 2018) and an earlier study from the station (station S7) where N₂ fixation by *Pseudanabaena* spp. was not measurable (Klawonn et al., 2016). However, these observations may originate from the fact that only some strains of Baltic *Pseudanabaena* have *nifH* and thereby the ability to carry out N₂ fixation (Acinas et al., 2009). Interestingly, the cell-specific N₂ fixation by *Pseudanabaena* was measured during a late-stage cyanobacterial bloom (Eigemann et al., 2019), hinting at *Pseudanabaena* spp. as an opportunistic N₂-fixer, becoming active only under specific conditions. Such strategy could potentially explain our observation of *Pseudanabaena* spp. dominating the *nifH* gene pool, while accounting for only a minor fraction of the *nifH* gene transcripts. Moreover, it underlines the need to unravel spatio-temporal dynamics of N₂ fixation in *Pseudanabaena* spp. for determining its importance for nitrogen cycling in the Baltic Sea.

The UCYN-A/haptophyte symbiosis has been recovered from coastal areas worldwide (e.g. Short and Zehr, 2007; Mulholland et al., 2019; Messer et al., 2015; Shiozaki et al., 2015; Cabello et al., 2020; Hallström et al., 2022). The sublineage UCYN-A2 is thought to have a coastal association (Turk-Kubo et al., 2017, 2021) and has previously been detected in the southern Baltic Sea (Reeder et al., 2022) and connected areas west of the Baltic Sea (Bentzon-Tilia et al., 2015b; Scavotto et al., 2015). Consistent with these findings, we only found high UCYN-A2 *nifH* expression (19–25% of the transcripts) at the relatively saline (7.93) and westernmost station (Mo8). We, therefore, speculate that UCYN-A2 makes a hitherto unaccounted for contribution to N₂ fixation in the westernmost part of the Baltic Sea but that it does not thrive in the low salinities of the central and northern parts. Deltaproteobacteria was the dominant class of NCDs (on average 20.6% of relative abundance) but accounted on average for only 0.1% of the *nifH* gene transcripts. However, the only NCD to show substantial *nifH* expression was *Pseudomonas*-like ASV 325. *Pseudomonas* accounted on average for 5% of the overall *nifH* genes, and was found at all stations at varying depth and size-fractions, except at the most western station (Mo8). Almost identical (99.7%) sequences have been found in other waters (Turk et al., 2011; JN097381.1 GenBank). Identical or similar (99.7%) sequences have been recovered from the Baltic Sea (Farnelid et al., 2013, 2014; Bentzon-Tilia et al., 2014). In addition, a Baltic *Pseudomonas stutzeri* strain (BAL361, identical to ASV 325) can sustain growth by N₂ fixation (Bentzon-Tilia et al., 2015a). The *Pseudomonas nifH* gene expression coincided with an upwelling event (as reported elsewhere; Siegel and Gerth, 2016), in our dataset suggested by elevated DIP, lowest surface temperature, lowest surface bacterial abundances and at depth the smallest HNA to LNA ratio. Indeed, cool and nutrient rich conditions have been suggested to select for gammaproteobacterial NCDs in the Baltic (Farnelid et al., 2009) and may, therefore, have stimulated the high *Pseudomonas nifH* gene expression at these specific coastal sites. Since upwelling is a reoccurring phenomenon along the southern coast of the Baltic Sea (Lehmann and Myrberg, 2008; Zhang et al., 2022), we speculate that *Pseudomonas* regularly contributes to N₂ fixation in this region.

To gain insight into patterns of *nifH* expression across the Baltic Sea, we compared five stations sampled at the same time of day (afternoon). *Aphanizomenon* accounted for between 3 and 88% of the transcripts and showed increased relative expression at lower salinity, consistent with its known preference for low salinity (Lehtimäki et al., 1997) and prevalence in the low saline northern Baltic Sea (Stal et al., 2003; Kownacka et al., 2020). In contrast, *Nodularia* is known to be more common in the south and central Baltic Sea (Kownacka et al., 2020), which corresponded to *Nodularia* (ASV1) showing higher relative expression in the south. These data should, however, be interpreted with caution as diurnal patterns in *nifH* expression may differ between taxa.

5. Concluding remarks

Our survey of nitrogenase genes and their expression revealed presence of several important groups, beyond the filamentous heterocyst-forming taxa, that are well-documented by microscopy. NCDs accounted for 43% of the diazotroph community composition, but only 1% of the total *nifH* gene transcripts. Furthermore, *Pseudanabaena* was the dominating cyanobacteria, but was not among the main *nifH* gene expressing taxa. This highlights that composition and expression of *nifH* can be remarkably different, something that has been repeatedly observed (Short and Zehr, 2007; Bentzon-Tilia et al., 2015a; Yang et al., 2019; Hallström et al., 2022). This discrepancy, as well as the differences in diazotroph composition between the offshore and coastal environments, emphasize the importance of including both composition and expression, as well as distinguishing between coastal and offshore environments, when monitoring diazotrophs and their activity in the Baltic Sea.

Funding

This work was supported by a Fulbright travel grant to ERSS, grants 217-00089B (Independent Research Fund Denmark) and 9096-00008B (The Danish Agency for Science and Higher Education) to LR, grant SAW-2014-IOW-2 (German Leibniz Association) to ML, and KTK was partially supported by NSF (OCE-2023498).

CRedit authorship contribution statement

Ellen R. Salamon Slater: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Data curation. **Kendra A. Turk-Kubo:** Writing – review & editing, Supervision, Methodology. **Søren Hallström:** Writing – review & editing, Methodology, Data curation. **Katharina Keszy:** Writing – review & editing, Methodology, Investigation, Data curation. **Peeter Laas:** Writing – review & editing, Methodology, Investigation, Data curation. **Jonathan Magasin:** Writing – review & editing, Methodology, Data curation. **Jonathan P. Zehr:** Writing – review & editing, Project administration. **Matthias Labrenz:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Lasse Riemann:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We thank the crew and captain on the R/V Poseidon, chief scientist Sonja Oberbeckmann, as well as Andreas Müller, Stephanie Mothes, Christina Hensler, Christian Burmeister, Lars Kreuzer, Birgit Sadkowiak and Jenny Jeschek from Leibniz Institute for Baltic Sea Research Warnemünde for helping with sampling and nutrient, DOC and Chl *a* measurements. In addition, we thank members of the Riemann and Zehr research groups for support and insightful input. We also acknowledge the invaluable help from The Data Science Lab at the University of Copenhagen.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecss.2023.108527>.

References

- Acinas, S.G., Haverkamp, T.H.A., Huisman, J., Stal, L.J., 2009. Phenotypic and genetic diversification of *Pseudanabaena* spp. (cyanobacteria). *ISME J.* 3 (1), 31–46. <https://doi.org/10.1038/ismej.2008.78>.
- Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. URL: www.bioinformatics.babraham.ac.uk/projects/fastqc.
- Barnett, D., Arts, I., Penders, J., 2021. microViz: an R package for microbiome data visualization and statistics. *J. Open Source Softw.* 6 (63), 3201. <https://doi.org/10.21105/joss.03201>.
- Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Software* 67 (1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Bentzon-Tilia, M., Farnelid, H., Jürgens, K., Riemann, L., 2014. Cultivation and isolation of N₂-fixing bacteria from sub-oxic waters in the Baltic Sea. *FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Ecol.* 88 (2), 358–371. <https://doi.org/10.1111/1574-6941.12304>.
- Bentzon-Tilia, M., Severin, I., Hansen, L.H., Riemann, L., 2015a. Genomics and ecophysiology of heterotrophic nitrogen fixing bacteria isolated from estuarine surface water. *mBio* 6 (4), e00929. <https://doi.org/10.1128/mbio.00929-15>.
- Bentzon-Tilia, M., Traving, S.J., Mantikci, M., Knudsen-Leerbeck, H., Hansen, J.L.S., Markager, S., Riemann, L., 2015b. Significant N₂ fixation by heterotrophs, photoheterotrophs and heterocystous cyanobacteria in two temperate estuaries. *ISME J.* 9 (2), 273–285. <https://doi.org/10.1038/ismej.2014.119>.
- Berner, C., Bertos-Fortis, M., Pinhassi, J., Legrand, C., 2018. Response of microbial communities to changing climate conditions during summer cyanobacterial blooms in the Baltic Sea. *Front. Microbiol.* 9, 1562. <https://doi.org/10.3389/fmicb.2018.01562>.
- Bertics, V.J., Löscher, C.R., Salonen, I., Dale, A.W., Gier, J., Schmitz, R.A., Treude, T., 2013. Occurrence of benthic microbial nitrogen fixation coupled to sulfate reduction in the seasonally hypoxic Eckernförde Bay, Baltic Sea. *Biogeosciences* 10 (3), 1243–1258. <https://doi.org/10.5194/bg-10-1243-2013>.
- Bombar, D., Paerl, R.W., Riemann, L., 2016. Marine non-cyanobacterial diazotrophs: moving beyond molecular detection. *Trends Microbiol.* 24 (11), 916–927. <https://doi.org/10.1016/j.tim.2016.07.002>.
- Boström, K.H., Riemann, L., Kühl, M., Hagström, Å., 2007a. Isolation and gene quantification of heterotrophic N₂-fixing bacterioplankton in the Baltic Sea. *Environ. Microbiol.* 9 (1), 152–164. <https://doi.org/10.1111/j.1462-2920.2006.01124.x>.
- Boström, K.H., Riemann, L., Zweifel, U.L., Hagström, Å., 2007b. *Nodularia* sp. nifH gene transcripts in the Baltic Sea proper. *J. Plankton Res.* 29 (4), 391–399. <https://doi.org/10.1093/plankt/fbm019>.
- Braun, S.T., Proctor, L.M., Zani, S., Mellon, M.T., Zehr, J.P., 1999. Molecular evidence for zooplankton-associated nitrogen-fixing anaerobes based on amplification of the *nifH* gene. *FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Ecol.* 28 (3), 273–279. <https://doi.org/10.1111/j.1574-6941.1999.tb00582.x>.
- Cabello, A.M., Turk-Kubo, K.A., Hayashi, K., Jacobs, L., Kudela, R.M., Zehr, J.P., 2020. Unexpected presence of the nitrogen-fixing symbiotic cyanobacterium UCYN-A in Monterey Bay, California. *J. Phycol.* 56 (6), 1521–1533. <https://doi.org/10.1111/jpy.13045>.
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13 (7), 581–583. <https://doi.org/10.1038/nmeth.3869>.
- Chakraborty, S., Andersen, K.H., Visser, A.W., Inomura, K., Follows, M.J., Riemann, L., 2021. Quantifying nitrogen fixation by heterotrophic bacteria in sinking marine particles. *Nat. Commun.* 12, 4085. <https://doi.org/10.1038/s41467-021-23875-6>.
- Charvet, S., Riemann, L., Alneberg, J., Andersson, A.F., von Borries, J., Fischer, U., Labrenz, M., 2019. AFISys - an autonomous instrument for the preservation of brackish water samples for microbial metatranscriptome analysis. *Water Res.* 149, 351–361. <https://doi.org/10.1016/j.watres.2018.11.017>.
- Church, M.J., Short, C.M., Jenkins, B.D., Karl, D.M., Zehr, J.P., 2005. Temporal patterns of nitrogenase gene (*nifH*) expression in the oligotrophic North Pacific Ocean. *Appl. Environ. Microbiol.* 71 (9), 5362–5370. <https://doi.org/10.1128/AEM.71.9.5362-5370.2005>.
- Davies-Colley, R.J., Smith, D.G., 2001. Turbidity, suspended sediment, and water clarity: a review. *JAWRA J. Am. Water Resour. Assoc.* 37, 1085–1101. <https://doi.org/10.1111/j.1752-1688.2001.tb03624.x>.
- Debeljak, P., Blain, S., Bowie, A., van der Merwe, P., Bayer, B., Obernosterer, I., 2021. Homeostasis drives intense microbial trace metal processing on marine particles. *Limnol. Oceanogr.* 66 (10), 3842–3855. <https://doi.org/10.1002/lno.11923>.
- Dupont, C.L., Larsson, J., Yooseph, S., Ininbergs, K., Goll, J., Asplund-Samuelsson, J., McCrow, J.P., Celepi, N., Allen, L.Z., Ekman, M., Lucas, A.J., Hagström, Å., Thiagarajan, M., Brindel, B., Richter, A.R., Andersson, A.F., Tenney, A., Lundin, D., Tovchigrechko, A., Nylander, J.A.A., Bami, D., Badger, J.H., Allen, A.E., Rusch, D.B., Hoffman, J., Norrby, E., Friedman, R., Pinhassi, J., Venter, J.C., Bergman, B., 2014. Functional tradeoffs underpin salinity-driven divergence in microbial community composition. *PLoS One* 9 (2), e89549. <https://doi.org/10.1371/journal.pone.0089549>.
- Edgar, R., 2016. UCHIME2: improved chimera prediction for amplicon sequencing. *bioRxiv*, 074252. <https://doi.org/10.1101/074252>. Preprint.
- Edgar, R., 2021. MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping. Preprint bioRxiv 2021. <https://doi.org/10.1101/2021.06.20.449169>.
- Eigemann, F., Vogts, A., Voss, M., Zoccarato, L., Schulz-Vogt, H., 2019. Distinctive tasks of different cyanobacteria and associated bacteria in carbon as well as nitrogen fixation and cycling in a late stage Baltic Sea bloom. *PLoS One* 14 (12), e0223294. <https://doi.org/10.1371/journal.pone.0223294>.
- European Environment Agency, 2006. The changing faces of Europe's coastal areas. Publications Office. URL: <https://op.europa.eu/s/y0c1>.
- Evans, A.M., Gallon, J.R., Jones, A., Staal, M., Stal, L.J., Villbrandt, M., Walton, T.J., 2000. Nitrogen fixation by Baltic cyanobacteria is adapted to the prevailing photon flux density. *New Phytol.* 147 (2), 285–297. <https://doi.org/10.1046/j.1469-8137.2000.00696.x>.
- Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., Kurilshikov, A., Bonder, M.J., Valles-Colomer, M., Vandeputte, D., Tito, R.Y., Chaffron, S., Rymenans, L., Verspecht, C., de Sutter, L., Lima-Mendez, G., D'hoë, K., Jonckheere, K., Homola, D., Garcia, R., Tigcheel, E., Eeckhaert, L., Fu, J., Henckaerts, L., Zhernakova, A., Wijmenga, C., Raes, J., 2016. Population-level analysis of gut microbiome variation. *Science* 352 (6285), 560–564. <https://doi.org/10.1126/science.123503>.
- Farnelid, H., Öberg, T., Riemann, L., 2009. Identity and dynamics of putative N₂-fixing picoplankton in the Baltic Sea proper suggest complex patterns of regulation. *Environ. Microbiol. Rep.* 1 (2), 145–154. <https://doi.org/10.1111/j.1758-2229.2009.00021.x>.
- Farnelid, H., Andersson, A.F., Bertilsson, S., Al-Soud, W.A., Hansen, L.H., Sørensen, S., Steward, G.F., Hagström, Å., Riemann, L., 2011. Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. *PLoS One* 6 (4), e19223. <https://doi.org/10.1371/journal.pone.0019223>.
- Farnelid, H., Bentzon-Tilia, M., Andersson, A.F., Bertilsson, S., Jost, G., Labrenz, M., Jürgens, K., Riemann, L., 2013. Active nitrogen-fixing heterotrophic bacteria at and below the chemocline of the central Baltic Sea. *ISME J.* 7 (7), 1413–1423. <https://doi.org/10.1038/ismej.2013.26>.
- Farnelid, H., Harder, J., Bentzon-Tilia, M., Riemann, L., 2014. Isolation of heterotrophic diazotrophic bacteria from estuarine surface waters. *Environ. Microbiol.* 16 (10), 3072–3082. <https://doi.org/10.1111/1462-2920.12335>.
- Farnelid, H., Turk-Kubo, K., del Carmen Muñoz-Marín, M., Zehr, J.P., 2016. New insights into the ecology of the globally significant uncultured nitrogen-fixing symbiont UCYN-A. *Aquat. Microb. Ecol.* 77 (3), 128–138. <https://doi.org/10.3354/ame01794>.
- Farnelid, H., Turk-Kubo, K., Ploug, H., Ossolinski, J.E., Collins, J.R., Van Mooy, B.A.S., Zehr, J.P., 2019. Diverse diazotrophs are present on sinking particles in the North Pacific Subtropical Gyre. *ISME J.* 13, 170–182. <https://doi.org/10.1038/s41396-018-0259-x>.
- Finn, R.D., Clements, J., Eddy, R.S., 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39 (2), W29–W37. <https://doi.org/10.1093/nar/gkr367>.
- Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28 (23), 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>.
- Gerritsen, H., 2018. mapplots: data Visualisation on Maps (R package version 1.5.1). URL: <https://CRAN.R-project.org/package=mapplots>.
- Grasshoff, K., Ehrhardt, M., Kremling, K., Anderson, L.G., 1999. Methods of Sea-Water Analysis, third ed. Wiley-VCH, Weinheim, Germany, p. 576. <https://doi.org/10.1002/9783527613984>.
- Hallström, S., Benavides, M., Salamon, E.R., Evans, C.W., Potts, L.J., Granger, J., Tobias, C.R., Moisaner, P.H., Riemann, L., 2022. Pelagic N₂ fixation dominated by sediment diazotrophic communities in a shallow temperate estuary. *Limnol. Oceanogr.* 67 (2), 364–378. <https://doi.org/10.1002/lno.11997>.
- Herlemann, D.P.R., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J.J., Andersson, A.F., 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* 5 (10), 1571–1579. <https://doi.org/10.1038/ismej.2011.41>.
- Huang, Y., Niu, B., Gao, Y., Fu, L., Li, W., 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26 (5), 680–682. <https://doi.org/10.1093/bioinformatics/btq003>.
- IOC, 1994. SCOR protocols for the joint global ocean flux study (JGOFS) core measurements. In: IOC. Manuals and Guides, vol. 29. UNESCO, Paris, France. URL: <https://unesdoc.unesco.org/ark:/48223/pf0000099739.locale=en>.
- Janson, S., Granéli, E., 2002. Phylogenetic analyses of nitrogen-fixing cyanobacteria from the Baltic Sea reveal sequence anomalies in the phycocyanin operon. *Int. J. Syst. Evol. Microbiol.* 52 (4), 1397–1404. <https://doi.org/10.1099/00207713-52-4-1397>.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30 (4), 772–780. <https://doi.org/10.1093/molbev/mst010>.
- Katoh, K., Misawa, K., Kuma, K.-I., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30 (14), 3059–3066. <https://doi.org/10.1093/nar/gkf436>.
- Klawon, I., Nahar, N., Walve, J., Andersson, B., Olofsson, M., Svedén, J.B., Littmann, S., Whitehouse, M.J., Kuypers, M.M.M., Ploug, H., 2016. Cell-specific nitrogen- and carbon-fixation of cyanobacteria in a temperate marine system (Baltic Sea). *Environ. Microbiol.* 18 (12), 4596–4609. <https://doi.org/10.1111/1462-2920.13557>.
- Kownacka, J., Busch, S., Göbel, J., Gromisz, S., Hällfors, H., Högländer, H., Huseby, S., Jaanus, A., Jakobsen, H.H., Johansen, M., Johansson, M., Jurgensone, I., Liebecke, N., Krasniewski, W., Kremp, A., Lehtinen, S., Olenina, I., Weber, M. v., Wasmund, N., 2020. Cyanobacteria biomass 1990-2019. HELCOM Baltic Sea environment fact sheets 2020. URL: www.helcom.fi/baltic-sea-trends/environment-fact-sheets/eutrophication/cyanobacteria-biomass/.

- Kullenberg, G., Jacobsen, T.S., 1981. The Baltic Sea: an outline of its physical oceanography area definition and topography. *Mar. Pollut. Bull.* 12 (6), 183–186. [https://doi.org/10.1016/0025-326X\(81\)90168-5](https://doi.org/10.1016/0025-326X(81)90168-5).
- Lahti, L., Shetty, S., 2019. Microbiome R package. URL: <http://microbiome.github.com/microbiome>.
- Larsson, U., Hajdu, S., Walve, J., Elmgren, R., 2001. Baltic Sea nitrogen fixation estimated from the summer increase in upper mixed layer total nitrogen. *Limnol. Oceanogr.* 46 (4), 811–820. <https://doi.org/10.4319/lo.2001.46.4.0811>.
- Legrand, C., Fridolfsson, E., Bertos-Fortis, M., Lindehoff, E., Larsson, P., Pinhassi, J., Andersson, A., 2015. Interannual variability of phyto-bacterioplankton biomass and production in coastal and off-shore waters of the Baltic Sea. *Ambio* 44, 427–438. <https://doi.org/10.1007/s13280-015-0662-8>.
- Lehmann, A., Myrberg, K., 2008. Upwelling in the Baltic Sea - a review. *J. Mar. Syst.* 74, 3–12. <https://doi.org/10.1016/j.jmarsys.2008.02.010>.
- Lehtimäki, J., Moisaner, P., Sivonen, K., Kononen, K., 1997. Growth, nitrogen fixation, and nodularin production by two Baltic Sea cyanobacteria. *Appl. Environ. Microbiol.* 63 (5), 1647–1656. <https://doi.org/10.1128/aem.63.5.1647-1656.1997>.
- Letunic, I., Bork, P., 2021. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49 (W1), W293–W296. <https://doi.org/10.1093/nar/gkab301>.
- Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22 (13), 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>.
- Liesirova, T., Aarenstrup-Launbjerg, T., Hallström, S., Bittner, M., Riemann, L., Voss, M., 2023. Nitrogen-fixing sulfate reducing bacteria in shallow coastal sediments under simulated resuspension. *Estuar. Coast Shelf Sci.* 280, 108165. <https://doi.org/10.1016/j.ecss.2022.108165>.
- Lysiak-Pastuszak, E., Andersens, H.C., 2004. ICES. Chemical measurements in the Baltic Sea: guidelines on quality assurance. *ICES Tech. Mar. Environ. Sci.* 35, 149. <https://doi.org/10.17895/ices.pub.5059>.
- Man-Aharonovich, D., Kress, N., Zeev, E.B., Berman-Frank, I., Bèjà, O., 2007. Molecular ecology of *nifH* genes and transcripts in the eastern Mediterranean Sea. *Environ. Microbiol.* 9 (9), 2354–2363. <https://doi.org/10.1111/j.1462-2920.2007.01353.x>.
- Marie, D., Brussaard, C.P.D., Thyrhaug, R., Bratbak, G., Vaulot, D., 1999. Enumeration of marine viruses in culture and natural samples by flow cytometry. *Appl. Environ. Microbiol.* 65 (1), 45–52. <https://doi.org/10.1128/AEM.65.1.45-52.1999>.
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17 (1), 10–12. <https://doi.org/10.14806/ej.17.1.200>.
- Martínez-Pérez, C., Mohr, W., Löscher, C.R., Dekazemacker, J., Littmann, S., Yilmaz, P., Lehnen, N., Fuchs, B.M., Lavik, G., Schmitz, R.A., LaRoche, J., Kuypers, M.M.M., 2016. The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nat. Microbiol.* 1, 16163. <https://doi.org/10.1038/nmicrobiol.2016.163>.
- McMurdie, P.J., Holmes, S., 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8 (4), e61217. <https://doi.org/10.1371/journal.pone.0061217>.
- McMurdie, P.J., Holmes, S., 2014. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* 10 (4) <https://doi.org/10.1371/journal.pcbi.1003531>.
- Meier, H.E.M., 2006. Baltic Sea climate in the late twenty-first century: a dynamical downscaling approach using two global models and two emission scenarios. *Clim. Dynam.* 27 (1), 39–68. <https://doi.org/10.1007/s00382-006-0124-x>.
- Messer, L.F., Doubell, M., Jeffries, T.C., Brown, M.V., Seymour, J.R., 2015. Prokaryotic and diazotrophic population dynamics within a large oligotrophic in-verse estuary. *Aquat. Microb. Ecol.* 74 (1), 1–15. <https://doi.org/10.3354/ame01726>.
- Messer, L.F., Brown, M.V., van Ruth, P.D., Doubell, M., Seymour, J.R., 2021. Temperate southern Australian coastal waters are characterised by surprisingly high rates of nitrogen fixation and diversity of diazotrophs. *PeerJ* 9, e10809. <https://doi.org/10.7717/peerj.10809>.
- Moisaner, P.H., Morrison, A.E., Ward, B.B., Jenkins, B.D., Zehr, J.P., 2007. Spatial-temporal variability in diazotroph assemblages in Chesapeake Bay using an oligonucleotide *nifH* microarray. *Environ. Microbiol.* 9 (7), 1823–1835. <https://doi.org/10.1111/j.1462-2920.2007.01304.x>.
- Mulholland, M.R., Bernhardt, P.W., Blanco-Garcia, J.L., Mannino, A., Hyde, K., Mondragon, E., Turk, K., Moisaner, P.H., Zehr, J.P., 2012. Rates of dinitrogen fixation and the abundance of diazotrophs in North American coastal waters between Cape Hatteras and Georges Bank. *Limnol. Oceanogr.* 57 (4), 1067–1083. <https://doi.org/10.4319/lo.2012.57.4.1067>.
- Mulholland, M.R., Bernhardt, P.W., Widner, B.N., Selden, C.R., Chappell, P.D., Clayton, S., Mannino, A., Hyde, K., 2019. High rates of N₂ fixation in temperate, western North Atlantic coastal waters expand the realm of marine diazotrophy. *Global Biogeochem. Cycles* 33 (7), 826–840. <https://doi.org/10.1029/2018GB006130>.
- Niemistö, L., Melvasalo, T., Niemi, A., Rinne, I., 1989. Blue-green algae and their nitrogen fixation in the Baltic Sea in 1980, 1982 and 1984. *Meri* 17, 1–59. URL: http://hel.sinkki.fi/bitstream/handle/10138/157388/Meri_17_1989.pdf?is-Allowed=y&sequence=1.
- Olofsson, M., Suikkanen, S., Kobos, J., Wasmund, N., Karlson, B., 2020. Basin-specific changes in filamentous cyanobacteria community composition across four decades in the Baltic Sea. *Harmful Algae* 91, 101685. <https://doi.org/10.1016/j.hal.2019.101685>.
- Paerl, H.W., Carlont, R.G., Kellogg, W.K., 1988. Control of nitrogen fixation by oxygen depletion in surface-associated microzones. *Nature* 332, 260–262. <https://doi.org/10.1038/332260a0>.
- Pedersen, J.N., Bombar, D., Paerl, R.W., Riemann, L., 2018. Diazotrophs and N₂-Fixation associated with particles in coastal estuarine waters. *Front. Microbiol.* 9, 2759. <https://doi.org/10.3389/fmicb.2018.02759>.
- R Core Team, 2021. R: A Language and Environment for Statistical Computing (R Version 4.1.1). R Foundation for Statistical Computing. URL: www.R-project.org/.
- Raymond, J., Siefert, J.L., Staples, C.R., Blankenship, R.E., 2004. The natural history of nitrogen fixation. *Mol. Biol. Evol.* 21 (3), 541–554. <https://doi.org/10.1093/molbev/msh047>.
- Redfield, B.H., Ketchum, A.C., Richards, F.A., 1963. The influence of organisms on the composition of sea water. In: Hill, M.N. (Ed.), *The Sea: Ideas and Observations on Progress in the Study of the Seas*, vol. 2. Interscience Publishers, New York, pp. 26–77. URL: www.vliz.be/en/imis?mod=ule=ref&refid=28944&printver=sio_n=1&dropIMIStitle=1.
- Reeder, C.F., Stoltenberg, I., Javidpour, J., Löscher, C.R., 2022. Salinity as a key control on the diazotrophic community composition in the southern Baltic Sea. *Ocean Sci.* 18 (2), 401–417. <https://doi.org/10.5194/os-18-401-2022>.
- Rho, M., Tang, H., Ye, Y., 2010. FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res.* 38 (20), e191. <https://doi.org/10.1093/nar/gkq747>.
- Riemann, L., Leitet, C., Pommier, T., Simu, K., Holmfeldt, K., Larsson, U., Hagström, Å., 2008. The native bacterio-plankton community in the central Baltic Sea is influenced by freshwater bacterial species. *Appl. Environ. Microbiol.* 74 (2), 503–515. <https://doi.org/10.1128/AEM.01983-07>.
- Riemann, L., Rahav, E., Passow, U., Grossart, H.-P., de Beer, D., Klawonn, I., Eichner, M., Benavides, M., Bar-Zeev, E., 2022. Planktonic aggregates as hotspots for heterotrophic diazotrophy: the plot thickens. *Front. Microbiol.* 13, 1092. <https://doi.org/10.3389/fmicb.2022.875050>.
- 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>.
- RStudio Team, 2021. RStudio: integrated development environment for R. RStudio, PBC. URL: www.rstudio.com/.
- Scavotto, R.E., Dzallas, C., Bentzon-Tilia, M., Riemann, L., Moisaner, P.H., 2015. Nitrogen-fixing bacteria associated with copepods in coastal waters of the North Atlantic Ocean. *Environ. Microbiol.* 17 (10), 3754–3765. <https://doi.org/10.1111/1462-2920.12777>.
- Schlitzer, R., 2021. Ocean Data View. Alfred Wegener Institute for Polar and Marine Research: Bremerhaven, Germany. URL: www.odv.awi.de.
- Shiozaki, T., Nagata, T., Jichi, M., Furuya, K., 2015. Nitrogen fixation and the diazotroph community in the temperate coastal region of the northwestern North Pacific. *Biogeosciences* 12 (15), 4751–4764. <https://doi.org/10.5194/bg-12-4751-2015>.
- Short, S.M., Zehr, J.P., 2007. Nitrogenase gene expression in the Chesapeake Bay estuary. *Environ. Microbiol.* 9 (6), 1591–1596. <https://doi.org/10.1111/j.1462-2920.2007.01258.x>.
- Short, S.M., Jenkins, B.D., Zehr, J.P., 2004. Spatial and temporal distribution of two diazotrophic bacteria in the Chesapeake Bay. *Appl. Environ. Microbiol.* 70 (4), 2186–2192. <https://doi.org/10.1128/AEM.70.4.2186-2192.2004>.
- Siegel, H., Gerth, M., 2016. Development of Sea Surface Temperature (SST) in the Baltic Sea 2015, vol. 2016. HELCOM Baltic Sea Environment Fact Sheets. URL: <https://helcom.fi/wp-content/uploads/2019/08/SST-development-in-2015-BSEFS-2016.pdf>.
- Sigman, D.M., Hain, M.P., 2012. The biological productivity of the ocean. *Nat. Educ.* 3 (10), 21. URL: <https://www.nature.com/scitable/knowledge/library/the-biological-productivity-of-the-ocean-70631104/>.
- Simon, M., Grossart, H.-P., Schweitzer, B., Ploug, H., 2002. Microbial ecology of organic aggregates in aquatic ecosystems. *Aquat. Microb. Ecol.* 28, 175–211. <https://doi.org/10.3354/ame028175>.
- Sivonen, K., Halinen, K., Sihvonen, L.M., Koskeniemi, K., Sinkko, H., Rantasärkkä, K., Moisaner, P.H., Lyra, C., 2007. Bacterial diversity and function in the Baltic Sea with an emphasis on cyanobacteria. *Ambio* 36 (2), 180–185. [https://doi.org/10.1579/0044-7447\(2007\)36\[180:BDAFTI\]2.0.CO;2](https://doi.org/10.1579/0044-7447(2007)36[180:BDAFTI]2.0.CO;2).
- Stal, L.J., Albertano, P., Bergman, B., von Bröckel, K., Gallon, J.R., Hayes, P.K., Sivonen, K., Walsby, A.E., 2003. BASIC: Baltic Sea cyanobacteria. An investigation of the structure and dynamics of water blooms of cyanobacteria in the Baltic Sea - responses to a changing environment. *Contin. Shelf Res.* 23 (17–19), 1695–1714. <https://doi.org/10.1016/j.csr.2003.06.001>.
- Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30 (9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
- STATECON, 2021. Guidelines for Monitoring of Phytoplankton Species Composition, Abundance and Biomass. URL: helcom.fi/wp-content/uploads/2020/01/HELCOM-Guidelines-for-monitoring-of-phytoplankton-species-composition-abundance-and-biomass.pdf.
- Torgo, L., 2016. Data Mining with R, Learning with Case Studies, second ed. Chapman and Hall/CRC, p. 426.
- Turk, K.A., Rees, A.P., Zehr, J.P., Pereira, N., Swift, P., Shelley, R., Lohan, M., Woodward, E.M.S., Gilbert, J., 2011. Nitrogen fixation and nitrogenase (*nifH*) expression in tropical waters of the eastern North Atlantic. *ISME J.* 5 (7), 1201–1212. <https://doi.org/10.1038/ismej.2010.205>.
- Turk-Kubo, K.A., Farnelid, H.M., Shilova, I.N., Henke, B., Zehr, J.P., 2017. Distinct ecological niches of marine symbiotic N₂-fixing cyanobacterium *Candidatus Atelocyanobacterium thalassa* sublineages. *J. Phycol.* 53 (2), 451–461. <https://doi.org/10.1111/jpy.12505>.
- Turk-Kubo, K.A., Mills, M.M., Arrigo, K.R., van Dijken, G., Henke, B.A., Stewart, B., Wilson, S.T., Zehr, J.P., 2021. UCYN-A/haptophyte symbioses dominate N₂ fixation in the Southern California Current System. *ISME Commun.* 1 (1), 42. <https://doi.org/10.1038/s43705-021-00039-7>.

- Turner, A., Millward, G.E., 2002. Suspended particles: their role in estuarine biogeochemical cycles. *Estuarine. Coast. Shelf Sci.* 55 (6), 857–883. <https://doi.org/10.1006/ecss.2002.1033>.
- Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics with S*, fourth ed. Springer, p. 495. URL: www.stats.ox.ac.uk/pub/MASS4/.
- Vintila, S., Selao, T., Norén, A., Bergman, B., El-Shehaw, R., 2011. Characterization of *nifH* gene expression, modification and rearrangement in *Nodularia spumigena* strain AV1. *FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Ecol.* 77 (2), 449–459. <https://doi.org/10.1111/j.1574-6941.2011.01127.x>.
- Wacklin, P., Hoffmann, L., Komárek, J., 2009. Nomenclatural validation of the genetically revised cyanobacterial genus *Dolichospermum* (Ralfs ex Bornet et Flahault) comb. nova. *Fottea* 9 (1), 59–64. <https://doi.org/10.5507/fot.2009.005>.
- Wasmund, N., Voss, M., Lochte, K., 2001. Evidence of nitrogen fixation by non-heterocystous cyanobacteria in the Baltic Sea and re-calculation of a budget of nitrogen fixation. *Mar. Ecol. Prog. Ser.* 214, 1–14. <https://doi.org/10.3354/meps214001>.
- Wasmund, N., Nausch, G., Schneider, B., Nagel, K., Voss, M., 2005. Comparison of nitrogen fixation rates determined with different methods: a study in the Baltic Proper. *Mar. Ecol. Prog. Ser.* 297, 23–31. <https://doi.org/10.3354/meps297023>.
- Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York, p. 182. URL: <https://ggplot2.tidyverse>.
- Włodarska-Kowalczyk, M., Balazy, P., Kobos, J., Wiktor, J., Zajączkowski, M., Moskal, W., 2014. Large red cyanobacterial mats (*Spirulina subsalsa* Oersted ex Gomont) in the shallow sublittoral of the southern Baltic. *Oceanologia* 56 (3), 661–666. <https://doi.org/10.5697/oc.55-3.661>.
- Wojtasiewicz, B., Stoń-Egiert, J., 2016. Bio-optical characterization of selected cyanobacteria strains present in marine and freshwater ecosystems. *J. Appl. Phycol.* 28 (4), 2299–2314. <https://doi.org/10.1007/s10811-015-0774-3>.
- Xie, Y., 2021. Knitr: a General-Purpose Package for Dynamic Report Generation in R (R Package Version 1.37.). URL: <https://yihui.org/knitr/>.
- Yang, Q.S., Dong, J. de, Ahmad, M., Ling, J., Zhou, W.G., Tan, Y.H., Zhang, Y.Z., Shen, D. D., Zhang, Y.Y., 2019. Analysis of *nifH* DNA and RNA reveals a disproportionate contribution to nitrogenase activities by rare plankton-associated diazotrophs. *BMC Microbiol.* 19, 188. <https://doi.org/10.1186/s12866-019-1565-9>.
- Zani, S., Mellon, M.T., Collier, J.L., Zehr, J.P., 2000. Expression of *nifH* genes in natural microbial assemblages in Lake George, New York, detected by reverse transcriptase PCR. *Appl. Environ. Microbiol.* 66 (7), 3119–3124. <https://doi.org/10.1128/AEM.66.7.3119-3124.2000>.
- Zehr, J.P., Capone, D.G., 2020. Changing perspectives in marine nitrogen fixation. *Science* 368, aay9514. <https://doi.org/10.1126/science.aay9514>.
- Zehr, J.P., McReynolds, L.A., 1989. Use of degenerate oligonucleotides for amplification of the *nifH* gene from the marine cyanobacterium *Trichodesmium thiebautii*. *Appl. Environ. Microbiol.* 55 (10), 2522–2526. <https://doi.org/10.1128/aem.55.10.2522-2526.1989>.
- Zehr, J.P., Mellon, M.T., Zani, S., 1998. New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (*nifH*) genes. *Appl. Environ. Microbiol.* 64 (9), 3444–3450. <https://doi.org/10.1128/AEM.64.9.3444-3450.1998>.
- Zehr, J.P., Jenkins, B.D., Short, S.M., Steward, G.F., 2003. Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environ. Microbiol.* 5 (7), 539–554. <https://doi.org/10.1046/j.1462-2920.2003.00451.x>.
- Zhang, S., Wu, L., Arnqvist, J., Hallgren, C., Rutgersson, A., 2022. Mapping coastal upwelling in the Baltic Sea from 2002 to 2020 using remote sensing data. *Int. J. Appl. Earth Obs. Geoinf.* 114, 103061 <https://doi.org/10.1016/j.jag.2022.103061>.
- Zhu, H., 2021. kableExtra: Construct Complex Table with “Kable” and Pipe Syntax (R Package Version 1.3.4). URL: <https://CRAN.R-project.org/package=kableExtra>.