Data in Brief 67 (2026) 112971

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Surface sediments prokaryotic communities: five years of 16S rRNA amplicon sequencing data from the northernmost part of the Mediterranean Sea

Noemi Tomasi[#],[*], Elisa Banchi[#], Vincenzo Manna, Mauro Celussi

National Institute of Oceanography and Applied Geophysics – OGS, Sgonico (Trieste), Italy

A R T I C L E I N F O

Article history:
Received 10 April 2026
Revised 26 May 2026
Accepted 10 June 2026
Available online 11 June 2026

Dataset link: PRJNA1442017 Surface sediment 16S rRNA amplicon sequencing of the northern Adriatic Sea, Jun 20 - May 25 (Original data)

Keywords:
Gulf of Trieste
Microbiome
Long-term-ecological research (LTER)
Bacteria
Archaea

A B S T R A C T

Surface sediments harbour diverse prokaryotic communities that play a key role in biogeochemical cycling and provide valuable insights when compared with water column communities, allowing for a more comprehensive understanding of marine ecosystem functioning. Specifically, this dataset presents prokaryotic community data from 16 surface sediment samples collected seasonally from June 2020 to May 2025 at the C1-LTER station (45°42′2.99″ N, 13°42′36.00″ E; DEIMS.iD https://deims.org/96969205-cfdf-41d8-979f-ff881ea8dc8b) in the Gulf of Trieste, located in the northeastern Adriatic Sea (Mediterranean Sea). Extracted DNA was sequenced following the 16S Metagenomic Sequencing Library Preparation protocol and run on an Illumina NovaSeq 6000 System. Raw reads were filtered and denoised with DADA2, and taxonomic assignment was performed against the Silva 138.2 99% reference database. The dataset provides useful insights into prokaryotic communities and their seasonal variability over five years. Moreover, a focus on specific taxa is provided, such as Cyanobacteriota and Archaea, highlighting patterns of community variability in the sediment. Finally, it shows seasonal stability and generally consistent taxa distribution over time, as indicated by the high proportion of shared taxa

---

[*] Corresponding author.
E-mail address: ntomasi@ogs.it (N. Tomasi).
[#] These authors contributed equally to this manuscript.

https://doi.org/10.1016/j.dib.2026.112971

at each taxonomic level. The raw data, deposited in the NCBI Sequence Read Archive (SRA) under BioProject PRJNA1442017, include two sets of sequencing reads obtained from surface sediment samples using the Illumina MiSeq and Illumina NovaSeq 6000 sequencing platforms, for a total of 27 16S rRNA gene sequencing FASTQ files. Overall, these data provide valuable insight into the surface sediment community in the northernmost part of the Mediterranean Sea, contributing to long-term research on sediment prokaryotic communities.

Specifications Table

Subject	Biology
Specific subject area	16S rRNA metabarcoding analysis of sediment prokaryotic community.
Type of data	Raw FASTQ sequence files and metadata files. Figures (bar plots, NMDS plot, Venn diagrams). Table (sample details).
Data collection	Surface sediment samples were processed using the DNeasy PowerSoil Pro Kit (Qiagen). Extracted DNA was quantified with a Qubit fluorometer (Thermo Fisher Scientific). Libraries were prepared following the 16S Metagenomic Sequencing Library Preparation protocol and run on an Illumina NovaSeq 6000 System. Primers were removed with cutadapt (v. 4.0). DNA metabarcoding analysis was performed using DADA2 (v. 1.36.0) to filter and denoise raw sequences and to infer amplicon sequence variants (ASVs). Taxonomy assignment to ASVs was performed against the Silva reference database (version 138.2).
Data source location	Sediment samples were collected by the OGS team from C1-LTER station (45°42′2.99″N, 13°42′36.00″E) (DEIMS.ID: https://deims.org/96969205-cfdf-41d8-979f-ff881ea8dc8b) in the Gulf of Trieste (Adriatic Sea), Italy.
Data accessibility	Repository name: NCBI Sequence Read Archive (SRA) Data identification number: BioProject ID PRJNA1442017 Direct URL to data: https://www.ncbi.nlm.nih.gov/sra/PRJNA1442017
Related research article	None.

1. Value of the Data

- These data describe the temporal succession of prokaryotic communities inhabiting the surface sediments of the northernmost basin of the Mediterranean Sea, based on 16S rRNA-based taxonomic analysis over five years.
- It offers detailed insights into the dynamics of key taxa, such as Cyanobacteriota and Archaea, highlighting the overall seasonal stability of surface sediment microbial communities across different taxonomic levels.
- The dataset provides a long-term record of the microbial community composition of sediment samples, which, combined with the supplied environmental context, could help detect possible long-term biogeochemical deviations.
- The dataset can serve as a starting point for researchers interested in the ecological functioning of prokaryotic communities in shallow marine sediments, supporting future independent investigations on their role in organic matter degradation, carbon cycling, benthic–pelagic coupling, and ecosystem functioning in temperate coastal areas.

2. Background

Surface sediments host diverse prokaryotic communities [1] playing a critical role in biogeochemical cycling [2,3]. Although sediment microbial communities are generally seasonally stable [4], they are pivotal for benthic–pelagic coupling [5]. Therefore, surface sediment samples can be investigated together with the water column communities to improve the overall understanding of marine ecosystem functioning [6]. Previous studies conducted at the C1-LTER station have highlighted the functional role of the surface sediment microbiome. Franzo et al. [2] demonstrated that prokaryotic carbon cycling in surface sediments strongly depends on temperature and substrate availability. Additionally, Cibic et al. [7] emphasised the role of benthic primary production and microphytobenthic dynamics in shaping benthic-pelagic coupling at the sampling site. However, long-term datasets on sediment microbial communities remain limited. The purpose of this study was to develop a comprehensive dataset describing surface sediment samples over five years using a standardized 16S rRNA amplicon sequencing workflow. The dataset is intended to support future independent research by enabling further exploration of the role of surface sediment prokaryotic communities.

3. Data Description

The dataset includes 27 16S rRNA gene amplicon sequencing data (FASTQ files) generated from 16 surface sediment samples collected seasonally in the Gulf of Trieste (Adriatic Sea). As shown in Table 1, the dataset includes 11 samples (4 time points) produced using the Illumina MiSeq System and 16 samples (all 16 time points) produced using the Illumina NovaSeq 6000 System. MiSeq-sequenced samples were re-sequenced on the NovaSeq 6000 platform to increase sequencing depth and data quality; consequently, the number of FASTQ files in the dataset exceeds the number of unique sediment samples collected. In addition, samples sequenced using the Illumina MiSeq System included two or three biological replicates, whereas those processed with the Illumina NovaSeq 6000 System included one pooled sample, obtained by pooling extracted DNA prior to sequencing, as detailed in the “Experimental Design, Materials and Methods” section. The complete list of FASTQ files and associated metadata is provided in Table 1. The data are publicly accessible in the NCBI Sequence Read Archive (SRA) under the accession number PRJNA1442017 (https://www.ncbi.nlm.nih.gov/sra/PRJNA1442017).

In both cases, sequencing targeted the V4-V5 variable region of the 16S rRNA gene, generating an average of 122,175 ± 59,269 and 588,347 ± 121,477 raw reads per sample for MiSeq and NovaSeq systems, respectively. On average, MiSeq-generated raw reads showed a mean quality score of 33.5 ± 0.27, while NovaSeq-generated raw reads showed a mean quality score of 34.5 ± 0.12. Detailed information on the Quality Scores of raw reads (Figure S1 and Table S1), as well as the number of sequencing reads retained at each major step of the bioinformatic pipeline (Table S2), are reported in the “Supplementary Materials”.

Illumina MiSeq data were initially used in a subset of samples (Table 1) to estimate the representativeness of the replicates (for a total of two replicates for April 2021 and three replicates for the remaining samples, see “Experimental Design, Materials And Methods” section), and all the samples were then (re)sequenced with Illumina Novaseq 6000. Therefore, the subsequent sections will be based on the description of data obtained by the Novaseq 6000 sequencing platform.

Novaseq sequencing yielded a total of 42,673 Amplicon Sequence Variants (ASVs), covering 86 phyla, 784 families and 1467 genera. Pseudomonadota and Planctomycetota were the most abundant phyla (Fig. 1a), with mean relative abundances across samples of 32.7 ± 3.4% and 16.2 ± 1.2%, respectively, and exhibited only minor seasonal variation across years (Fig. 1a). At the genus level, *Woeseia* and Candidatus *Nitrosopumilus* were the most predominant genera (Fig. 1b), accounting for a mean relative abundance of 5.5 ± 0.6% and 3.6 ± 2.1%, respectively. While *Woeseia* was almost equally present across samples, Candidatus *Nitrosopumilus* showed a higher degree of seasonal variability, with lower abundance in summer samples. Nevertheless,

Table 1
Metadata associated with the analysed samples, including Sample number (Nr), Sample name, NCBI Sequence Read Archive (SRA) BioSample accession number, Sampling date, Season, Replicate, and Sequencing platform(s) used for sequencing analyses.

Nr	Sample name	BioSample accession number	Date	Season	Replicate	Sequencing platform(s)
1	C1_jun20_summer_r1	SAMN60340584	25/06/2020	Summer	R1	Illumina MiSeq
2	C1_jun20_summer_r2	SAMN60340585	25/06/2020	Summer	R2	Illumina MiSeq
3	C1_jun20_summer_r3	SAMN60340586	25/06/2020	Summer	R3	Illumina MiSeq
4	C1_sep20_summer_r1	SAMN60340587	08/09/2020	Summer	R1	Illumina MiSeq
5	C1_sep20_summer_r2	SAMN60340588	08/09/2020	Summer	R2	Illumina MiSeq
6	C1_sep20_summer_r3	SAMN60340589	08/09/2020	Summer	R3	Illumina MiSeq
7	C1_jan21_winter_r1	SAMN60340590	18/01/2021	Winter	R1	Illumina MiSeq
8	C1_jan21_winter_r2	SAMN60340591	18/01/2021	Winter	R2	Illumina MiSeq
9	C1_jan21_winter_r3	SAMN60340592	18/01/2021	Winter	R3	Illumina MiSeq
10	C1_apr21_spring_r1	SAMN60340593	08/04/2021	Spring	R1	Illumina MiSeq
11	C1_apr21_spring_r2	SAMN60340594	08/04/2021	Spring	R2	Illumina MiSeq
12	C1_jun20_summer_pool	SAMN60340595	25/06/2020	Summer	Pool	Illumina NovaSeq 6000
13	C1_sep20_summer_pool	SAMN60340596	08/09/2020	Summer	Pool	Illumina NovaSeq 6000
14	C1_jan21_winter_pool	SAMN60340597	18/01/2021	Winter	Pool	Illumina NovaSeq 6000
15	C1_apr21_spring_pool	SAMN60340598	08/04/2021	Spring	Pool	Illumina NovaSeq 6000
16	C1_jul21_summer_pool	SAMN60340599	15/07/2021	Summer	Pool	Illumina NovaSeq 6000
17	C1_oct21_autumn_pool	SAMN60340600	22/10/2021	Autumn	Pool	Illumina NovaSeq 6000
18	C1_jan22_winter_pool	SAMN60340601	19/01/2022	Winter	Pool	Illumina NovaSeq 6000
19	C1_may22_spring_pool	SAMN60340602	03/05/2022	Spring	Pool	Illumina NovaSeq 6000
20	C1_jul22_summer_pool	SAMN60340603	25/07/2022	Summer	Pool	Illumina NovaSeq 6000
21	C1_dec22_autumn_pool	SAMN60340604	14/12/2022	Autumn	Pool	Illumina NovaSeq 6000
22	C1_apr23_spring_pool	SAMN60340605	05/04/2023	Spring	Pool	Illumina NovaSeq 6000
23	C1_jul23_summer_pool	SAMN60340606	13/07/2023	Summer	Pool	Illumina NovaSeq 6000
24	C1_jan24_winter_pool	SAMN60340607	23/01/2024	Winter	Pool	Illumina NovaSeq 6000
25	C1_apr24_spring_pool	SAMN60340608	08/04/2024	Spring	Pool	Illumina NovaSeq 6000
26	C1_nov24_autumn_pool	SAMN60340609	08/11/2024	Autumn	Pool	Illumina NovaSeq 6000
27	C1_may25_spring_pool	SAMN60340610	14/05/2025	Spring	Pool	Illumina NovaSeq 6000

taxonomic distribution remained generally consistent across time points, with small changes between years and seasons.

Focusing on specific taxa (Fig. 2), Cyanobacteriota's relative abundance ranged from 0.03% to 0.4%, with higher values recorded in July 2021 and November 2024 compared to the lowest value in April 2021 (Fig. 2a). At genus level (Fig. 2c), *Synechococcus* CC9902 was the most abundant taxon, with a mean relative abundance of 0.1 ± 0.1%, and was consistently detected across all seasons. The second most abundant genus was *Cyanobium* PCC-6307, with a mean relative abundance of 0.02 ± 0.03%, which varied across seasons, making it less temporally stable. Phyla belonging to the Archaea domain accounted for a higher relative abundance within the community, ranging from 1.59% in May 2022 to 8.7% in November 2024 (Fig. 2b). As antic-

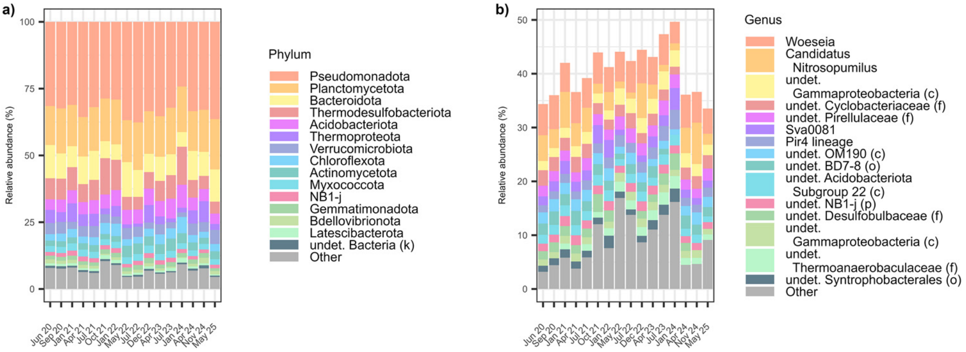

Fig. 1. Temporal changes in the relative abundance of the prokaryotic community composition of surface sediment samples at the phylum (a) and genus level (b; genera with relative abundance ≥ 1% in at least one sample). When taxonomic assignment is undetermined, the highest known taxonomic resolution is reported in brackets: kingdom (k), phylum (p), class (c), order (o) and family (f).

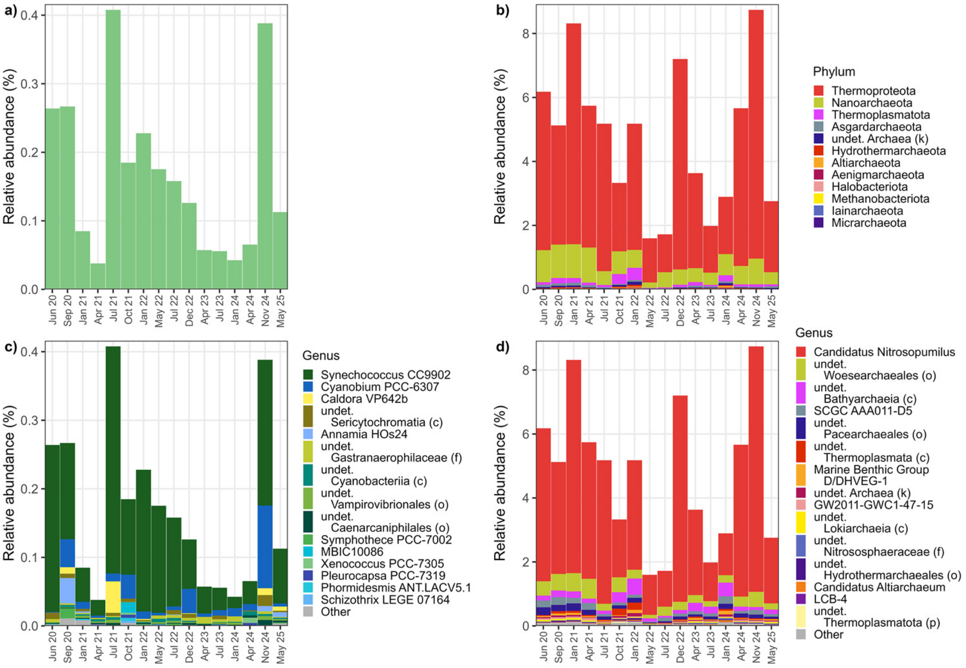

Fig. 2. Temporal changes in the relative abundance of the prokaryotic community composition of surface sediment samples, focusing on the phylum Cyanobacteriota (phylum: a; genus: c) and on the domain of Archaea (phylum: b; genus: d). When taxonomic assignment is undetermined, the highest known taxonomic resolution is reported in brackets: kingdom (k), phylum (p), class (c), order (o) and family (f).

ipated, Candidatus *Nitrosopumilus* was one of the most represented genera among samples and the most abundant within the Archaea domain, characterising the community across the five years (Fig 2d).

When considering the whole prokaryotic community, alpha diversity analysis (i.e., Shannon diversity) indicated no significant differences in surface sediment samples within seasons (Kruskal–Wallis test: $p = 0.145$). Likewise, non-metric multidimensional scaling (NMDS) based

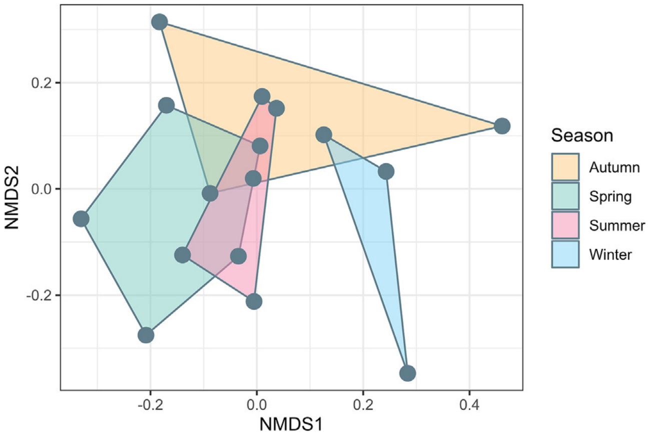

Fig. 3. Plot of non-metric multidimensional scaling (NMDS) ordination of surface sediment samples based on Bray–Curtis dissimilarities; polygons connect samples from the same astronomical season (stress = 0.123).

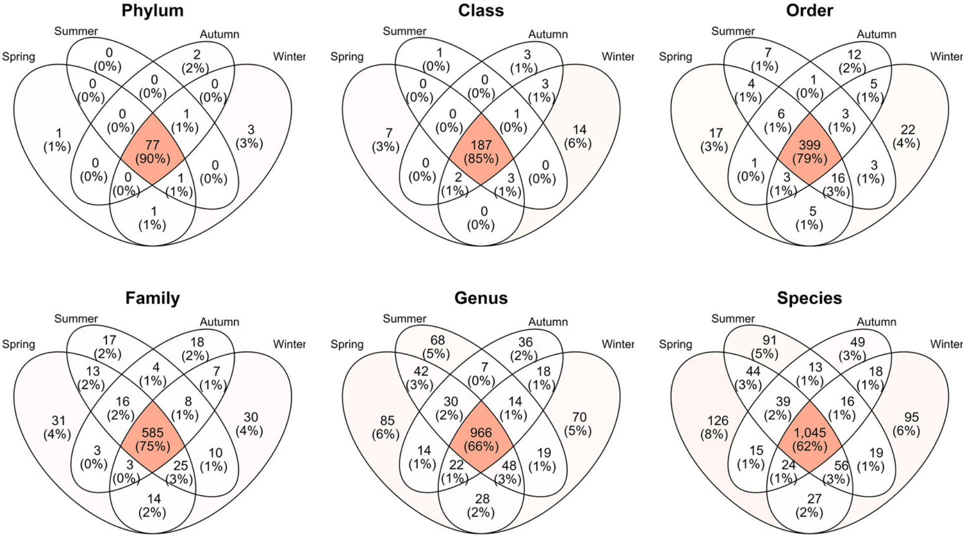

Fig. 4. Venn diagrams showing the overlap of taxa among seasons at each taxonomic level: phylum, class, order, family, genus and species. Darker colour indicates a higher proportion of shared taxa.

on Bray–Curtis dissimilarity showed considerable overlap among samples from different seasons, confirming a relatively stable community structure across seasons and years. The pattern was also confirmed by the results of the PERMANOVA analysis, which revealed no significant differences among seasons (PERMANOVA 9999 permutations, $R^2 = 0.25$, $p = 0.078$) (Fig. 3).

Taxonomic resolution illustrated by Venn diagrams showed the overlap of taxa among seasons (Fig. 4). At the phylum level, almost 90% of phyla were shared among seasons, with winter retaining the highest proportion of unique taxa (3%). At the class level, 85% of the taxa were shared, followed by order at 79%, family at 75%, and genus at 66%. At the species level, 62% of taxa were shared across all four seasons, with spring and winter registering the most unique species at 8% and 6%, respectively. Finally, to provide an environmental context for the compositional data presented in this article, the seasonal distribution across the years 2010–2019 [2,7] of temperature, salinity, dissolved inorganic macronutrients in bottom waters, as well as total nitrogen, total organic carbon, carbohydrates, proteins, and lipid contents have been reported in Figures S2 and S3.

4. Experimental Design, Materials and Methods

4.1. Sample collection, DNA extraction and sequencing

A total of 16 sediment samples were collected seasonally from June 2020 to May 2025 at the time points reported in Table 1 from C1-LTER station (45°42′2.99″N, 13°42′36.00″E) (DEIMS.iD https://deims.org/96969205-cfdf-41d8-979f-ff881ea8dc8b) from the Gulf of Trieste (Italy), located in the Adriatic Sea (northeastern Mediterranean Sea).

At each time point, three biological replicates consisting of three separated surface sediment cores were collected using a KC Haps bottom corer (KC-Denmark, Silkeborg, Denmark) with polycarbonate sample tubes (outside diameter 133 mm/internal diameter 127 mm, sampling surface of 127 cm^2) as reported in the protocol implemented by Franzo et al. [2]. In the laboratory, the cores were partially extruded, and the top sediment layers were collected, one replicate for each core, and immediately preserved at −80 °C until downstream analyses. Sediments were processed as detailed by Banchi et al. [8]. For each replicate, DNA extraction was performed using the DNeasy PowerSoil Pro Kit (Qiagen) from an average of 0.55 ± 0.04 g of sediment and subsequently quantified with a Qubit fluorimeter (Thermo Fisher Scientific). Extracted DNA samples for the Illumina MiSeq System platform were sequenced separately in two or three replicates (Table 1), whereas the three replicates of extracted DNA samples assigned to NovaSeq sequencing were pooled prior to sequencing (Table 1).

Then, the V4-V5 region of the 16S rRNA gene was amplified using the 515F-Y (5′-GTGYCAGCMGCCGCGGTAA-3′) and the 926R (5′-CCGYCAATTYMTTTRAGTTT-3′) primers [9], and libraries were prepared following the 16S Metagenomic Sequencing Library Preparation protocol. Firstly, surface sediment cores from the first four samplings (from June 2021 to April 2021; Table 1) were separately sequenced on an Illumina MiSeq System for a read length of 2 × 300 bp at Cogentech (Consortium for Genomic Technologies c/o IFOM-IEO Campus, Milan, Italy). To assess the reliability and robustness of the replicates, a correlation analysis (Spearman's ρ) was performed on the amplicon sequence variants (ASVs) table across the three replicates of each of the samples (two for April 2021). Within each sample, the replicates were significantly correlated ($p < 0.05$) with an average ρ of 0.62 ± 0.05, indicating an overall consistent community structure across biological replicates. As described above, the DNA extracted from the three replicates of each sample from June 2021 to May 2025 was pooled and sequenced using an Illumina Novaseq 6000 System for a read length of 2 × 250 bp at the Institute of Applied Genomics (IGA, Udine, Italy).

4.2. Bioinformatics processing

DNA metabarcoding analyses were performed as follows. Raw sequences obtained from the sequencing facility were first checked for quality with FastQC (v. 0.12.1) [10] and MultiQC (v. 1.29) [11]. Raw sequences were then quality filtered and denoised with DADA2 v. 1.36.0 [12] in R (v. 4.5.2) [13] using the pseudo-pooling method . Primers were removed with cutadapt v. 4.0 [14]. Reads were filtered (maxEE = c(2,2), truncQ = 2, maxN = 0) and trimmed (truncLen = c(220,220)), then denoised, merged using default DADA2 parameters. For Novaseq 6000 data, error rate learning was determined by modifying the *loess* function and enforcing monotonicity to account for binned quality scores. Chimeric sequences were identified and removed using the consensus method. Taxonomy was assigned to amplicon sequence variants (ASVs) using the sklearn naiv̈e Bayes taxonomy classifier [15] against the Silva 99% reference database with seven-level taxonomy (v. 138.2) [16]. ASVs belonging to Eukarya, mitochondria, chloroplasts, or with a frequency <2 (singletons) were removed.

4.3. Statistical analysis

Statistical analyses were performed in R (v. 4.5.2) [13]. Correlation analysis (Spearman's ρ) was performed using base functions from the *stats* package. ASV counts were not rarefied. For taxonomic composition analyses, ASV counts were transformed to relative abundances (%). For NMDS ordination, samples were transformed to relative abundances before calculating Bray–Curtis dissimilarities. Bar plots of temporal changes in the taxonomic composition of surface sediments and NMDS ordination plot were generated using *ggplot2* (v. 4.0.2) [17]. Alpha diversity (Shannon index) was calculated using the *phyloseq* package (v. 1.36.0) [18], while beta diversity was assessed using the *vegan* package (v. 2.7–2) [19]. Venn diagrams were generated using the *ggVennDiagram* package (v. 1.5.7) [20]. Graphs reported in the Supplementary Materials were generated from the FastQC [10] and MultiQC [11] output files using *ggplot2* (v. 4.0.2) [17].

Limitations

The study presents some limitations concerning the sampling design and metadata availability. Sampling was not performed consistently across years, leading to uneven temporal coverage, with some seasons missing in certain years. Due to logistical and operational constraints, sampling campaigns were carried out in different months, although efforts were made to preserve the astronomical seasonality of the sampling design. Additionally, samples lacked complete associated environmental metadata. However, long-term (2020–2019; [2,7]) seasonal trends of environmental variables at the sampling station in the Gulf of Trieste are provided in the Supplementary Material (Figure S2 and Figure S3).

Ethics Statement

The authors have read and follow the ethical requirements for publication in Data in Brief and confirm that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

CRediT Author Statement

Noemi Tomasi: Data curation; Formal analysis, Writing - Original Draft; **Elisa Banchi**: Data curation, Formal analysis, Writing - Review & Editing, Supervision; **Vincenzo Manna**: Data curation, Writing - Review & Editing, Supervision; **Mauro Celussi:** Conceptualization, Funding acquisition, Writing - Review & Editing, Supervision.

Data Availability

PRJNA1442017 Surface sediment 16S rRNA amplicon sequencing of the northern Adriatic Sea, Jun 20 - May 25 (Original data) (SRA - NCBI).

Acknowledgments

Samples were collected from the site "Gulf of Trieste (DEIMS.ID: https://deims.org/96969205-cfdf-41d8-979f-ff881ea8dc8b)" which belongs to the Long Term Ecological Research national and international networks (LTER-Italy, LTER-Europe and ILTER). The data generation was partially funded by the Friuli Venezia Giulia autonomous region (LR 15/2005) and by the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4–Call for tender No. 3138 of 16 December 2021, rectified by Decree n.3175 of 18 December 2021 of the Italian Ministry of University and Research funded by the European Union–NextGenerationEU; Project code CN_00000033, Concession Decree No. 1034 of 17 June 2022 adopted by the Italian Ministry of University and Research, CUP D33C22000960007, Project title "National Biodiversity Future Center–NBFC". We thank the Spazzamare crew and OGS staff for their contribution to sampling activities.

Declaration of Competing Interest

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

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2026.112971.

References

[1] P.K. Bech, K.L. Lysdal, L. Gram, M. Bentzon-Tilia, M.L. Strube, Marine sediments hold an untapped potential for novel taxonomic and bioactive bacterial diversity, MSystems 5 (2020), doi:10.1128/msystems.00782-20.

[2] A. Franzo, M. Celussi, M. Bazzaro, F. Relitti, P. Del Negro, Microbial processing of sedimentary organic matter at a shallow LTER site in the northern Adriatic Sea: an 8-year case study, Nat. Conserv. 34 (2019) 397–415, doi:10.3897/natureconservation.34.30099.

[3] M. Huettel, P. Berg, J.E. Kostka, Benthic exchange and biogeochemical cycling in permeable sediments, Ann. Rev. Mar. Sci. 6 (2014) 23–51, doi:10.1146/annurev-marine-051413-012706.

[4] S. Miksch, M. Meiners, A. Meyerdierks, D. Probandt, G. Wegener, J. Titschack, M.A. Jensen, A. Ellrott, R. Amann, K. Knittel, Bacterial communities in temperate and polar coastal sands are seasonally stable, ISME Commun. 1 (2021), doi:10.1038/s43705-021-00028-w.

[5] S. Ahmerkamp, C. Winter, K. Krämer, D. de Beer, F. Janssen, J. Friedrich, M.M.M. Kuypers, M. Holtappels, Regulation of benthic oxygen fluxes in permeable sediments of the coastal ocean, Limnol. Oceanogr. 62 (2017) 1935–1954, doi:10.1002/lno.10544.

[6] A. Tagliabue, K.O. Matterson, M. Ponti, E. Turicchia, M. Abbiati, F. Costantini, Sediment and bottom water eDNA metabarcoding to support coastal management, Ocean Coast. Manag. 244 (2023) 106785, doi:10.1016/j.ocecoaman.2023.106785.

[7] T. Cibic, L. Baldassarre, F. Cerino, C. Comici, D. Fornasaro, M. Kralj, M. Giani, Benthic and pelagic contributions to primary production: experimental insights from the Gulf of Trieste (Northern Adriatic Sea), Front. Mar. Sci. 9 (2022), doi:10.3389/fmars.2022.877935.

[8] E. Banchi, P. Del Negro, M. Celussi, F. Malfatti, Sediment features and Human activities structure the surface microbial communities of the Venice lagoon, Front. Mar. Sci. 8 (2021), doi:10.3389/fmars.2021.762292.

[9] A.E. Parada, D.M. Needham, J.A. Fuhrman, Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples, Environ. Microbiol. 18 (2016) 1403–1414, doi:10.1111/1462-2920.13023.

[10] S. Andrews, FastQC: a quality control tool for high throughput sequence data [Online], 2010. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

[11] P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report, Bioinformatics 32 (2016) 3047–3048, doi:10.1093/bioinformatics/btw354.

[12] B.J. Callahan, P.J. McMurdie, M.J. Rosen, A.W. Han, A.J.A. Johnson, S.P. Holmes, DADA2: high-resolution sample inference from Illumina amplicon data, Nat. Method. 13 (2016) 581–583, doi:10.1038/nmeth.3869.

[13] R Core Team, R: a language and environment for statistical computing, 2025. https://www.R-project.org/.

[14] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnet. J. 17 (2011) 10.

[15] N.A. Bokulich, M.R. Dillon, Y. Zhang, J.R. Rideout, E. Bolyen, H. Li, P.S. Albert, J.G. Caporaso, q2-longitudinal: longitudinal and paired-sample analyses of microbiome data, MSystems 3 (2018), doi:10.1128/msystems.00219-18.

[16] C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F.O. Glöckner, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools, Nucleic Acids Res. 41 (2013) D590–D596, doi:10.1093/nar/gks1219.

[17] H. Wickham, ggplot2, WIREs Comput. Statistic. 3 (2011) 180–185, doi:10.1002/wics.147.

[18] P.J. McMurdie, S. Holmes, phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data, PLoS One 8 (2013) e61217, doi:10.1371/journal.pone.0061217.

[19] J. Oksanen, G.L. Simpson, F.G. Blanchet, R. Kindt, P. Legendre, P.R. Minchin, R.B. O'Hara, P. Solymos, M.H.H. Stevens, E. Szoecs, H. Wagner, M. Barbour, M. Bedward, B. Bolker, D. Borcard, T. Borman, G. Carvalho, M. Chirico, M. De Caceres, S. Durand, H.B.A. Evangelista, R. FitzJohn, M. Friendly, B. Furneaux, G. Hannigan, M.O. Hill, L. Lahti, C. Martino, D. McGlinn, M.-H. Ouellette, E. Ribeiro Cunha, T. Smith, A. Stier, C.J.F. Ter Braak, J. Weedon, vegan: community Ecology Package, (2025). https://doi.org/10.32614/CRAN.package.vegan.

[20] C.-H. Gao, A. Dusa, ggVennDiagram: a “ggplot2” implement of Venn Diagram, (2026). https://doi.org/10.32614/CRAN.package.ggVennDiagram.