

RESEARCH

Open Access



Ecology of aerobic anoxygenic phototrophs on a fine-scale taxonomic resolution in Adriatic Sea unravelled by unsupervised neural network

Iva Stojan^{1,2}, Danijela Šantić^{1*}, Cristian Villena-Aleman^{3,4}, Željka Trumbić⁵, Frano Matić⁵, Ana Vrdoljak Tomaš¹, Ivana Lepen Pleić¹, Kasia Piwosz⁶, Grozdan Kušpilić¹, Živana Ninčević Gladan¹, Stefanija Šestanović¹ and Mladen Šolić¹

Abstract

Background Aerobic anoxygenic phototrophs are metabolically highly active, diverse and widespread polyphyletic members of bacterioplankton whose photoheterotrophic capabilities shifted the paradigm about simplicity of the microbial food chain. Despite their considerable contribution to the transformation of organic matter in marine environments, relatively little is still known about their community structure and ecology at fine-scale taxonomic resolution. Up to date, there is no comprehensive (*i.e.* qualitative and quantitative) analysis of their community composition in the Adriatic Sea.

Results Analysis was based on *pufM* gene metabarcoding and quantitative FISH-IR approach with the use of artificial neural network. Significant seasonality was observed with regards to absolute abundances (maximum average abundances in spring $2.136 \pm 0.081 \times 10^4$ cells mL⁻¹, minimum in summer 0.86×10^4 cells mL⁻¹), FISH-IR groups (*Roseobacter* clade prevalent in autumn, other Alpha- and Gammaproteobacteria in summer) and *pufM* sequencing data agglomerated at genus-level. FISH-IR results revealed heterogeneity with the highest average relative contribution of AAPs assigned to *Roseobacter* clade (37.66%), followed by Gammaproteobacteria (35.25%) and general Alphaproteobacteria (31.15%). Community composition obtained via *pufM* sequencing was dominated by Gammaproteobacteria clade NOR5/OM60, specifically genus *Luminiphilus*, with numerous rare genera present in relative abundances below 1%. The use of artificial neural network connected this community to biotic (heterotrophic bacteria, HNA and LNA bacteria, *Synechococcus*, *Prochlorococcus*, picoeukaryotes, heterotrophic nanoflagellates, bacterial production) and abiotic environmental factors (temperature, salinity, chlorophyll *a* and nitrate, nitrite, ammonia, total nitrogen, silicate, and orthophosphate concentration). A type of neural network, neural gas analysis at order-, genus- and ASV-level, resulted in five distinct best matching units (representing particular environments) and revealed that high diversity was generally independent of temperature, salinity, and trophic status of the environment, indicating a potentially dissimilar behaviour of aerobic anoxygenic phototrophs compared to the general bacterioplankton.

Conclusion This research represents the first comprehensive analysis of aerobic anoxygenic phototrophs in the Adriatic Sea on a trophic gradient during a year-round period. This study is also one of the first reports of their genus-level ecology linked to biotic and abiotic environmental factors revealed by unsupervised neural network algorithm,

*Correspondence:

Danijela Šantić
segvic@izor.hr

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



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

paving the way for further research of substantial contribution of this important bacterial functional group to marine ecosystems.

Background

Aerobic anoxygenic phototrophs (AAPs) are a polyphyletic group of bacteria capable of photoheterotrophy. They are omnipresent in diverse habitats, and were discovered in the 1970s [1]. In recent decades, they have been recognised as an important functional group of particular interest in microbial ecology, and extensive studies have been conducted on this topic [2–10]. AAPs are considered ubiquitous in the marine environment with first quantitative estimates of their abundances up to 11% of the total microbial community in the upper open ocean [3]. A recent study reports their record abundance in the Adriatic Sea with up to 30% of the bacterial population in the piconeuston community following open fire events [11].

AAPs are facultative photoheterotrophs that harvest light energy and generate ATP by photophosphorylation using a unique type of bacteriochlorophyll-*a*-containing reaction center. Nevertheless, they primarily rely on dissolved organic matter as an energy source [12]. They exhibit higher growth rates and larger cell volumes compared to other bacterioplankton, making them particularly vulnerable to predation [13–16]. Hence, their contribution to the transformation of both organic and inorganic matter in aquatic environments is substantial [5, 8, 10, 17]. Ecology of AAPs in different ecosystems is rather complex, not yet fully understood, and influenced by a plethora of factors, such as temperature, salinity, nutrient availability, light intensity, and the presence of predators [5, 7, 9, 14, 16]. Furthermore, the abundance and distribution of AAPs has been shown to vary significantly on a spatiotemporal scale with distinct seasonality and could be regulated by ocean circulation patterns and seasonal changes in sunlight availability [5, 9].

Based on the metagenomic approach and metabarcoding of the *pufM* gene, marine AAPs are so far taxonomically classified into the proteobacterial classes Alpha- and Gammaproteobacteria [5, 18, 19]. Sequencing amplicons of the *pufM* gene, encoding the M-chain of the photosynthetic reaction centre complex, is currently the most common approach in the investigation of their ecology. This preference stems from the realisation that metagenomic approach may overlook certain groups that occur at very low abundances [5]. In comparison with sequencing-based methodologies which do not provide quantitative estimates of specific taxa [20], infrared epifluorescence microscopy combined with fluorescence in situ hybridisation (FISH-IR) aims for quantification of specific AAP

groups [21]. Therefore, a combination of qualitative and quantitative methodologies is required for a comprehensive analysis.

Even though various studies have been conducted over the years on the topic of AAP abundance, distribution, ecology, and dynamics [11, 22–25], their community composition remains unknown in the Adriatic Sea. Hence, here we present the first comprehensive analysis of this community in the Adriatic Sea along a trophic gradient during a year-round period. This study: (i) elucidates patterns of AAP distribution and community composition based on *pufM* metabarcoding in the central Adriatic, (ii) quantifies abundances of the main AAP groups with FISH-IR, (iii) reveals biotic and abiotic environmental factors potentially affecting AAPs on a fine-scale taxonomic resolution using a neural gas algorithm, ultimately broadening our knowledge of their composition and ecology.

Methods

Study area, environmental parameters and plankton analysis

A total of 90 samples (Additional file 1: Table S1) were collected on board the R/V BIOS DVA, predominantly on a monthly basis (with the exception of July, September and October) from February 2021 to January 2022 on vertical profiles at three stations in the central Adriatic Sea: the coastal area ST101 (0 m and 35 m depths), the channel CJ007 (0 m, 30 m and 50 m depths) and the open sea CJ009 [0 m, 30 m, 50 m, 75 m, 100 m and deep chlorophyll maximum (dChlMax) depths] (Fig. 1).

Seasons were determined according to the criteria of a comprehensive historical hydrographic dataset of the Adriatic Sea, with the months January to April being considered winter, May and June spring, July to October summer and November and December autumn [26]. Temperature, salinity, nutrients concentrations and chlorophyll *a* (Chl *a*) were measured as described in a previous study [27] and values are shown in Fig. 2. Abundances of picoplankton community members, namely *Synechococcus* and *Prochlorococcus*, picoeukaryotes (PE), heterotrophic bacteria, high and low nucleic acid bacteria (HNA and LNA bacteria, respectively) and heterotrophic nanoflagellates (HNF), were measured by flow cytometry as previously described [27]. Bacterial production was estimated by the 3H-thymidine incorporation method as previously described [28].



Fig. 1 Study area of the central Adriatic Sea with sampling stations ST101, CJ007 and CJ009

Aerobic anoxygenic phototrophs (AAPs) abundance

Seawater samples for epifluorescent microscopy were collected using a Niskin bottle (5 L). After immediate fixation of 50 mL aliquotes with formaldehyde (pH 7.5, final concentration (f. c.) 2%) for 1 h at room temperature or overnight at 4 °C, samples were filtered on white 0.2 µm polycarbonate filters (47 mm diameter, Whatman® Nuclepore™ Track-Etched, Merck) and, after drying, stained with 4',6-diamidino-2-phenylindole (DAPI, f. c. 1 µg mL⁻¹) using a 3:1 mixture of Citifluor™ AF1 and Vectashield® [29]. AAP bacteria were counted using an Olympus BX51 microscope equipped with an Olympus UPlanSApo 100×/1.40 OIL, infra-red (IR) objective, a U-LH100H6 Hg lamp for excitation and image analysis software (CellSens). Images were taken with an XM10-IR camera (Olympus). Due to the rapid fading of the bacteriochlorophyll-*a* (Bchl *a*) autofluorescence, three epifluorescent filter sets were applied in a specific order: IR, DAPI and Chl *a*. The Chl *a* signal was subtracted from IR to obtain a net count of AAP cells.

Infrared epifluorescence microscopy and fluorescence in situ hybridisation (FISH-IR)

A combination of two epifluorescence-based methods, infrared epifluorescence microscopy and fluorescence

in situ hybridisation (FISH-IR) was applied to simultaneously detect infrared and probe signals, as described previously [21, 29, 30]. Probes targeting Alphaproteobacteria (probe ALF968), Gammaproteobacteria (probe GAM42a and unlabelled Bet42a competitor) and the *Roseobacter* clade (probe ROS537) were used. For detailed protocol, please see Additional file 2: Methods.

DNA extraction, *pufM* amplification, and Illumina amplicon sequencing

Seawater for DNA analyses was collected using a Niskin bottle (5 L), pre-filtered through a 20-µm plankton net and 1–2 L were immediately vacuum filtered on board through 0.22-µm polyethersulfone membrane filters (PES, 47 mm diameter, FiltraTECH, France). One liter of Milli-Q water represented a negative filtration control. Filters were frozen in liquid nitrogen and stored at –80 °C until further analyses. A modified DNeasy PowerWater kit (QIAGEN, The Netherlands) with an enhanced bead-beating step was used for DNA extraction [31]. Negative control (extraction blank) included empty 0.22-µm polyethersulfone membrane filter (PES, 47 mm diameter, FiltraTECH, France). Briefly, PES filters were cut in half with a sterilised scalpel and cut into smaller pieces, placed in 1.5 mL tubes filled with ceramic beads (MagNA

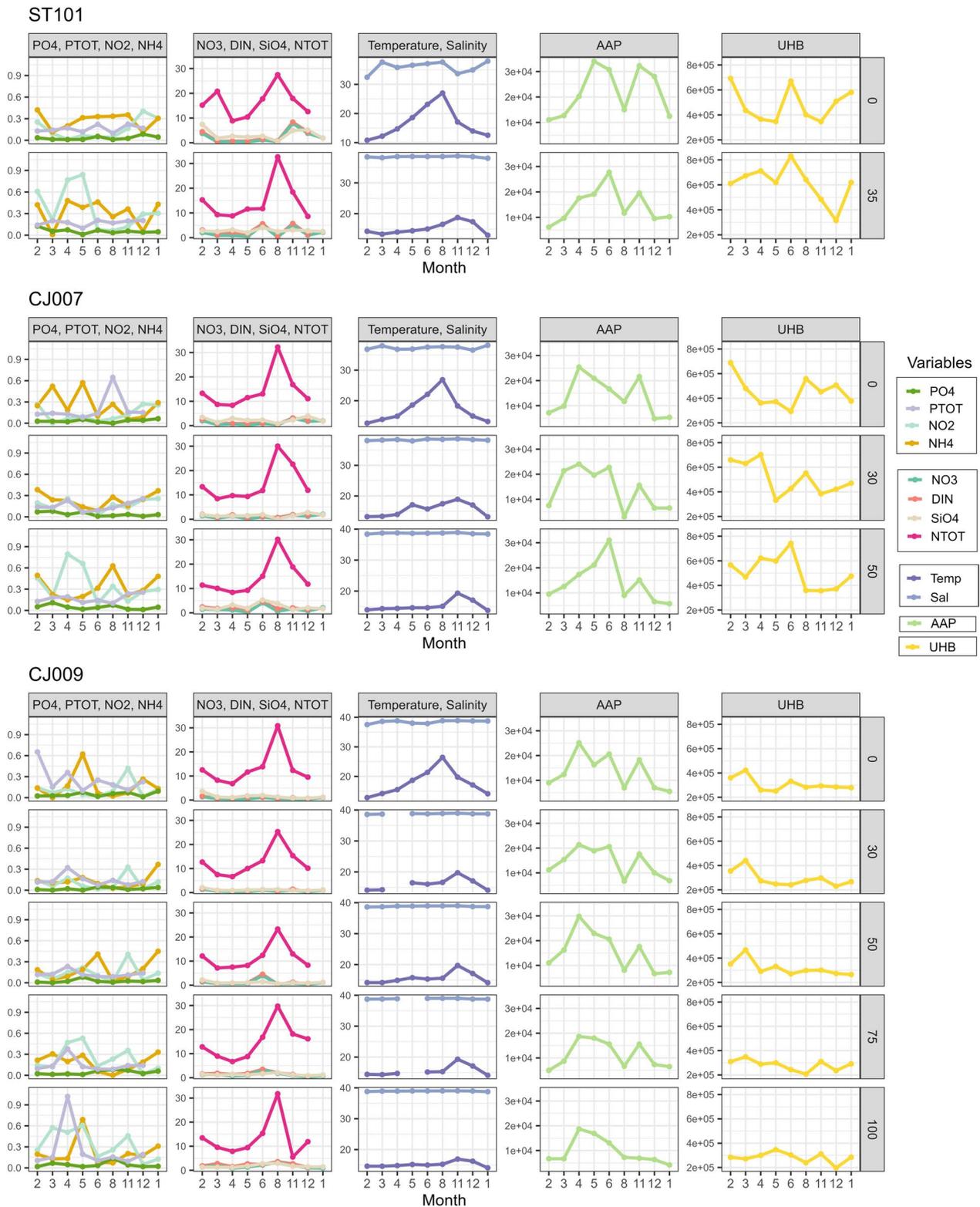


Fig. 2 Abiotic variables (temperature-Temp, salinity-Sal, nitrate- NO_3^- , nitrite- NO_2^- , ammonium ion- NH_4^+ , dissolved inorganic nitrogen-DIN, total nitrogen-NTOT, orthophosphate- PO_4^{3-} , total phosphorus-PTOT, silicate- SiO_4^{2-}) with absolute abundances of total heterotrophic bacteria (UHB) and aerobic anoxygenic phototrophs (AAPs) shown per station, month and depth

Lyser Green Beads, Roche, Switzerland), followed by rigorous homogenisation with the MagNALyser instrument, twice for 20 s at 9000 RCF. After homogenisation, the tubes were centrifuged at 5600 RCF for 1 min. The supernatant was transferred to a clean collection tube and the protocol was performed according to the manufacturer's instructions from step 8 in the Quick Start Protocol. The DNA was finally eluted in 35 μ L of the EB solution. Total extracted DNA was quantified and A260/A280 and A260/A230 absorbance ratios were measured using DS -11 spectrophotometer (Denovix, USA). Negative filtration and extraction controls showed DNA concentrations below the limit of detection. To analyse the composition of the AAP community, amplification of *pufM* gene (~204 bp) was performed with *pufM*_UniF (5'-GGNAAYTNTWYTAYAAAYCCNTTYCA) and *pufM*_WAW (5'-AYNGCRAACCACCANGCCCA) primers [16, 32, 33]. All samples were amplified in triplicate, with 25 μ L of reaction mix each containing 12.5 μ L of Q5[®] High-Fidelity 2X Master Mix (New England Biolabs, USA), 1.25 μ L of each primer at a final concentration of 0.5 μ M, 2 μ L of DNA template (10 ng/ μ L, final mass of DNA template in PCR reaction 20 ng) and 8 μ L of sterile, nuclease-free water. Cycling conditions were as follows: initial denaturation at 98 °C for 30 s, followed by 35 cycles of amplification at 98 °C for 7 s, 58 °C for 30 s and 72 °C for 30 s, with 2 min of final extension at 72 °C (T100 thermal cycler, Biorad, USA). The triplicates were purified from the agarose gel (2.0%) and pooled using the Wizzard SV Gel and PCR Clean-Up System (Promega, USA) according to the manufacturer's instructions and then quantified using the DS -11 spectrophotometer (Denovix, USA). All filtration/extraction blanks and PCR non-template controls were negative. All DNA extractions and PCR amplifications were performed in the same laboratory by the same person. Library preparation and amplicon pair-end sequencing (2 \times 250 bp) on the Illumina MiSeq were performed by the Genomics Core Facility of the Universitat Pompeu Fabra, Barcelona, Spain.

Bioinformatics, phylogenetic placement and calculation of diversity and evenness

A total of 7,823,083 input reads from Illumina Miseq were used for the bioinformatic analyses. The initial counts per sample are given in Additional file 1: Table S2. The quality of raw forward and reverse reads was assessed with FastQC v0.11.9. Primers were trimmed using cutadapt v4.1 [34]. Subsequent data processing was performed with the statistical software R v4.0.2 (<https://cran.r-project.org/>) using the R package DADA2 v1.16.0 [35]. Briefly, *filterAndTrim* function (`truncLen=c(200, 200)`, `maxN=0`, `maxEE=c(2,`

`2)`, `truncQ=2`, `rm.phix=TRUE`) removed low quality sequences and sequence tails. After error learning and sample inference algorithm, the paired-end reads were merged, resulting in the amplicon sequence variant (ASV) table. The *removeBimeraDenovo* function with "pooled" method discarded chimeras that accounted for 5.8% of the merged sequence reads. The naive Bayesian classification method [36] was used to assign taxonomy with the *assignTaxonomy* function based on the manually curated database, the most complete for AAPs to date [6].

Due to the high proportion of reads unclassified at genus level in the taxonomic assignment using DADA2, as high as 50% in some samples, a phylogenetic analysis was performed. ASVs occurring less than 2 times in at least 5% of samples and samples with an unacceptably low final number of reads per sample ($N < 2000$) were excluded from further analysis using the R package phyloseq v1.32.0 [37], resulting in 663 unique ASVs (number of reads per sample: median 58,227, min 2,039, max 32,5210) in 81 samples. The 663 ASVs were aligned using MAFFT [38] and 2 ASVs that poorly aligned were removed and further analysis was performed using 661 ASVs. A database of 3363 *pufM* sequences and their taxonomic assignment [6] was used for taxonomic assignment of the ASVs. ASVs were placed in the *pufM* gene phylogenetic tree [6] using the Evolutionary Placement Algorithm v0.3.5 [39] and Gappa [40] handled the taxonomic assignment of ASVs according to their position in the phylogenetic *pufM* tree with a consensus threshold higher than 50%. Sample data, untransformed final ASV matrix, taxonomy and reference *pufM* sequences (refseq) are given in Additional file 3. Since most of the initial *pufM* phylogroups [41] have already been assigned to a standard phylogenetic taxonomy [19], just the Rhodobacterales ASVs which could potentially belong to distinct *pufM* phylogroups E, F or G were aligned with phylogroup sequences using MAFFT [38]. A phylogenetic tree was built using FastTree [42] to obtain phylogroup affiliation or to facilitate the comparison with previous publications that used phylogroup taxonomic affiliation [5, 19, 41].

The R package ggplot2 v3.3.5 [43] was used to graphically visualize the composition of the AAP community in terms of relative abundances of a given taxon at class, order and genus levels. For agglomeration of ASVs to a certain taxonomic rank (order or genus); function *tax_glom* from R package phyloseq v1.32.0 was used. Since microbiome count data is of a compositional nature and should be treated as such [20, 44, 45], centered log-ratio (CLR) transformation [46] was performed using function *transform* from R package microbiome v1.10.0 [47], with introduced pseudo-counts (minimum relative abundance

divided by two) instead of zeros in ASV table before the transformation.

In order to account for significant discrepancies in read counts between samples, rarefying (*i.e.* random subsample of reads to equalise read depth) was used exclusively to estimate the diversity metrics. Rarefying to the smallest library size multiple times (function *rarefy_even_depth* from phyloseq v1.32.0 R package, rarefying threshold=2000, N(times)=100) was applied on the ASV matrix to estimate alpha diversity metrics calculated as mean values of multiple subsample calculations: the observed number of ASVs, Shannon's (H') and Pielou's (J') indices calculated with vegan v2.5.7 R package [48]. Rarefaction curve exhibiting sufficient sequencing depth for diversity estimates is given in Additional file 1: Fig. S1.

Statistical analyses and neural gas algorithms

Correlations between abiotic environmental variables were assessed using correlational (Draftsman) plots in PRIMER7 to estimate if a certain variable should be excluded in the case of a strong correlation (Additional file 1: Fig. S2). No variables were excluded from further analysis.

To assess whether there are differences in the spatiotemporal patterns of the AAP community, permutational multivariate analysis of variance (PERMANOVA) based on Aitchison distances (Euclidean distances on CLR transformed dataset) was performed on the *pufM* dataset in PRIMER7 software, with 'season' and 'region' as fixed factors and 'layer' as a random factor nested in 'region' (9999 permutations, sums of squares type: type II (conditional), permutation method: Unrestricted permutation of raw data) [49, 50]. Factor layer was defined as L1 (0–30 m), L2 (30–50 m), L3 (50–75 m) and L4 (75–100 m). Accompanying PERMANOVA, distance-based test for homogeneity of multivariate dispersions (PERMDISP) was performed on the same transformed dataset (9999 permutations, deviations from centroid). A significant result of PERMDISP would indicate that groups differ in dispersion [51]. Variance-based compositional principal component (PCA) biplot on Aitchison distances was generated based on genus-level agglomerated and CLR-transformed values with zero replacement using pseudo-counts with the R package microViz v0.10 [52].

AAP absolute and relative abundance data as well as FISH-IR microscopic counts given as relative abundances were square root transformed. Non-metric multidimensional scaling (nMDS) ordination plot based on Bray–Curtis distances was constructed to visualise differences in seasonal abundances for FISH-IR groups, followed by PERMANOVA to test the significance of observed differences (9999 permutations,

sums-of-squares type: type II (conditional), permutation method: Unrestricted permutation of raw data). PERMDISP (9999 permutations, deviations from centroid) was performed accompanying PERMANOVA on the same transformed dataset. To assess agreement between metabarcoding and FISH-IR data, Spearman correlation coefficients and their statistical significance were calculated between FISH-IR probe counts and corresponding pseudoabundances from *pufM* sequencing (*i.e.* relative abundance of Gammaproteobacteria class \times AAP absolute number/100) using *cor.test* function from R package stats v.3.6.2 (Additional file 4).

An artificial neural network/unsupervised topology learning algorithm, neural gas (NG), was applied to a CLR-transformed *pufM* dataset to estimate characteristic AAP patterns associated with specific biotic/abiotic environmental factors [53]. NG has been successfully used for this type of data in previous studies due to its suitability for modelling anomalies and the mean distribution of microbiological parameters [25, 27, 54]. In our study, separate NG models were created on agglomerated and CLR-transformed datasets on *pufM* sequencing data, carried out at the order-, genus- and ASV-levels. NG models used the *pufM* sequencing data for quantifying data space to generate "best-match units" (BMUs). For all models, the ecological factors (biotic: heterotrophic bacteria-UHB, High nucleic acid bacteria-HIGH, *Synechococcus*-SYN, *Prochlorococcus*-PROCHL, picoeukaryotes-PE, bacterial production-BP, heterotrophic nanoflagellates-HNF, aerobic anoxygenic phototrophs-AAP; and abiotic: temperature-Temp, salinity-Sal, nitrates- NO_3^- , nitrites- NO_2^- , ammonium ion- NH_4^+ , dissolved inorganic nitrogen-DIN, total nitrogen-NTOT, soluble reactive phosphorus-SRP, total phosphorus-PTOT, silicate- SiO_4^{2-} , Chlorophyll *a*-Chl *a*) were calculated as average values for a specific BMU. All models resulted in five characteristic distributions (BMUs). The models were initialised by setting the number of training epochs to 1000, the initial step size to 0.5 and the initial decay constant to 4.5 using SOM Toolbox version 2.0 for MATLAB (E. Alhoniemi, J. Himberg, J. Parhankangas and J. Vesanto, Helsinki University of Technology, Finland: <http://www.cis.hut.fi/projects/somtoolbox>).

Heatmap presentation of taxa in each BMU was generated using conditional formatting in Microsoft Excel v.16.0: values of environmental variables were coloured according to their average values for every BMU, and taxa according to their average CLR-transformed values. Further, *heatmap.2* function of gplots v3.1.3 R package was used to construct heatmaps representing community composition based on Aitchison distances and Ward.D2 dendrogram agglomeration method [55].

Results

Abundance of AAP bacteria

AAPs were observed in all 90 samples (Fig. 2) and their mean absolute abundance in the study area was $1.43 \pm 0.75 \times 10^4$ cells mL⁻¹. The lowest value was recorded at station CJ007 in August at a depth of 30 m (0.30×10^4 cells mL⁻¹), while the highest value was observed at station ST101 in May at the sea surface (3.41×10^4 cells mL⁻¹). The average relative abundance was $3.86\% \pm 2.27\%$ (minimum 0.55% at station CJ007 in August at 30 m depth; maximum 10.26% at CJ009 in April at 50 m depth). At the monthly level, the highest mean abundance was recorded in June with $2.19 \pm 0.62 \times 10^4$ cells mL⁻¹, followed by almost identical values in April with $2.18 \pm 0.41 \times 10^4$ cells mL⁻¹, while the lowest mean abundance was observed in January 2022 with $0.7 \pm 0.25 \times 10^4$ cells mL⁻¹ and in February 2021 with $0.84 \pm 0.22 \times 10^4$ cells mL⁻¹.

Absolute and relative abundances of AAPs significantly differed on a seasonal scale (PERMANOVA pseudo-F=23.13, $p=0.0001$, unique permutations=9957; pseudo-F=27.98, $p=0.0001$, unique permutations=9948 respectively), but not with respect to region or depth (Additional file 1: Table S3A and S3B respectively), although a decreasing trend was observed towards the open sea (Fig. 3). The pairwise comparisons showed that spring was the only season different from others (Additional file 1: Table S3). However, seasonal absolute abundances also differed in dispersion (PERMDISP test, group factor: Season, $F=5.05$, $p=0.0067$, 9999 permutations),

while relative abundances did not (PERMDISP test, group factor: Season, $F=1.51$, $p=0.27$), suggesting differences also stem from larger variances between samples.

FISH-IR

To quantify which proportion of AAPs belong to the Alphaproteobacteria and Gammaproteobacteria classes or to the *Roseobacter* clade, a method based on epifluorescence microscopy, FISH-IR, was used. In the study area, gammaproteobacterial AAPs had a mean relative abundance of 35.25%, Alphaproteobacteria 31.15% and *Roseobacter* 37.66%. Complete results are presented and visualized in Additional file 2. During data acquisition and processing, several limitations of the ALF968 probe, targeting all Alphaproteobacteria were observed: percentage of AAP cells hybridized with the ROS537 probe (*Roseobacter* clade within Alphaproteobacteria) generally exceeded the number of cells detected with the ALF968 probe, and a total sum of all probes was often greater than 100%. Therefore, in silico coverage and specificity of ALF968 and ROS537 probes for the order Rhodobacterales were estimated against SILVA database. The ALF968 had larger coverage than the ROS537 probe (96.3% vs. 92.9%), but showed reduced specificity (92.96% vs. 99.32% respectively), which might explain this phenomenon (Additional file 5). After correlating FISH-IR and *pufM* metabarcoding data, we observed strong positive correlation ($r=0.745$) for Gammaproteobacteria, moderate positive correlation ($r=0.651$) for Alphaproteobacteria and strong positive correlation ($r=0.717$) for

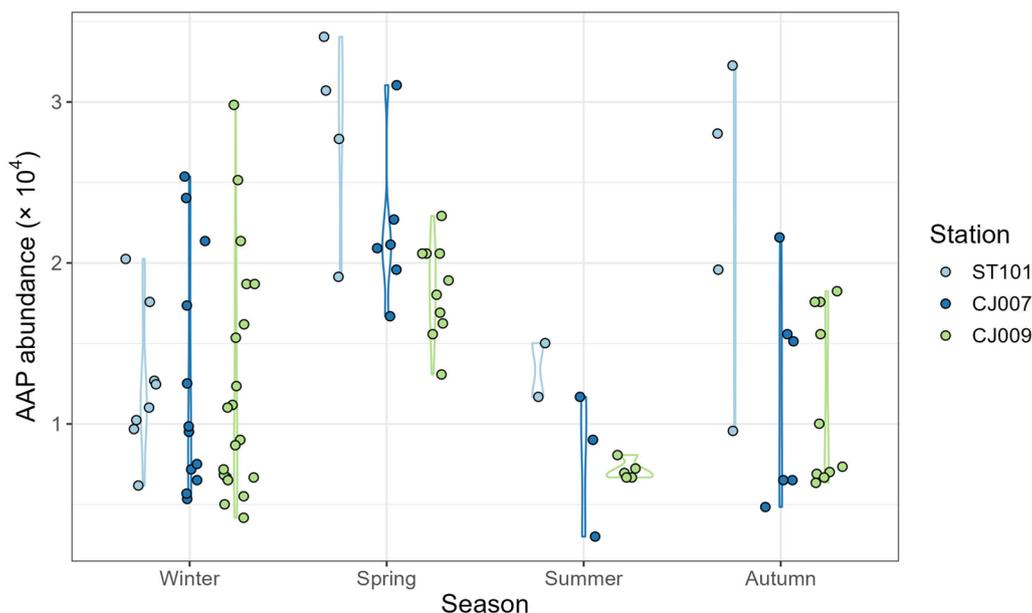


Fig. 3 Absolute abundances of AAPs shown per station (ST101, CJ007 and CJ009) and season

Roseobacter, indicating consistency in the relative abundance patterns between these two methods. All correlations were statistically significant ($p < 0.001$) (Additional file 4). We conclude that results from FISH-IR probes Gam42a and ROS537 targeting class Gammaproteobacteria and *Roseobacter* clade respectively can be trusted and contribute additional value to data quantitative interpretation.

AAPs belonging to *Roseobacter* clade showed the highest relative abundance (77.77%) at station ST101 at 35 m in August and the lowest one (14.29%) at CJ007 station at 50 m in August. The highest average relative abundances were recorded in autumn and the lowest average ones in summer, except for ST101 station (Additional file 2: Figs. S2 and S3). Gammaproteobacteria AAPs had the highest relative abundances in samples collected from the surface (0 m) of open sea station CJ009 in February (78.9%), followed by samples from station CJ007 at 30 m depth in August (75%) and station ST101 at 35 m depth also in August (71.4%). The lowest value (8.3%) was measured at the sea surface in January at station CJ009. In terms of seasonal distribution, they were on average predominant in summer at all stations, with the lowest contribution in winter (Additional file 2: Figs. S2 and S3).

AAP community composition on a spatiotemporal scale obtained via metabarcoding

When considering the composition of the AAP community obtained from the *pufM* sequencing dataset agglomerated at the genus level, statistically significant seasonal differences were observed (PERMANOVA, pseudo-F = 2.06, $p = 0.0037$, unique permutations = 9890, Additional file 1: Table S4). Similarly to total AAP abundances, there were no differences between spatial or vertical profiles. Pairwise comparisons showed that summer differed from other seasons (Additional file 1: Table S4). There were no differences in dispersion between seasons (PERMDISP, $F = 1.89$, $p = 0.2035$, 9999 permutations). Slight separation of summer samples from others was observed in PCA biplot, with possible subdivision within the group (Fig. 4). Taxa contributing most to separation of summer from other seasons (with higher contribution during summer) were *Puniceibacterium* genus, UBA868, Xanthobacteraceae and Maricaulaceae family (Fig. 4).

Phylum Proteobacteria dominated in all samples. Interestingly, the phylum Gemmatimonadota, family Gemmatimonadaceae, represented by two ASVs (ASV550 and ASV559) unclassified at genus level, was observed in a very low relative abundance (mean 0.002%) in a sample collected in April at 30 m depth at station CJ007, as well as in few others in a very low number of reads. For details, please see Additional file 3 (sheets "ASV_matrix" and "Taxonomy").

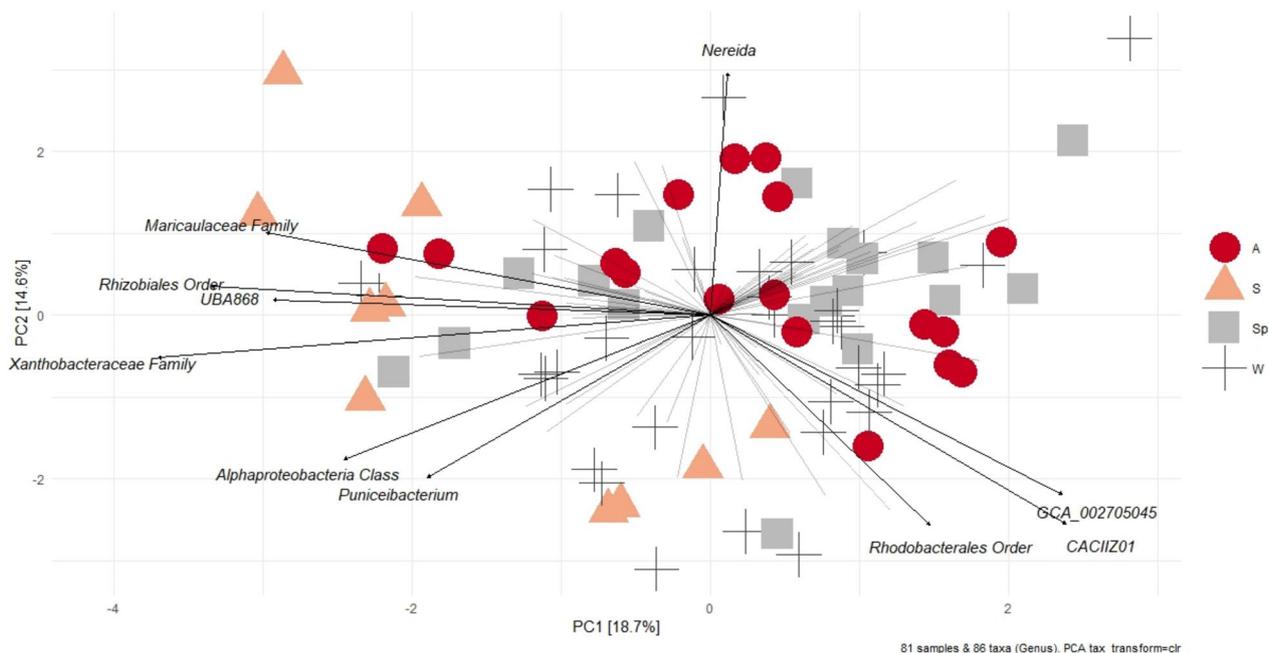


Fig. 4 Variance-based compositional principal component (PCA) biplot on Aitchison distances of *pufM* dataset (agglomerated at genus-level and CLR transformed) showing groupings by season (W-Winter, Sp-Spring, S-Summer, A-Autumn) where names of top 10 taxa by the longest loading vector length are indicated

At the class level, the AAP community composition was dominated by Gammaproteobacteria in all samples (75.9% on average), with the highest values in December at stations ST101 and CJ007 and in November at station CJ009. In contrast, the lowest relative contribution was recorded in February at the sea surface at station ST101, when their relative abundance dropped <40%, and in March, also at the sea surface at station CJ007 with relative abundances of ~50% (Additional file 1: Fig. S3). The alphaproteobacterial AAPs had an average relative abundance across all samples of 24.1% with the highest contribution to the total number of reads among the sea surface samples in February at ST101 (~60%) and in March at CJ007 (~40%).

Looking at AAP orders, the gammaproteobacterial Pseudomonadales dominated at all stations and in all months during the sampling period (Fig. 5), except for the lowest contribution in February at the sea surface at station ST101 (<40%) and in March, also at the sea surface at station CJ007 (~50%). Burkholderiales had higher relative abundance (~10%) in February at 0 m at station ST101 and in May at dChlMax (non-standard oceanographic depths, 68 m) at station CJ009. Order Rhodobacterales (family Rhodobacteraceae), which represented 96% of Alphaproteobacteria, was also omnipresent at all stations and in all months, with the highest relative contribution recorded in February at the sea surface at station ST101 (~60%) and in March, also at the sea surface at station CJ007 (~40%). In contrast, the

alphaproteobacterial orders Rhizobiales, Sphingomonadales and Caulobacterales occurred occasionally in relative abundances >1% (Fig. 5).

Numerous AAP genera (71 of total 86) occurred in relative abundances of less than 1%, while the 15 most prevalent ones accounted for 99% of the sequencing data (Fig. 6). In general, genus *Luminiphilus* (phylogroup K, order Pseudomonadales, family Halielaceae) dominated the AAP community, with relative abundance exceeding 90% in some samples. In total, 146 different ASVs were recorded for this genus, with different relative contribution at specific stations, depths and months. Details about the distribution of *Luminiphilus* ASVs are given in Additional file 6. *Limnohabitans* was detected in abundance >1% at sea surface station ST101 in February, but not towards the open sea. The Alphaproteobacteria *CACIIZ01* and Rhodobacteraceae *Puniceibacterium* were not recorded at costal station ST101, and sporadically on other stations (Fig. 6). The genus *Planktomarina* showed the highest relative abundance (~20%) at station ST101 in February at the sea-surface and higher values in November regardless of depth at stations ST101 and CJ007. Genus *MED_G52* of the Rhodobacteraceae family was present in abundance up to 10% at all stations, in all seasons and depths, except in December at stations ST101 and CJ007 (present in >1% only at 0 m) and in November at CJ009 (<1%).

The lowest average number of observed ASVs was recorded in winter (124.34 ± 50.9), while the highest

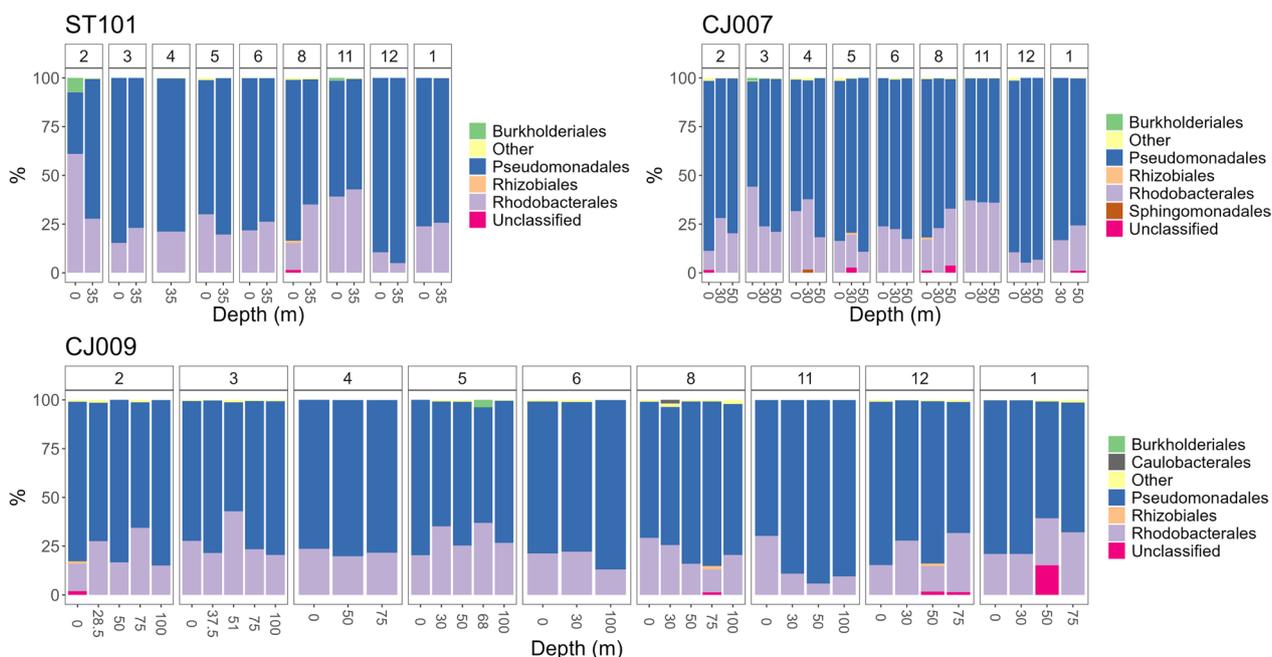


Fig. 5 AAP orders detected in the study area via *puFM* metabarcoding shown per station (ST101, CJ007 and CJ009), month and depth

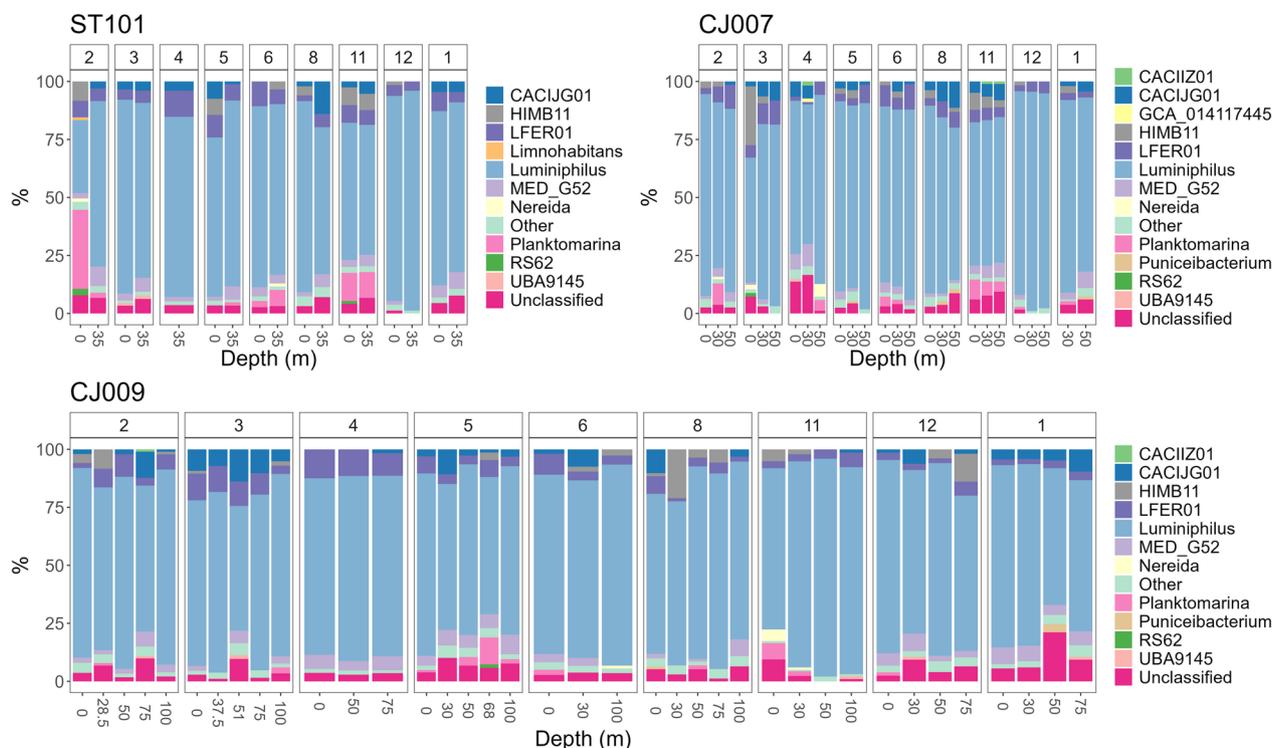


Fig. 6 AAP community composition via *pufM* metabarcoding in the study area at the genus level shown per station (ST101, CJ007, CJ009), month and depth

one was recorded in autumn (132.6 ± 56.38). The average Shannon diversity indices were highest in summer (3.5 ± 0.44) and lowest in autumn (3.39 ± 0.68), while the average values for winter and spring months were in between (3.45 ± 0.39 and 3.47 ± 0.35 , respectively). The average Pielou's evenness showed the same trend as Shannon indices, with the highest average values in summer (0.73 ± 0.05) and the lowest in autumn (0.70 ± 0.08) (Additional file 1: Table S5).

Neural gas analyses of *pufM* gene metabarcoding data

To gain a deeper insight into AAP community composition at a finer taxonomic resolution and their structuring in the light of biotic and environmental parameters, neural gas (NG) analysis, an artificial neural network algorithm robust to outliers, was applied to *pufM* sequencing dataset at order-, genus- and ASV-levels. Agglomeration of dataset to higher taxonomic rank (order) was necessary for comparison to previous literature, and the rationale behind running multiple analyses on lower taxonomic ranks was to reveal if a particular genus within order, or ASVs of certain genera have consistent overall behaviour (relative abundance) in a certain environment or not. Another rationale behind running the order-level neural network was that a lot of AAP genera, even with the most complete taxonomic database up to date, are

still unclassified (e.g. genera of Rhizobiales order). Additionally, identification at species level was not possible with methodology used here, so there is an important taxonomic link missing between genus and ASVs, challenging the interpretation of this model. Hence, we focus on order- and genus- level data.

AAPs agglomerated at the order level were clustered into five distinctive BMUs according to their relative contribution to community (Fig. 7, Additional file 7: Fig. S1, Additional file 8). Detailed description of each BMU unit is given in Additional file 7. BMUs generally described most nutrient-enriched (BMU1), warmest and shallowest (BMU2), lowest ammonia (BMU3), maximum nitrite (BMU4) and coldest, nutrient-depleted, deepest and most saline (BMU5) environment (Fig. 7). In all units, two orders were omnipresent and showed consistently high contribution regardless of environmental conditions: Pseudomonadales and Rhodobacterales. In contrast, orders Acetobacterales, UBA8317, Gemmatimonadales, Xanthomonadales and Thalassobaculales occurred transiently in higher, but mostly in very low values (Fig. 7).

Based on the *pufM* dataset agglomerated at genus level, five distinct BMUs were also formed covering a total of 86 unique genera (Fig. 8, Additional file 7: Fig. S2 and Additional file 9). Detailed descriptions for each

	BMU1	BMU2	BMU3	BMU4	BMU5
UHB	0.4522	0.4096	0.4079	0.4157	0.3121
HIGH	25.343	29.288	28.892	25.472	26.529
SYN	19.526	19.249	15.154	15.651	11.931
PROCHL	1.6171	1.4425	1.1958	118.79	0.2671
PE	5.4614	3.3938	3.195	8.9533	2.71
BP	0.1274	0.128	0.1168	0.1565	0.2385
HNF	0.7497	0.6049	0.5127	0.5195	0.4681
AAP	1.6645	1.4277	1.0538	1.5327	0.852
Temp	16.439	16.628	16.178	15.371	15.017
Sal	37.967	38.191	38.457	38.357	38.71
NO ₃ ⁻	1.7343	0.8821	1.017	1.4039	0.9068
NO ₂ ⁻	0.222	0.1695	0.174	0.2806	0.1557
NH ₄ ⁺	0.2802	0.1933	0.1496	0.2102	0.3534
DIN	2.2685	1.2449	1.3407	1.8499	1.0243
PO ₄ ³⁻	0.0425	0.0373	0.0344	0.0357	0.0378
PTOT	0.1556	0.1773	0.1599	0.1873	0.1263
SiO ₄ ²⁻	2.1471	1.7983	1.5683	2.3332	1.5799
Chl <i>a</i>	0.3336	0.35	0.2733	0.335	0.2514
N/P	53.418	33.379	38.928	51.846	27.118
Alphaproteobacteria (%)	28.835	33.585	37.47	26.637	34.947
Gammaproteobacteria (%)	31.18	36.226	42.314	31.701	41.578
<i>Roseobacter</i> (%)	39.882	36.226	27.681	38.34	39.527
	BMU1	BMU2	BMU3	BMU4	BMU5
Pseudomonadales	9.3592	8.0675	8.786	10.43	8.6969
Rhodobacterales	8.4198	6.7861	7.2894	8.6113	7.8996
Unclassified (Alphaproteobacteria)	1.2974	2.2351	3.1981	-1.1265	4.9998
Rhizobiales	-0.1422	2.668	3.2207	-0.6277	1.4436
Burkholderiales	3.469	1.2635	-2.9749	0.1449	1.9003
Sphingomonadales	-0.7269	0.2067	0.3326	-1.3543	-0.9239
Caulobacterales	-2.3245	1.4653	1.7458	-1.7381	-2.8427
Arenicellales	-1.543	0.2733	0.5741	-1.711	-1.2847
SP197	0.2661	-0.5334	-2.297	-1.0703	-1.6652
UBA8366	-2.51	-2.2166	-1.9464	-1.64	2.2497
Acetobacterales	-1.6929	-3.004	-3.3936	-0.9584	-3.4122
UBA8317	-2.8029	-3.6686	-2.272	-1.9278	-3.4122
Gemmatimonadales	-2.7578	-2.9368	-3.5081	-1.4468	-3.4122
Unclassified (Proteobacteria)	-2.7055	-3.3779	-2.6793	-1.9278	-3.4122
Xanthomonadales	-2.8029	-3.6213	-3.0015	-1.73	-3.4122
Thalassobaculales	-2.8029	-3.6069	-3.0739	-1.9278	-3.4122

Fig. 7 Neural gas analysis of *puflM* dataset agglomerated at the order level clustered into five distinctive BMUs according to their relative contribution to AAP community (expressed as average CLR-transformed values) and connected average values of biotic (UHB, HIGH, SYN, PROCHL, PE, BP, HNF, AAP, percentages of FISH-IR groups) and abiotic (Temp, Sal, NO₃⁻, NO₂⁻, NH₄⁺, DIN, PO₄³⁻, PTOT, SiO₄²⁻, Chl *a*, N/P) variables of the environment. Colour gradient from red to green represents the lowest and highest average values respectively

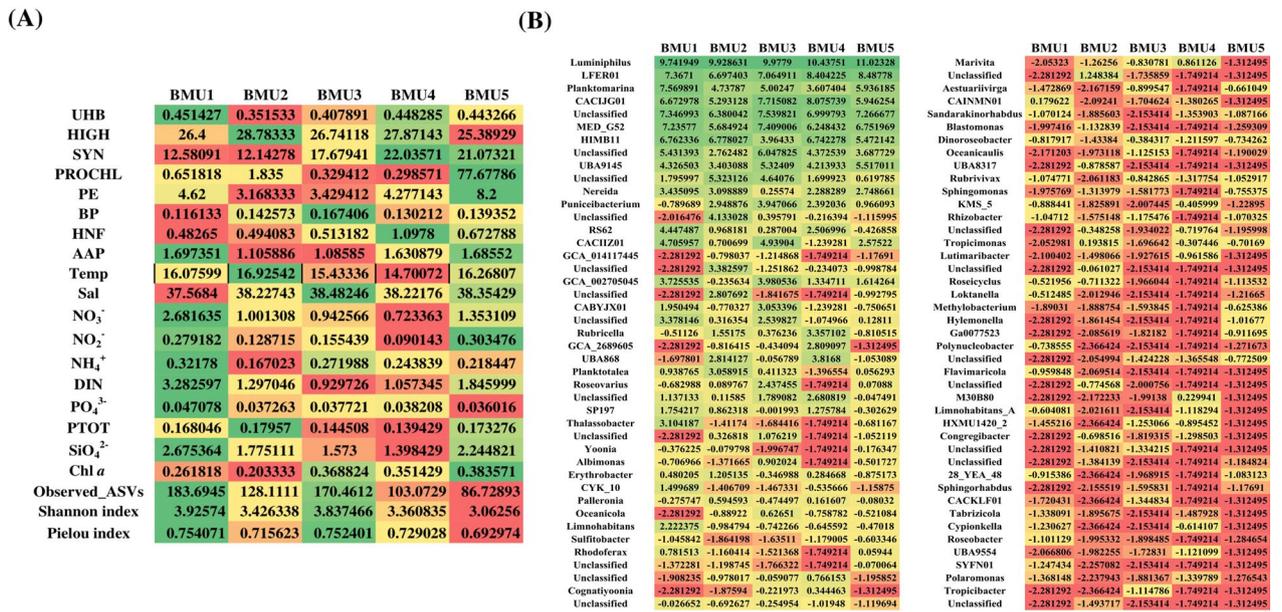


Fig. 8 Neural gas analysis results of *pufM* dataset agglomerated at the genus-level and CLR-transformed and CLR-transformed, clustered into five BMUs. Biotic (UHB, HIGH, SYN, PROCHL, PE, BP, HNF, AAP, diversity metrics) and abiotic (Temp, Sal, NO₃⁻, NO₂⁻, NH₄⁺, DIN, PO₄³⁻, Chl a, N/P) variables of the environment are given in (A) as average value for each unit. Relative contribution of specific genera to AAP community (expressed as average CLR-transformed values) is shown in (B). Colour gradient from red to green represents the lowest and highest average values respectively

BMU are given in Additional file 7. Genera *Luminiphilus* (family Halieaceae, order Pseudomonadales, class Gammaproteobacteria), the uncultured *LFER01* as well as *Planktomarina* and *CACIJG01* (family Rhodobacteraceae, order Rhodobacterales, class Alphaproteobacteria) had almost the same high average values across all units. In nutrient-enriched and low-salinity habitat (BMU1), higher representation of *Nereida*, *CYK10*, *Thalassobacter*, and the Gammaproteobacteria *RS62*, *Rhodofera* and *Limnohabitans* was noted, coupled with the highest abundance of AAPs, heterotrophic bacteria and picoeukaryotes (Fig. 8). Genera *Erythrobacter* (order Sphingomonadales), *Planktotalea* and *Palleronia* (Rhodobacteraceae) clustered with the lowest abundances of total heterotrophic bacteria, *Synechococcus*, picoeukaryotes and low abundances of AAPs as well as HNF, preferring the warmest habitat with low ammonia, Chl a, nitrate and nitrite (BMU2). The most saline yet diverse habitat scarce with nitrate and dissolved inorganic nitrogen, where AAP abundance was lowest, was dominated by the highest value of genera *Roseovarius*, *Punicibacterium* and *CACIZ01* (Rhodobacteraceae) and the gammaproteobacterial *CABYX01* (Halieaceae, Pseudomonadales), indicating these genera do not appear to be controlled by a low-N environment (BMU3). The shallowest unit with a clear spatiotemporal pattern (almost all winter samples from transitional/open sea stations) with the lowest average temperature, scarce with

nitrate, silicate, absolute abundances of *Prochlorococcus* but abundant with *Synechococcus*, HNF and AAPs were characterised by higher incidence of *CACIJG01*, *Rubricella*, *GCA2689605* (Alphaproteobacteria, family Hyphomicrobiaceae) and *UBA868* (Arenicellales), indicating their preference for colder environments (BMU4). Genera *Roseobacter* and *Sphingomonas*, rarely found in other units, preferred nutrient-rich habitat (BMU5), characterised by very high average salinity and temperature above 16 °C, rich in nitrite, silicate and Chl a, with maximum absolute abundances of *Prochlorococcus*, picoeukaryotes, high abundances of *Synechococcus* and AAPs, but with the lowest diversity (Fig. 8). In respect to order-level model, it was noticed that the most dominant genera inside a particular order (e.g. genus *Luminiphilus* of order Pseudomonadales and *LFER01* of order Rhodobacterales) masked rare ones, which did not follow general behaviour of the order. Examples include *Nereida* and *Punicibacterium* of order Rhodobacterales which were inconsistently present in high values across units. Another example is rare genus *Rubrivivax* which expressed different behaviour from Burkholderiales order it belongs to (Figs. 7 and 8).

ASV-level NG model based on an entire ASV dataset (661 ASVs) also revealed five distinct units (BMUs). ASVs of the same genus exhibited different behaviours, but it was difficult to describe each specific variant of each genus. The most prominent variants of major

genera that showed specific behaviour in certain BMU are shown in Additional file 7: Fig. S3. What can be observed is that in this model, we generally perceive more variation in respect to order- and genus-level models, *i.e.* certain ASVs appear only in some BMUs and are missing from others, which was masked at higher taxonomic levels.

Discussion

Since AAPs represent a fascinating group of ubiquitous photoheterotrophs important for carbon cycling, they have been extensively studied in various environments in recent decades, especially in aquatic ones [10]. Previous research in the Adriatic Sea focused on their abundances estimated from Bchl *a* concentrations along a latitudinal transect [56], and on counts from coastal and estuarine waters along the eastern Adriatic coast in summer using infrared epifluorescence microscopy [22]. Subsequent studies based on one-year sampling along the trophic gradient described their spatial and vertical distribution with maximum abundances reported in late winter (April) [23]. A recent review article collected seven years of data from 34 sites to gain insights into their distribution and controlling environmental factors in the Adriatic Sea [24].

However, a comprehensive analysis of AAP community structure in the Adriatic Sea on a fine taxonomic scale has not been yet described. In this study, we present their community composition during one year in the central Adriatic from three stations along the trophic gradient using *pufM* metabarcoding and quantitative FISH-IR approach. In addition, neural gas algorithms were applied to agglomerated at order- and genus-level sequencing dataset (CLR-transformed after agglomeration) and ASV-level to reveal patterns of the AAP community in respect to specific physicochemical and biological variables.

In the study area, significant differences in seasonality were observed in terms of AAP abundances (maximum average value in spring: $2.136 \pm 0.081 \times 10^4$ cells mL⁻¹, minimum in summer: 0.86×10^4 cells mL⁻¹), FISH-IR groups (*Roseobacter* prevalent in autumn, Alpha- and Gammaproteobacteria in summer) and *pufM* genus-agglomerated metabarcoding data. However, no clear spatial pattern emerged, as previously reported for the Mediterranean coastal lagoon [57]. The pronounced seasonality is consistent with the first long-term study based on *pufM* amplicon dataset from the northwestern Mediterranean, which was conducted over a decadal period and found that this community is highly seasonal, with certain taxa showing recurrent patterns [5].

AAP abundances

As for the absolute abundances, the average values of $1.427 \pm 0.754 \times 10^4$ cells mL⁻¹ correspond to those previously reported for the central Adriatic. However, the range of 0.30×10^4 to 3.41×10^4 cells mL⁻¹ was narrower in our study, compared to studies that sampled Adriatic estuaries [22–24]. The average relative abundance of AAPs was $3.862\% \pm 2.266\%$ of total prokaryotes (minimum of 0.548%, maximum of 10.258%), which is comparable to previous results from the Adriatic [22, 23] and the Mediterranean [4, 18, 57], as well as the Global Ocean Sampling Expedition [41]. As in the study by Ferrera et al. [18] where maximum abundances of AAPs were found in summer, a clear and significant seasonality was observed in our study. In contrast, we have detected the maximum average abundance in spring, the only season significantly different from the others in the pairwise comparison, and the minimum average abundances in summer. The maxima observed in April and June are comparable to a previous study in the Adriatic, where maximum AAP abundances were also reached in April at all stations, with increased phosphorus concentrations and lower N/P ratio as possible explanations [23]. Depth in the range from 0 to 100 m as well as dChlMax did not prove to be a significant environmental factor controlling AAP abundance in this study, suggesting that, contrary to previous results [23, 58], sea transparency of the central Adriatic was not a crucial factor affecting this community. Inconsistency in spatial and vertical distribution pattern in our study compared to previous research in Adriatic could be possibly explained by the trophic status of study area. Previous research covered more eutrophic coastal and estuarine area to the oligotrophic open sea [23], whilst in our study differences in trophic status were less enhanced and overall oligotrophic at all depths. Potential explanation for inconsistency in spatial distribution compared to previous results could be attributed to geographic proximity of stations studied in our research, while in previous ones larger study area was covered [22, 23].

Quantitative estimates from FISH-IR

Quantitative estimates of specific bacterial groups resulting from fluorescence in situ hybridisation combined with catalysed reporter deposition (CARD-FISH) have been widely reported in numerous ecological studies for years [14, 27, 59–61]. Nevertheless, quantitative estimates of AAPs taxonomically assigned to specific clades or groups are rare [21, 29], especially for environmental samples. In our study, FISH-IR was applied to quantify the percentage of AAPs assigned to either the Gammaproteobacteria class, the Alphaproteobacteria class or the *Roseobacter* clade. Overall, the AAP community

in the central Adriatic was quite heterogeneous according to these three probes, with the highest mean relative abundances of AAPs associated with *Roseobacter* clade, followed by gammaproteobacterial AAPs and alphaproteobacterial ones. A distinct seasonal pattern emerged, with AAPs assigned to *Roseobacter* clade generally less abundant in warmer months and more prevalent in autumn, while the patterns of general Alphaproteobacteria and Gammaproteobacteria were inverse to those of *Roseobacter*. The proposed term *Roseobacter* clade stands for a group of marine Rhodobacteraceae (members of the Rhodobacterales order that share 89% identity of 16S rRNA gene sequences), which are known to be ubiquitous in seawater and crucial for biogeochemical cycling of various elements [62]. Most of the cultured members have quite large genomes, high GC content (~60%) and a wide diversity of metabolic capabilities. Both free-living and/or associated with phytoplankton and microalgae, *Roseobacter*-like bacteria could account for up to 30% of bacterial communities in marine environments [27, 62, 63]. The lower contribution of AAPs assigned to this clade observed in summer contrasts with previous findings where the *Roseobacter* clade (represented as % of total bacteria, determined via CARD-FISH) had the highest average relative abundances in summer [27, 64]. However, the capability for aerobic anoxygenic phototrophy, recruited via horizontal gene transfer, is considered to be present in at least 14 phylogenetically distinct strains of roseobacters, but not in all *Roseobacter*-clade members [62, 65]. Therefore, we hypothesize that the seasonal behaviour of AAPs that belong to *Roseobacter* clade may potentially differ from other *Roseobacter*-clade members due to AAPs' advantage of photoheterotrophy (*i.e.* heterotrophic energy acquisition through oxidation of organic matter coupled with light utilization) [62]. Nevertheless, FISH-IR data in conjunction with the NG order-level model revealed that AAP roseobacters prefer a warm, less saline and nutrient-enriched environment (order-level BMU1, without distinct seasonality observed), associated with the highest absolute abundances of AAPs, suggesting that *Roseobacter*-like bacteria dominate the AAP community in the described environment. Another pattern significant for roseobacters was their preference for ammonia regardless of ambient temperature, with their lowest counts in BMU3 coinciding with the lowest ammonia concentrations and vice versa in an ammonia-enriched habitat (BMU5). This is to be expected as many organisms in the *Roseobacter* clade can utilise nitrogen exclusively in a reduced form, as they do not possess genes encoding enzymes for nitrate and nitrite reduction, but gene clusters encoding enzymes for the uptake of ammonium, amino acids, spermidine and urea [66]. The observed seasonal dynamics of Alphaproteobacteria

AAPs, characterised by the highest average relative abundances in summer and the lowest ones in autumn, is in contrast to previously reported seasonal behaviour of total Alphaproteobacteria estimated with CARD-FISH [27, 64, 67]. These results contradict the assumption that Alphaproteobacteria are generally good competitors in oligotrophic environments (predominant in winter months), suggesting that the behaviour of Alphaproteobacteria AAPs may differ from general bacterial community patterns, preferring eutrophic habitats despite their phototrophic capabilities [7, 10, 68]. In contrast, the peak abundances of gammaproteobacterial AAPs in summer are comparable to previous findings about general Gammaproteobacteria, indicating their preference for higher nutrient concentrations (mainly total nitrogen), assimilation of organic carbon sources and generally higher temperatures [27, 68].

When discussing these results, it is crucial to bear in mind the substantial limitations of FISH-IR method as well as the coverage and specificity of oligonucleotide probes. As mentioned in Results, percentage of AAP cells hybridized with the ROS537 probe was generally higher than the number of cells detected with the ALF968 probe, sometimes leading to a total sum of relative abundances greater than 100%. Similar results were also observed in other studies for same probes, but for CARD-FISH results [27, 69, 70]. Riou et al. [71] re-evaluated the *in silico* specificity of eleven bacterial and eukaryotic probes regularly used to enumerate marine microbes via CARD-FISH and found that both the Ros537 and Roseo536R probes identified 91% of the alphaproteobacterial *Roseobacter* clade, but also targeted 2.9% of sequences classified as non-*Roseobacter* (other members of the Rhodobacterales order). Similar result was observed in our study when *in silico* evaluating ROS537 probe, with coverage of 92.9% and specificity of 99.3%. In addition, previous studies reported false-positive hits and misclassification of marine Gammaproteobacteria with the Gam42a probe and of Alphaproteobacteria with the ALF968 probe, with group coverage of 76% and 81%, respectively [72]. In our study, ALF968 probe showed higher *in silico* coverage (96.3%) of Rhodobacterales order than ROS537 probe, but with lower specificity (~93%). Hence, we speculate that superior complementary base pairing close to 100% of ROS537 probe, especially for genera *Nereida*, *Planktomarina* and *Puniceibacterium* detected in higher relative abundances with *pufM* metabarcoding, could explain higher relative abundances observed with ROS537 than ALF968. This may indicate an urgent need for FISH probes redesign, especially ALF968. Furthermore, almost all Alphaproteobacteria (96%) detected in this study belong to the *Roseobacter* clade (*i.e.* Rhodobacterales order, Rhodobacteraceae family) as revealed by

metabarcoding results. Considering decreased specificity of the ALF968 probe, this might explain why numbers fall below that of ROS537 probe. Other major disadvantages of FISH, besides the coverage and affinity of the probes for complementary nucleic acids, are low and variable number of ribosomes (targets) in certain bacterial cells, the lack of cells' permeabilization and discrepancies in the binding of fully complementary oligonucleotides, discussed in detail elsewhere [72]. Another substantial limitation of this method that may have affected the results was the rapid fading of Bchl *a* autofluorescence when infrared and probe signals were simultaneously detected, which often required rapid and repeated manipulation when images were acquired. Nevertheless, positive correlation between FISH-IR and *pufM* pseudoabundances for corresponding taxonomic levels indicate that FISH-IR data image well relative patterns observed in AAP community, while Gam42a and ROS537 probes contribute solid quantitative estimates of these groups. This becomes important when considering inherent primer bias in metabarcoding, especially in respect to Gammaproteobacteria (see further below) where FISH-IR data provide alternative estimate and added value to the dataset (35.3% FISH-IR Gam42 vs. 75.9% *pufM* metabarcoding). These findings pave the way for protocol optimizations and new probe designs targeting AAPs.

Community structure from *pufM* gene metabarcoding

Although we have obtained taxonomic resolution at the genera level, for the purpose of comparison to previously published studies, taxonomic phylogroups proposed by Yutin et al. [41] will be referred to if they have a taxonomically described representative in our dataset. The community composition determined by *pufM* gene sequencing is consistent with the results of Auladell et al. [5] from the northwestern Mediterranean region mentioned earlier, where Gammaproteobacteria belonging to the NOR5/OM60 group (phylogroup K; Pseudomonadales) dominated the AAP composition over a decade with a mean relative abundance of 83.8%, while their mean relative contribution in our study was 75.9%. Interestingly, the similar decrease in the contribution of Gammaproteobacteria as in Auladell et al. [5] for months of February and March (59.6% and 52% on average, respectively) was also observed in our study, with a relative respective contribution of <40% and 50%. In these months, we detected an increased incidence of Alphaproteobacteria, specifically order Rhodobacterales, in February for ST101 and in March for CJ007. Order Rhodobacterales in the study area was associated with phylogroups E, F and G. Occasional peaks of certain AAPs with relative contributions >1%, as reported in other studies, were also present in our survey. For instance,

Sphingomonadales peaked in April, Caulobacterales in August, and the genus *Planktomarina* (phylogroup E) in February and November [5, 7]. However, a long-term survey is required to assess whether these occurrences are repeatable.

At the genus level, community composition in our study area was dominated by Gammaproteobacterium *Luminiphilus* (phylogroup K, OM60/NOR5 clade; order Pseudomonadales, family Halieaceae), a mesophilic and moderately halophilic photoheterotroph common in seawater and surface layers of littoral marine sediments [73]. Besides *Luminiphilus*, taxon with the highest relative contribution to AAP community composition was *LFERO1*, belonging to Rhodobacterales order (phylogroup G, family Rhodobacteraceae). Of interest is the report of two Gemmatimonadota ASVs at transition station CJ007, the first report of Gemmatimonadota AAPs in marine habitat, although they were present in a very low relative abundance. This indicates that amplicon sequencing provides a suitable tool for detection of very low abundant (0.1–10%) taxa, compared to metagenomics [19]. Gemmatimonadota is a poorly studied bacterial phylum, with only six cultivated representatives reported to date, whose photoheterotrophic and heterotrophic members are commonly found in soil, euphotic zones of freshwaters, sewage treatment plants and sediments, however their photoheterotrophic members were not yet detected in marine habitats [74, 75]. Two ASVs observed in our study were assigned to unclassified genera of the family Gemmatimonadaceae. Currently, only two cultivated representatives, *Gemmatimonas phototrophica* and *Gemmatimonas groenlandica*, are known to possess the capacity for anoxygenic photosynthesis. However, metagenome-assembled genome analyses have shown that anoxygenic photosynthesis is also present in uncultivated lineages of Gemmatimonadota [74, 76]. We hypothesise that the Gemmatimonadota detected in our study may originate from the river basin of the Adriatic karst rivers, namely Jadro and Cetina, or are passively transported from the sediment.

As previously confirmed by different methodological approaches such as metagenomics [5, 77], amplicon sequencing [5, 7] and microscopy [18], our results also indicate an inverse relationship between AAP abundance and diversity metrics in summer. The highest AAP diversity in relation to the Shannon index and Pielou's evenness was found when AAP abundance was the lowest in August. This contrasts with previous results from a long-term study by Auladell et al. [5] (NW Mediterranean), where diversity was highest (when abundance was lowest) in winter. In our study, no clear inverse relationship was found between AAP abundance and diversity measures in other seasons.

Although amplicon sequencing is cost-effective compared to metagenomics, primer biases due to high variability in protein-coding gene sequences, such as in the *pufM* gene, remain a major drawback of this approach [19]. A recent article by Gazulla et al. [19] compared metagenomic and metabarcoding approaches using existing and newly designed primers (*pufMF/pufM* WAW, *UniF/UniR* and *pufMF_Y/pufM* WAW) and showed that all primer pairs are biased towards Gammaproteobacteria and even some members of Alphaproteobacteria. In our study, we used *UniF/WAW* primer pair combination and obtained successful PCR amplification on marine samples, unlike Gazulla et al. [19] (see Additional file 1: Table S2). However, it is reasonable to expect these primers also possibly overestimate the Gammaproteobacterial AAPs (Pseudomonadales, phylogroup K) to the detriment of uncultured representatives and results should be interpreted with caution. As mentioned previously, *pufM* metabarcoding estimates relative contribution of Gammaproteobacteria to 75.9% while FISH-IR Gam42 to 35.3%. Moreover, PCR reaction optimisation was lengthy and tedious, possibly due to the very low GC content of *UniF*, the low melting temperature and ten degenerate nucleotides in the primer sequence. Instead, the newly designed forward primer *pufMF_Y* in combination with the reverse primer WAW and the universal *UniF/UniR* primer pair are currently proposed to obtain a less biased taxonomic representation of marine AAP [19].

Environmental variables affecting AAP community composition

Neural gas models linking sequencing data to biotic/abiotic environmental factors have already been successfully applied to 16S rRNA gene datasets [25, 27]. Here, they were performed at order-, genus- and ASV-levels of the *pufM* dataset to extract characteristic patterns that could potentially elucidate still unresolved ecology of AAPs. In general, the Adriatic Sea is an oligotrophic environment, scarce in nitrogen and phosphorus, with low productivity and elevated average salinity [78, 79]. In order- and genus-NG models, the maximum abundance of AAPs was detected in more productive environments (BMU1 in both and BMU5 in the genus-NG model), namely those with the highest concentrations of soluble reactive phosphorus, nitrate and dissolved inorganic nitrogen, consistently coupled with the highest abundance of total heterotrophic bacteria [10]. The gammaproteobacterial order Burkholderiales (phylogroup I), represented by the genera *Rhodospirillum* and *Limnochlamydomonas*, as well as the alphaproteobacterial lineage *SP197* and genera belonging to order Rhodobacterales, family Rhodobacteraceae *Planctomarina* (phylogroup E), *Nereida* (phylogroup G), *CYK10* and *Thalassobacter* (phylogroup E) were

present in increased relative abundances in nutrient-enriched environments with the highest AAP diversity. Conversely, the abundance of AAPs was lowest in nutrient-poor habitats with low nitrogen and phosphorus, according to order- and genus-model. Another feature evident in aforementioned models is the inverse relationship between AAP abundance and salinity as a potentially important controlling factor: the lowest abundance was found in the most saline habitats and vice versa. However, in contrast to previous research from saline lakes of the Tibetan Plateau, neither clear nor inverse relationship between diversity measures and salinity was found in our study [80]. The observed number of ASVs and Shannon index reached a maximum in habitats with the lowest salinity, while a similarly high number of ASVs, Shannon index and highest Pielou's evenness were found in habitats with the highest salinity (BMU3), suggesting that AAP diversity is not necessarily directly impacted by salinity. Furthermore, the high AAP diversity in different habitats, both nutrient-rich (BMU1) and nutrient-poor one (BMU3), was independent of temperature or salinity in the studied area, again indicating a different behaviour of AAPs compared to the general bacterial population. Previous results focusing on bacteria and archaea in the Adriatic Sea pointed to the "*plankton paradox*": archaeal and bacterial diversity is lowest in the environment with the highest abundance of picoplankton members, bacterial production and chlorophyll concentration, while, on the contrary, the highest bacterial diversity was measured in deep and saline environments abundant with nitrates, nitrites and soluble reactive phosphorus [25, 27]. Nevertheless, a high average salinity (38.35) combined with a temperature greater than 16 °C associated with the highest concentrations of nitrites, silicates and Chl *a* negatively affected AAP diversity (BMU 5). In contrast to previous studies [5, 24], there was neither a clear nor a linear relationship between Chl *a* (coupled with cyanobacterial abundance peaks), and the abundance of AAPs in our order- and genus-level models.

The alphaproteobacterial orders Rhizobiales (phylogroup J), Sphingomonadales, Caulobacterales and Arenicellales occurred simultaneously in higher incidence in habitats scarce with nitrate/ammonia and temperatures above 16 °C (BMUs 2 and 3). Conversely, they had low values in nitrogen-enriched environments (BMUs 1 and 4), indicating an inverse relationship between their relative abundances and nitrogen compounds and possibly their preference for low-nitrogen habitats. Caulobacterales, Rhizobiales and Sphingomonadales also co-occurred in a recent study conducted in freshwater lakes, possibly indicating a preference of these orders for the same environmental conditions, with their maxima observed in spring [7]. Recently, the contribution

of AAP photoheterotrophic activity to carbon fluxes in a freshwater lake was quantified and associated with higher relative abundance of Caulobacterales and Sphingomonadales, possibly implying an important role of these orders in aquatic habitats [17]. Interestingly, the uncultured alphaproteobacterial lineage UBA8366 and an unclassified order of Alphaproteobacteria showed the highest incidence only in the coldest, scarcest, and deepest environment (order-level BMU5) with an average depth of ~60 m and maximum average salinity (38.71), which coincided with the lowest absolute abundances of AAPs, total heterotrophic bacteria, *Synechococcus*, *Prochlorococcus*, picoeukaryotes, HNF and diversity metrics. This result is consistent with a previous study that found a dominance of Alphaproteobacteria in deeper and saltier seawater [27]. A recent article dealing with the comparative genomics of Alphaproteobacteria MAGs from the Arctic Ocean also showed that the uncultured UBA8366 lineage is involved in the degradation of aromatic compounds, especially humic substances of terrestrial origin [81].

To our knowledge, there is no previous literature to compare data obtained at genus- and ASV-level analysis in respect to sea-water environmental conditions. In addition to the euryhaline genera *Luminiphilus* (phylogroup K, OM60/NOR5 clade; order Pseudomonadales, family Halieaceae) and the uncultured *LFER01* (phylogroup G, order Rhodobacterales, family Rhodobacteraceae), which were prevalent in all five distinct environments (BMU1-5 at the genus level), alphaproteobacterial genera *Planktomarina* (phylogroup E), *Nereida* (phylogroup G), *CYK10*, *Thalassobacter* (phylogroup E) from family Rhodobacteraceae and the Gammaproteobacteria *RS62*, *Rhodoferrax* and *Limnohabitans* (phylogroup I, family Burkholderiaceae) appeared in highest values in nutrient-enriched habitat (genus-level BMU1), coupled with the highest abundance of AAPs, heterotrophic bacteria and picoeukaryotes. These genera preferred low-salinity environments enriched in nitrates, ammonia, dissolved inorganic nitrogen, phosphorus and silicates. As already mentioned, this is to be expected as *Roseobacter* group members, such as genus *Planktomarina*, utilise nitrogen only in reduced form as they lack the metabolic pathways for nitrate and nitrite reduction and rely on the uptake of ammonium, amino acids, spermidine and urea [66]. Regarding ASV-level neural network, it was observed that different variants of the same genus behaved differently when clustered in specific environment. However, working at the lowest taxonomic resolution possible did not result in easily interpretable/straightforward model which could be compared to previous literature and order- and genus-level BMUs. We contribute this to the fact that we are missing species

level taxonomic identification and interpretation was focused on higher taxonomic rank models instead.

Conclusion

Overall, significant seasonality was observed regarding AAP abundances (maximum average values in spring and minimum in summer), FISH-IR data (*Roseobacter* clade prevalent in autumn, Alpha- and Gammaproteobacteria in summer) and *pufM* sequencing dataset (highest diversity metrics in summer). Community composition in Adriatic Sea based on *pufM* metabarcoding was dominated by Gammaproteobacteria belonging to the NOR5/OM60 clade, namely the genus *Luminiphilus*, while numerous other genera were present in low relative abundances <1%. An inverse relationship between AAP abundance and diversity metrics was observed in summer. Neural gas models revealed potentially important controlling variables of AAP abundance, community structure, and diversity measures. The high AAP diversity was independent of temperature or salinity in various trophic environments, indicating a different behaviour of AAPs compared to the general bacterial population. We emphasize there is need for redesign of FISH-IR Alphaproteobacterial hybridization probe, however, positive correlation between FISH-IR and *pufM* metabarcoding datasets indicate that FISH-IR provides solid quantitative estimates of relative abundances of Gammaproteobacteria (Gam42a probe) and *Roseobacter* clade (ROS537 probe). This represents an added value of combining qualitative and quantitative approaches, especially when considering inherent primer bias towards Gammaproteobacteria in *pufM* metabarcoding.

Abbreviations

AAPs	Aerobic anoxygenic phototrophs
ASV	Amplicon sequence variant
BMU	Best match unit
BChl a	Bacteriochlorophyll-a
BP	Bacterial production
Chl a	Chlorophyll a
CLR	Centered log-ratio
DAPI	4',6-Diamidino-2-phenylindole
dChlMax	Deep chlorophyll maximum
DIN	Dissolved inorganic nitrogen,
FISH	Fluorescence in situ hybridisation
FISH-IR	Infrared epifluorescence microscopy and fluorescence in situ hybridisation
HNA	High nucleic acid
HNF	Heterotrophic nanoflagellates
LNA	Low nucleic acid
NG	Neural gas
NH ₄ ⁺	Ammonium ion
NO ₂ ⁻	Nitrites
NO ₃ ⁻	Nitrates
NTOT	Total nitrogen
PE	Picoeukaryotes
PROCHL	<i>Prochlorococcus</i>
PTOT	Total phosphorus
Sal	Salinity
SiO ₄ ²⁻	Silicate

SRP	Soluble reactive phosphorus
SYN	<i>Synechococcus</i>
Temp	Temperature
UHB	Total heterotrophic bacteria

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40793-024-00573-6>.

Additional file 1. Table S1. Description of samples analysed in this study ordered by month (date of collection, W-Winter, Sp-Spring, S-Summer, A-Autumn); **Table S2.** Number of reads per sample that passed through dada2 pipeline; input: number of raw reads, filtered: filtered out low quality sequences and tails, denoisedF and denoisedR: denoised sequences after the core sample inference algorithm of forward and reverse reads respectively, merged: merged sequences which are output if the forward and reverse reads overlap by at least 12 bases and are identical to each other in the overlap region, nonchim: sequences after chimera removal [1]. Samples with an unacceptably low final number of reads per sample ($N < 2000$) were excluded from further analysis; **Figure S1.** Rarefaction curve exhibiting sufficient sequencing depth (Sample Size) for AAP diversity metrics estimates. Rarefying to the smallest library size was performed multiple times (function `rarefy_even_depth` from `phyloseq` v1.32.0 R package [2], rarefying threshold = 2000, $N(\text{times}) = 100$); **Figure S2.** Draftsman plots used to estimate correlation of abiotic environmental variables, made in PRIMER7 [3]. Variables are: temperature-Temp, salinity-Sal, nitrates-NO₃⁻, nitrites-NO₂⁻, ammonium ion-NH₄⁺, dissolved inorganic nitrogen-DIN, total nitrogen-NTOT, soluble reactive phosphorus-SRP, total phosphorus-PTOT, silicate-SiO₄⁻; **Table S3.** Permutational multivariate analysis of variance (PERMANOVA) on Euclidean distance of square-root transformed AAPs' absolute (A) and relative (B) abundance dataset. Factors: Se-Season (fixed), Re-Region (fixed), La- Layer (nested in Region, L1 (0-30m), L2 (30-50m), L3 (50-75m), L4 (75-100m)). Pairwise comparisons for significant seasonality (W-Winter, Sp-Spring, S-Summer, A-Autumn) are given in the right part of the table. PERMANOVA was performed in PRIMER7 with 9999 permutations, unrestricted permutation of raw data, sums of squares type: Type II (conditional) [3]; **Table S4.** Permutational multivariate analysis of variance (PERMANOVA) on Aitchison distance of centered-log ratio transformed pufM sequencing dataset agglomerated on genus-level. Factors: Se- Season (fixed), Re-Region (fixed), La- Layer (nested in Region, L1 (0-30m), L2 (30-50m), L3 (50-75m), L4 (75-100m)). Pairwise comparisons for significant seasonality (W-Winter, Sp-Spring, S-Summer, A-Autumn) are given in the right part of the table. PERMANOVA was performed in PRIMER7 with 9999 permutations, unrestricted permutation of raw data, sums of squares type: Type II (conditional) [3]; **Figure S3.** AAP community composition obtained via pufM metabarcoding in the study area shown at the class level per station (ST101, CJ007, CJ009), month and depth.; **Table S5.** Alpha diversity metrics based on rarefied pufM dataset (Observed number of ASVs, Shannon and Pielou's index) given as average values with standard deviations (SD) calculated per season (vegan v2.5.7 R package [4]).

Additional file 2. Detailed description of Infrared epifluorescence microscopy and fluorescence in situ hybridisation (FISH-IR). **Table S1.** Permutational multivariate analysis of variance (PERMANOVA) on Bray-Curtis similarity for square-root transformed FISH-IR relative abundance dataset. Factors: Se- Season (fixed), Re-Region (fixed), La- Layer (nested in Region, L1 (0-30m), L2 (30-50m), L3 (50-75m), L4 (75-100m)). Pairwise comparisons for significant seasonality (W-Winter, Sp-Spring, S-Summer, A-Autumn) are given in the right part of the table. PERMANOVA was performed in PRIMER7 with 9999 permutations, Unrestricted permutation of raw data, sums of squares type: Type II (conditional) [5]; **Figure S1.** Two-dimensional nonmetric multidimensional scaling (NMDS) ordination plot of FISH-IR square root transformed dataset (relative abundances of AAPs assigned to Alphaproteobacteria, Gammaproteobacteria and Roseobacter clade) on a seasonal scale (W-Winter, Sp-Spring, S-Summer, A-Autumn); **Figure S2.** Spatio-temporal distribution of FISH-IR groups (probes for Alphaproteobacteria, Gammaproteobacteria, Roseobacter) given as relative abundances of AAPs ($n=90$ samples), shown per station

(ST101, CJ007 and CJ009), month, and depth.; **Figure S3.** Seasonal distribution (W-winter, Sp-Spring, S-Summer, A-Autumn) of FISH-IR groups (Alphaproteobacteria, Gammaproteobacteria, Roseobacter) given as average relative abundances of AAPs.

Additional file 3. Sample data, untransformed final ASV matrix, taxonomy and reference pufM sequences (refseq).

Additional file 4. Correlation between metabarcoding and FISH-IR data; Spearman correlation coefficients between pseudoabundances from pufM sequencing (i.e. relative abundance of Gammaproteobacteria class \times AAP absolute number/100) and FISH-IR counts.

Additional file 5. In silico coverage and specificity of ALF968 and ROS537 probes.

Additional file 6. Spatio-temporal distribution of Luminiphilus ASVs. **Figure 1.** Genus Luminiphilus composition via pufM metabarcoding at ST101 station shown per month and depth. Category "Other" represents ASVs with relative abundances below 5%; **Figure 2.** Genus Luminiphilus composition via pufM metabarcoding at CJ007 station shown per month and depth. Category "Other" represents ASVs with relative abundances below 5%. **Figure 3.** Genus Luminiphilus composition via pufM metabarcoding at CJ009 station shown per month and depth. Category "Other" represents ASVs with relative abundances below 5%.

Additional file 7. Detailed description of neural gas analysis results.

Figure S1. Heatmap of order-agglomerated and CLR-transformed pufM dataset resulting in five distinct BMUs. Samples are shown for each unit; **Figure S2.** Genus-agglomerated and CLR-transformed pufM dataset resulting in five distinct BMUs. Samples are shown for each BMU; **Figure S3.** Neural gas analysis results of pufM dataset at clr-transformed ASV-level, clustered into five BMUs. Biotic (UHB, HIGH, SYN, PROCHL, PE, BP, HNF, AAP) and abiotic (Temp, Sal, NO₃⁻, NO₂⁻, NH₄⁺, DIN, NTOT, PO₄³⁻, PTOT, SiO₄²⁻, Chl a) variables of each environment are given in (A) as average value for each unit. Colour gradient from red to green represents the lowest and highest average values respectively. Relative contribution expressed as CLR-transformed value of specific ASV is shown in (B). ASVs are ordered by genus, with most dominant genera ordered alphabetically.

Additional file 8. Neural gas analysis at the order level: sample data with resulting BMUs for each sample, ASV matrices and taxonomy table.

Additional file 9. Neural gas analysis at the genus level: sample data with resulting BMUs for each sample, ASV matrices and taxonomy table.

Acknowledgements

We sincerely thank Ante Marasović and David Udovičić for help with sampling aboard the R/V BIOS DVA and Ivan Vučić for preparing Figure 1. We also wish to thank the Laboratory of Chemical Oceanography and Sedimentology of the Sea for the physicochemical analyses and the Laboratory of Plankton and Shellfish Toxicity for providing the chlorophyll data.

Author contributions

IS and AVT performed the sampling. Experimental procedures were carried out by IS, ILP, ŽT and AVT while DŠ, SŠ, ŽNG and GK provided environmental variables. Bioinformatic and statistical analyses were performed by IS, CVA, FM and ŽT. CVA provided taxonomic database. IS and ŽT performed data visualization. DŠ provided funding. IS wrote the original draft and all the authors helped reviewing and editing the manuscript.

Funding

This work was fully supported by the ADRISAAP project (UIP-2019-04-8401: Ecology of Aerobic Anoxygenic phototrophs in the Adriatic Sea).

Data availability

The datasets supporting the results of this article are included within the article and its additional files. The raw sequencing dataset supporting the conclusions of this article is available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) as a part of BioProject PRJNA912619 under accession numbers SAMN37596707-SAMN37596796.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Institute of Oceanography and Fisheries, Šetalište Ivana Meštrovića 63, Split, Croatia. ²Doctoral Study of Biophysics, Faculty of Science, University of Split, Ruđera Boškovića 37, Split, Croatia. ³Laboratory of Anoxygenic Phototrophs, Institute of Microbiology, Czech Academy of Sciences, 379 81 Třeboň, Czechia. ⁴Department of Ecosystem Biology, Faculty of Science, University of South Bohemia, České Budějovice, Czechia. ⁵University Department of Marine Studies, University of Split, Ruđera Boškovića 37, Split, Croatia. ⁶Department of Fisheries, Oceanography and Marine Ecology, National Marine Fisheries Research Institute, Gdynia, Poland.

Received: 19 December 2023 Accepted: 22 April 2024

Published online: 29 April 2024

References

- Harashima K, Shiba T, Murata N. Aerobic photosynthetic bacteria. 1989.
- Kolber ZS, Van Dover CL, Niederman RA, Falkowski PG. Bacterial photosynthesis in surface waters of the open ocean. *Nature*. 2000;407:177–9.
- Kolber ZS, Plumley FG, Lang AS, Beatty JT, Blankenship RE, VanDover CL, et al. Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. *Science* (80-). 2001;292:2492–5.
- Hojerová E, Mašín M, Brunet C, Ferrera I, Gasol JM, Koblížek M. Distribution and growth of aerobic anoxygenic phototrophs in the Mediterranean Sea. *Environ Microbiol*. 2011;13:2717–25.
- Auladell A, Sánchez P, Sánchez O, Gasol JM, Ferrera I. Long-term seasonal and interannual variability of marine aerobic anoxygenic photoheterotrophic bacteria. *ISME J*. 2019;13:1975–87.
- Villena-Aleman C, Mujakić I, Fecskeová LK, Woodhouse J, Auladell A, Dean J, et al. Phenology and ecological role of Aerobic Anoxygenic Phototrophs in fresh waters. *Microbiome*. 2024;12:65.
- Villena-Aleman C, Mujakić I, Porcal P, Koblížek M, Pivosz K. Diversity dynamics of aerobic anoxygenic phototrophic bacteria in a freshwater lake. *Environ Microbiol Rep*. 2023;15:60–71.
- Sánchez O, Ferrera I, Mabrito I, Gazulla CR, Sebastián M, Auladell A, et al. Seasonal impact of grazing, viral mortality, resource availability and light on the group-specific growth rates of coastal Mediterranean bacterioplankton. *Sci Rep*. 2020. <https://doi.org/10.1038/s41598-020-76590-5>.
- Ferrera I, Sánchez O, Kolářová E, Koblížek M, Gasol JM. Light enhances the growth rates of natural populations of aerobic anoxygenic phototrophic bacteria. *ISME J*. 2017;11:2391–3.
- Koblížek M. Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiol Rev*. 2015;39:854–70.
- Vrdoljak Tomaš A, Šantić D, Šolić M, Skejić S, Milinković A, Cvitešić Kušan A, et al. How do open coastal fire episodes' impact sea surface microlayer neuston communities? *Sci Total Environ*. 2023;861:160593.
- Eiler A. Evidence for the ubiquity of mixotrophic bacteria in the upper ocean: implications and consequences. *Appl Environ Microbiol*. 2006;72:7431–7.
- Pernthaler J. Predation on prokaryotes in the water column and its ecological implications. *Nat Rev Microbiol*. 2005;3:537–46.
- Ferrera I, Gasol JM, Sebastián M, Hojerová E, Koblížek M. Comparison of growth rates of aerobic anoxygenic phototrophic bacteria and other bacterioplankton groups in coastal mediterranean waters. *Appl Environ Microbiol*. 2011;77:7451–8.
- Stegman MR, Cottrell MT, Kirchman DL. Leucine incorporation by aerobic anoxygenic phototrophic bacteria in the Delaware estuary. *ISME J*. 2014;8:2339–48.
- Fecskeová LK, Pivosz K, Šantić D, Šestanović S, Vrdoljak Tomaš A, Hanusova M, et al. Lineage-specific growth curves document large differences in response of individual groups of marine bacteria to the top-down and bottom-up controls. *mSystems*. 2021;6:e00934-21.
- Pivosz K, Villena-Aleman C, Mujakić I. Photoheterotrophy by aerobic anoxygenic bacteria modulates carbon fluxes in a freshwater lake. *ISME J*. 2022;16:1046–54.
- Ferrera I, Borrego CM, Salazar G, Gasol JM. Marked seasonality of aerobic anoxygenic phototrophic bacteria in the coastal NW Mediterranean Sea as revealed by cell abundance, pigment concentration and pyrosequencing of *pufM* gene. *Environ Microbiol*. 2014;16:2953–65.
- Gazulla CR, Cabello AM, Sánchez P, Gasol JM, Sánchez O, Ferrera I. A metagenomic and amplicon sequencing combined approach reveals the best primers to study marine aerobic anoxygenic phototrophs. *Microb Ecol*. 2023;86:2161–72.
- Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. *Front Microbiol*. 2017;8:1–6.
- Kasalický V, Zeng Y, Pivosz K, Šimek K, Kratochvilová H, Koblížek M. Aerobic anoxygenic photosynthesis is commonly present within the genus *Limnohabitans*. *Appl Environ Microbiol*. 2018. <https://doi.org/10.1128/AEM.02116-17>.
- Šantić D, Šestanović S, Vrdoljak A, Šolić M, Kušpilić G, Ninčević Gladan Ž, et al. Distribution of aerobic anoxygenic phototrophs in the Eastern Adriatic Sea. *Mar Environ Res*. 2017;130:134–41.
- Vrdoljak Tomaš A, Šantić D, Šolić M, Ordulj M, Jozić S, Šestanović S, et al. Dynamics of aerobic anoxygenic phototrophs along the trophic gradient in the central Adriatic Sea. *Deep Res Part II Top Stud Oceanogr*. 2018;2019(164):112–21.
- Vrdoljak Tomaš A, Šantić D, Stojan I, Šolić M. Aerobic anoxygenic phototrophs of the Adriatic Sea. *Acta Adriat*. 2023;64:1–10.
- Šantić D, Pivosz K, Matic F, Vrdoljak Tomaš A, Arapov J, Dean JL, et al. Artificial neural network analysis of microbial diversity in the central and southern Adriatic Sea. *Sci Rep*. 2021;11:1–15.
- Artegiani A, Bregant D, Paschini E, Pinardi N, Raicich F, Russo A. The Adriatic Sea general circulation. Part I: air-sea interactions and water mass structure. *J Phys Oceanogr*. 1997;27:1492–514.
- Šantić D, Stojan I, Matic F, Trumbić Ž, Tomaš AV, Fredotović Ž, et al. Picoplankton diversity in an oligotrophic and high salinity environment in the central Adriatic Sea. *Sci Rep*. 2023. <https://doi.org/10.1038/s41598-023-34704-9>.
- Fuhrman JA, Azam F. Thymidine incorporation as a measure of heterotrophic bacterioplankton production in marine surface waters: evaluation and field results. *Mar Biol*. 1982;66:109–20.
- Mašín M, Zdun A, Stoń-Egiert J, Nausch M, Labrenz M, Moulisová V, et al. Seasonal changes and diversity of aerobic anoxygenic phototrophs in the Baltic Sea. *Aquat Microb Ecol*. 2006;45:247–54.
- Mašín M, Nedoma J, Pechar L, Koblížek M. Distribution of aerobic anoxygenic phototrophs in temperate freshwater systems. *Environ Microbiol*. 2008;10:1988–96.
- Stojan I, Trumbić Ž, Lepen Pleić I, Šantić D. Evaluation of DNA extraction methods and direct PCR in metabarcoding of mock and marine bacterial communities. *Front Microbiol*. 2023. <https://doi.org/10.3389/fmicb.2023.1151907>.
- Béjà O, Suzuki M, Heidelberg J, Nelson W, Preston C, Hamada T, et al. Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature*. 2002;415:630–3.
- Yutin N, Suzuki MT, Béjà O. Novel primers reveal wider diversity among marine aerobic anoxygenic phototrophs. *Appl Environ Microbiol*. 2005;71:8958–62.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal*. 2011;17:10–2.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13:581–3.
- Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol*. 2007;73:5261–7.
- McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8:e61217.

38. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30:772–80.
39. Barbera P, Kozlov AM, Czech L, Morel B, Darriba D, Flouri T, et al. EPA-ng: massively parallel evolutionary placement of genetic sequences. *Syst Biol.* 2019;68:365–9.
40. Czech L, Barbera P, Stamatakis A. Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data. *Bioinformatics.* 2020;36:3263–5.
41. Yutin N, Suzuki MT, Teeling H, Weber M, Venter JC, Rusch DB, et al. Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the global ocean sampling expedition metagenomes. *Environ Microbiol.* 2007;9:1464–75.
42. Price MN, Dehal PS, Arkin AP. Fasttree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol.* 2009;26:1641–50.
43. Wickham H. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York; 2016. <https://ggplot2.tidyverse.org>
44. Sisk-Hackworth L, Kelley ST. An application of compositional data analysis to multiomic time-series data. *NAR Genomics Bioinforma.* 2020. <https://doi.org/10.1093/nargab/lqaa079>.
45. Tsilimigras MCB, Fodor AA. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Ann Epidemiol.* 2016;26:330–5.
46. Aitchison J. The statistical analysis of compositional data. *J R Stat Soc Ser B.* 1982;44:139–60.
47. Lahti L, Shetty S, Blake T, Salojärvi J. microbiome R package. 2017.
48. Oksanen AJ, Blanchet FG, Friendly M, Kindt R, Legendre P, Mcglinn D, et al. vegan community ecology package version 2.5–7 November 2020.
49. Clarke KR, Gorley RN. Primer: user manual/tutorial. Prim Ltd, Plymouth, UK. 2015; 93.
50. Anderson MJ. Permutational Multivariate Analysis of Variance (PERMANOVA). *Wiley StatsRef Stat Ref Online.* 2017; 1–15.
51. Anderson MJ. Distance-based tests for homogeneity of multivariate dispersions. *Biometrics.* 2006;62:245–53.
52. Barnett DJM, Arts ICW, Penders J. microViz : an R package for microbiome data visualization and statistics. *J Open Source Softw.* 2021;6:3201.
53. Martinez TM, Berkovich SG, Schulten KJ. “Neural-Gas” network for vector quantization and its application to time-series prediction. *IEEE Trans Neural Netw.* 1993;4:558–69.
54. Šolić M, Šantić D, Šestanović S, Kušpilić G, Matić F, Vrdoljak Tomaš A, et al. Changing ecological conditions in the marine environment generate different microbial food web structures in a repeatable manner. *Front Mar Sci.* 2022. <https://doi.org/10.3389/fmars.2021.811155>.
55. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. ggplots: Various R programming tools for plotting Data.R package version 3.1.3. 2022.
56. Celussi M, Gallina AA, Ras J, Gianni M, Del NP. Effect of sunlight on prokaryotic organic carbon uptake and dynamics of pigments relevant to photoheterotrophy in the Adriatic Sea. *Aquat Microb Ecol.* 2015;74:235–9.
57. Lamy D, De Carvalho-Maalouf P, Cottrell MT, Lami R, Catala P, Oriol L, et al. Seasonal dynamics of aerobic anoxygenic phototrophs in a Mediterranean coastal lagoon. *Aquat Microb Ecol.* 2011;62:153–63.
58. Waidner LA, Kirchman DL. Aerobic anoxygenic phototrophic bacteria attached to particles in turbid waters of the Delaware and Chesapeake estuaries. *Appl Environ Microbiol.* 2007;73:3936–44.
59. Pernthaler A, Pernthaler J, Amann R. Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl Environ Microbiol.* 2002;68:3094–101.
60. Kubota K. CARD-FISH for environmental microorganisms: Technical advancement and future applications. *Microbes Environ.* 2013;28:3–12.
61. Magalhães C, Semedo M, Vezzi A, Øvreås L, Fadeev E, Cardozo-Mino MG, et al. Comparison of two 16S rRNA primers (V3–V4 and V4–V5) for studies of Arctic microbial communities. *Front Microbiol.* 2021. <https://doi.org/10.3389/fmicb.2021.637526>.
62. Luo H, Moran MA. Evolutionary ecology of the marine roseobacter clade. *Microbiol Mol Biol Rev.* 2014;78:573–87.
63. Simon M, Scheuner C, Meier-Kolthoff JP, Brinkhoff T, Wagner-Döbler I, Ulbrich M, et al. Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation to marine and non-marine habitats. *ISME J.* 2017;11:1483–99.
64. Mary I, Cummings DG, Biegala IC, Burkill PH, Archer SD, Zubkov MV. Seasonal dynamics of bacterioplankton community structure at a coastal station in the western English Channel. *Aquat Microb Ecol.* 2006;42:119–26.
65. Koblížek M, Moulisová V, Muroňová M, Oborník M. Horizontal transfers of two types of puf operons among phototrophic members of the Roseobacter clade. *Folia Microbiol (Praha).* 2015;60:37–43.
66. Giebel HA, Wolterink M, Brinkhoff T, Simon M. Complementary energy acquisition via aerobic anoxygenic photosynthesis and carbon monoxide oxidation by *Planktomarina temperata* of the Roseobacter group. *FEMS Microbiol Ecol.* 2019;95:1–9.
67. Teira E, Martínez-García S, Lønborg C, Álvarez-Salgado XA. Growth rates of different phylogenetic bacterioplankton groups in a coastal upwelling system. *Environ Microbiol Rep.* 2009;1:545–54.
68. Pinhasi J, Berman T. Differential growth response of colony-forming α - and γ -proteobacteria in dilution culture and nutrient addition experiments from Lake Kinneret (Israel), the Eastern Mediterranean Sea, and the Gulf of Eilat. *Appl Environ Microbiol.* 2003;69:199–211.
69. Alonso-Sáez L, Gasol JM. Seasonal variations in the contributions of different bacterial groups to the uptake of low-molecular-weight compounds in Northwestern Mediterranean coastal waters. *Appl Environ Microbiol.* 2007;73:3528–35.
70. Simonato F, Gómez-Pereira PR, Fuchs BM, Amann R. Bacterioplankton diversity and community composition in the Southern Lagoon of Venice. *Syst Appl Microbiol.* 2010;33:128–38.
71. Riou V, Périot M, Biegala IC. Specificity re-evaluation of oligonucleotide probes for the detection of marine picoplankton by tyramide signal amplification-fluorescent in situ hybridization. *Front Microbiol.* 2017;8:1–13.
72. Amann R, Fuchs BM. Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nat Rev Microbiol.* 2008;6:339–48.
73. Spring S, Riedel T, Spröer C, Yan S, Harder J, Fuchs BM. Taxonomy and evolution of bacteriochlorophyll a-containing members of the OM60/NOR5 clade of marine gammaproteobacteria: Description of *Luminiphilus syltensis* gen. nov., sp. nov., reclassification of *Hallea rubra* as *Pseudohallea rubra* gen. nov., comb. nov. *BMC Microbiol.* 2013;13:1–21.
74. Mujakić I, Cabello-Yeves PJ, Villena-Aleman Y, Piewosz K, Rodriguez-Valera F, Picazo A, et al. Multi-environment ecogenomics analysis of the cosmopolitan phylum Gemmatimonadota. *Environ Microbiol.* 2023. <https://doi.org/10.1128/spectrum.01112-23>.
75. Mujakić I, Andrei A-Ş, Shabarova T, Fecskeová LK, Salcher MM, Piewosz K, et al. Common presence of phototrophic gemmatimonadota in temperate freshwater lakes. *mSystems.* 2021. <https://doi.org/10.1128/mSystems.01241-20>.
76. Mujakić I, Piewosz K, Koblížek M. Phylum gemmatimonadota and its role in the environment. *Microorganisms.* 2022;10:1–17.
77. Galand PE, Pereira O, Hochart C, Auguet JC, Debroas D. A strong link between marine microbial community composition and function challenges the idea of functional redundancy. *ISME J.* 2018;12:2470–8.
78. Vilibić I, Matijević S, Šepić J, Kušpilić G. Changes in the Adriatic oceanographic properties induced by the Eastern Mediterranean Transient. *Biogeosciences.* 2012;9:2085–97.
79. Beg Paklar G, Vilibić I, Grbec B, Matić F, Mihanović H, Džoić T, et al. Record-breaking salinities in the middle Adriatic during summer 2017 and concurrent changes in the microbial food web. *Prog Oceanogr.* 2020;185:102345.
80. Jiang H, Dong H, Yu B, Lv G, Deng S, Wu Y, et al. Abundance and diversity of aerobic anoxygenic phototrophic bacteria in saline lakes on the Tibetan plateau. *FEMS Microbiol Ecol.* 2009;67:268–78.
81. Grevesse T, Guéguen C, Onana VE, Walsh DA. Degradation pathways for organic matter of terrestrial origin are widespread and expressed in Arctic Ocean microbiomes. *Microbiome.* 2022;10:1–21.

Publisher's Note

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