Next-generation monitoring & mapping tools to assess marine ecosystems & biodiversity

Deliverable D2.4

Taxonomic and functional characterization of the examined biocommunities

Greece 2.0
NATIONAL RECOVERY AND RESILIENCE PLAN

Funded by the European Union
NextGenerationEU

This project is carried out within the framework of the National Recovery and Resilience Plan Greece 2.0, funded by the European Union – NextGenerationEU (Implementation body: HFRI).

DOCUMENT INFORMATION AND VERSION CONTROL

Project Acronym	NEMO-Tools
Project Title	Next-generation monitoring and mapping tools to assess marine ecosystems and biodiversity
Project Number	016035
Work Package	WP2
Related Task(s)	T2.3
Deliverable Number	D2.4
Deliverable Name	Taxonomic and functional characterization of the examined biocommunities
Due Date	14 December 2025
Date Delivered	
Dissemination Level	Public — fully open (automatically posted online on the Project Results platforms)

VERSION CONTROL

Revision-N°	Date	Description	Prepared By	Reviewed By
	20/11/2025	1st Draft	S. Genitsaris	C. Gubili
		Final Draft		

Executive Summary

This Deliverable 2.4 – "Taxonomic and functional characterization of the examined biocommunities" – presents results from a large-scale environmental DNA (eDNA) survey conducted across three coastal regions of Greece (Saronikos, Thermaikos, and Kavala) and four offshore sites in the Aegean Sea. The study aimed to assess microbial and fish biodiversity and compare the performance of three eDNA sampling methods: Sterivex filtration, active filtering on nucleopore filters, and passive gauze collection. Sampling included surface waters and the lower euphotic zone to evaluate spatial and vertical variation in marine microbial communities. Using 16S (bacteria), 18S (eukaryotes) and 12S (fishes) rRNA gene sequencing, the analysis revealed high diversity across all regions, with offshore waters showing the highest richness and community evenness.

Alpha diversity indices confirm that Thermaikos and Saronikos coastal waters also host diverse bacterial assemblages, while Kavala showed comparatively lower richness. Across all areas, Alphaproteobacteria were the dominant bacterial class, with Cyanobacteria, Gammaproteobacteria, and Bacteroidia consistently present as secondary key groups. Community structure varied significantly by location, reflecting environmental gradients and coastal influences. Eukaryotic communities were more variable and generally less diverse than bacteria. Diversity patterns were strongly shaped by dominance events, often driven by blooms of dinoflagellates and other planktonic taxa. Offshore samples displayed the highest eukaryotic richness and evenness, while Saronikos exhibited the most frequent low-diversity dominance signatures. After removing multicellular taxa from analysis, unicellular phytoplankton groups, particularly Dinophyceae, diatoms, haptophytes, and picoalgae, emerged as key components of the microbial food web. Similarly, Thermaikos and Saronikos gulfs host more diverse fish assemblages than Kavala, with the latter showing the lowest species richness. Anchovies and sardines dominated most stations. Species of the families Sparidae, Scombridae, Mugilidae, Gobidae, and Callionymidae were also identified.

Functional profiles were similar among gulfs, indicating environmental gradients driving community functions. The dominance of metabolic pathways related to carbon turnover and biosynthesis suggests active processing of organic matter across all regions. Collectively, the data indicate stable metabolic capability across the gulfs with minor localized enhancements tied to specific biogeochemical processes.

Comparison of sampling methodologies showed that Sterivex and active filtration methods produced more consistent and reliable biodiversity profiles, particularly for eukaryotes. Gauze sampling yielded the highest variability and occasional under-representation of microbial diversity, indicating its reduced suitability for quantitative biodiversity monitoring.

Overall, the findings demonstrate clear spatial structuring of marine communities in the Aegean Sea, shaped by geography, sampling depth, and environmental conditions, but common functional roles dominating. Thus, NEMO-Tools provides a

robust baseline dataset for future biomonitoring efforts and supports eDNA as a scalable method for high-resolution taxonomic and functional biodiversity assessment across marine ecosystems.

TABLE OF CONTENTS

CONTRIBUTORS

TABLE 1 NAMES AND ROLES OF CONTRIBUTORS TO THIS DELIVERABLE.

Name	Affiliation	WP Lead	Task Lead
Savvas Genitsaris	National and Kapodistrian University of Athens		2.3
Chrysoula Gubili	Hellenic Agricultural Organization – DIMITRA - Fisheries Research Institute	2	
Alexandra Zachariadou	National and Kapodistrian University of Athens		
Panagiota Xanthopoulou	Hellenic Agricultural Organization – DIMITRA - Fisheries Research Institute		
Antonios Mazaris	Aristotle University of Thessaloniki		

1. Introduction

Environmental DNA (eDNA) monitoring has emerged as a powerful tool for characterizing biodiversity in aquatic ecosystems, enabling detection of organisms across trophic levels without relying on traditional sampling or visual identification. This deliverable describes the outputs of eDNA metabarcoding towards assessing microbial and fish taxonomic diversity and community functional diversity across key marine regions of the Aegean Sea, including three coastal gulfs (Saronikos, Thermaikos, and Kavala) and offshore waters, by combining multiple sampling methods and sequencing approaches targeting bacterial (16S rRNA), eukaryotic (18S rRNA), fish (12S rRNA) communities and functional profiles (shotgun metagenomics). The sampling, bioinformatic and ecological modelling pipelines are described in detail in previous WP2 deliverables. The current D2.4 deliverable aims to: (1) establish a baseline of microbial biodiversity across contrasting marine environments, (2) evaluate spatial and depth-related ecological patterns, (3) summarize the functional profiles of the entire communities in multi-stressored coastal areas; and (3) assess the performance and consistency of commonly used eDNA collection techniques. The results contribute to advancing marine biomonitoring capability in the region and support the integration of eDNA tools into long-term ecological assessment and management frameworks.

2. Methods

Detailed methods are described in D2.1 for samplings and D2.2 for downstream processing of samples.

Briefly, water sampling was carried out in three coastal gulfs (Saronikos, Thermaikos, and Kavala) each surveyed during a dedicated cruise. In every gulf, six stations were selected to capture spatial variability, and samples were collected from two depths: the surface and the lower boundary of the euphotic zone. At each station, three eDNA sampling methods were employed to assess potential differences in marine microbial biodiversity: active filtering with a vacuum pump, syringe sterivex filtering, and gauze passive sampling. For the active and Sterivex filtrations, seawater was first collected with a Niskin bottle, sieved through a 200 μm mesh to exclude large metazoans, and then filtered through 0.2 μm pore-size polycarbonate (Nuclepore) or Sterivex filters, respectively. For the passive filtration, a gauze sheet was mounted on a perforated plastic frame and immersed at the sampling site for approximately 10 minutes, allowing seawater to flow through. In addition to the coastal sites, four offshore stations across the Aegean Sea were sampled during separate cruises. Offshore samples were collected only with the Filter method at the surface. All filters and gauze samples were immediately stored at −20 °C until DNA extraction. Additional 0.5 L water samples were collected for phytoplankton abundance and biomass measurements, preserved with Lugol's iodine solution, and stored in the dark until microscopic analysis. In situ abiotic parameters, including temperature (T), salinity

(S), dissolved oxygen (DO), and pH, were measured using a CTD multiparameter probe at each station and depth.

In the lab, the eDNA was extracted using the Macherey-Nagel NucleoSpin® Soil kit (Macherey-Nagel) and the DNeasy PowerWater Kit (Qiagen) according to the manufacturer's instructions, and the concentration and quality of the recovered DNA was confirmed using a ThermoScientific™ NanoDrop™ spectrophotometer and a Qubit 4 Fluorometer (Invitrogen). PCR amplification targeted the following marker regions. The V3–V4 hypervariable region of the 16S rRNA gene for bacterioplankton, using primers S-DBact-0341-b-S-17 (CCTACGGGNGGCWGCAG) and S-D-Bact-0785-a-A-21 (GACTACHVGGGTATCTAATCC) (Klindworth et al., 2013), and the V4 region of the 18S rRNA gene for planktonic protists, using primers E572F (CYGCGGTAATTCCAGCTC) and E1009R (AYGGTATCTRATCRTCTTYG) (Comeau et al., 2011). Additionally, a ~170 bp barcoding region of the 12S rRNA was amplified using the fish specific MiFish primer set (Forward: 5' -GTCGGTAAAACTCGTGCCAGC- 3'; Reverse: 5' -CATAGTGGGGTATCTAATCCCAGTTTG- 3'; Miya et al., 2015).

Table 1. Location and characteristics of the sampling sites across the study area

Location	Sampling Site	Sampling point code	Depth	Method	Latitude	Longtitude
Saronikos	S1	**S_S1s**	surface	Sterivex	37.9929	23.5482
		S_S1d	deep	Sterivex		
		F_S1s	surface	Filter		
		F_S1d	deep	Filter		
		G_S1s	surface	Gauze		
		G_S1d	deep	Gauze		
	S2	**S_S2s**	surface	Sterivex	37.8979	23.5471
		S_S2d	deep	Sterivex		
		F_S2s	surface	Filter		
		F_S2d	deep	Filter		
		G_S2s	surface	Gauze		
		G_S2d	deep	Gauze		
	S3	**S_S3s**	surface	Sterivex	37.8503	23.433
		S_S3d	deep	Sterivex		
		F_S3s	surface	Filter		
		F_S3d	deep	Filter		
		G_S3s	surface	Gauze		
		G_S3d	deep	Gauze		
	S4	**S_S4s**	surface	Sterivex	37.9028	23.3803
		S_S4d	deep	Sterivex		
		F_S4s	surface	Filter		
		F_S4d	deep	Filter		
		G_S4s	surface	Gauze		
		G_S4d	deep	Gauze		
	S5	**S_S5s**	surface	Sterivex	37.981	23.4139
		S_S5d	deep	Sterivex		
		F_S5s	surface	Filter		

		F_S5d	deep	Filter		
		G_S5s	surface	Gauze		
		G_S5d	deep	Gauze		
	S6	**S_S6s**	surface	Sterivex	38.0132	23.4955
		S_S6d	deep	Sterivex		
		F_S6s	surface	Filter		
		F_S6d	deep	Filter		
		G_S6s	surface	Gauze		
		G_S6d	deep	Gauze		
Thermaikos	T1	**S_T1s**	surface	Sterivex	40.5469	22.7625
		S_T1d	deep	Sterivex		
		G_T1s	surface	Gauze		
		G_T1d	deep	Gauze		
		F_T1d	deep	Filter		
		F_T1s	surface	Filter		
	T2	**S_T2s**	surface	Sterivex	40.6209	22.8958
		S_T2d	deep	Sterivex		
		G_T2s	surface	Gauze		
		G_T2d	deep	Gauze		
		F_T2d	deep	Filter		
		F_T2s	surface	Filter		
	T3	**S_T3s**	surface	Sterivex	40.6248	22.938
		S_T3d	deep	Sterivex		
		G_T3s	surface	Gauze		
		G_T3d	deep	Gauze		
		F_T3d	deep	Filter		
		F_T3s	surface	Filter		
	T4	**S_T4s**	surface	Sterivex	40.5652	22.954
		S_T4d	deep	Sterivex		
		G_T4s	surface	Gauze		
		G_T4d	deep	Gauze		
		F_T4d	deep	Filter		
		F_T4s	surface	Filter		
	T5	**S_T5s**	surface	Sterivex	40.5139	22.8481
		S_T5d	deep	Sterivex		
		G_T5s	surface	Gauze		
		G_T5d	deep	Gauze		
		F_T5d	deep	Filter		
		F_T5s	surface	Filter		
	T6	**S_T6s**	surface	Sterivex	40.3729	22.9285
		S_T6d	deep	Sterivex		
		G_T6s	surface	Gauze		
		G_T6d	deep	Gauze		
		F_T6d	deep	Filter		
		F_T6s	surface	Filter		
Kavala	K1	**S_K1s**	surface	Sterivex	40.7291	24.2547
		S_K1d	deep	Sterivex		
		G_K1s	surface	Gauze		

		G_K1d	deep	Gauze		
		G2_K1s	surface	Gauze		
		G2_K1d	deep	Gauze		
		F_K1d	deep	Filter		
		F_K1s	surface	Filter		
	K2	**S_K2s**	surface	Sterivex	40.731	24.2116
		S_K2d	deep	Sterivex		
		G_K2s	surface	Gauze		
		G_K2d	deep	Gauze		
		F_K2d	deep	Filter		
		F_K2s	surface	Filter		
	K3	**S_K3s**	surface	Sterivex	40.7722	24.3521
		S_K3d	deep	Sterivex		
		G_K3s	surface	Gauze		
		G_K3d	deep	Gauze		
		F_K3d	deep	Filter		
		F_K3s	surface	Filter		
	K4	**S_K4s**	surface	Sterivex	40.817	24.4126
		S_K4d	deep	Sterivex		
		G_K4s	surface	Gauze		
		G_K4d	deep	Gauze		
		G2_K4s	surface	Gauze		
		G2_K4d	deep	Gauze		
		F_K4d	deep	Filter		
		F_K4s	surface	Filter		
	K5	**S_K5s**	surface	Sterivex	40.836	24.3559
		S_K5d	deep	Sterivex		
		G_K5s	surface	Gauze		
		G_K5d	deep	Gauze		
		F_K5d	deep	Filter		
		F_K5s	surface	Filter		
	K6	**S_K6s**	surface	Sterivex	40.838	24.3212
		S_K6d	deep	Sterivex		
		G_K6s	surface	Gauze		
		G_K6d	deep	Gauze		
		F_K6d	deep	Filter		
		F_K6s	surface	Filter		
Offshore	A1	**A1**	surface	Filter	37.594635	23.528138
	DK3	**DK3s**	surface	Filter	36.656816	27.139265
	DK4	**DK4s**	surface	Filter	36.55826	27.17592
	DS3	**DS3s**	surface	Filter	36.456274	25.410807

Amplicons were sequenced on an Illumina MiSeq platform using 2 × 300 bp paired-end chemistry (16S and 18S) and an Illumina NovaSeq 6000 platform (12S), which provides overlapping reads that can be merged into full-length sequences.

Microbial raw amplicon reads were processed using mothur v1.34.0 following the MiSeq SOP (Schloss et al., 2009, 2011). Taxonomic assignment was performed using

SINA searches against the SILVA database for bacteria (Pruesse et al., 2012) and the classify.seqs command in mothur against the Protist Ribosomal Reference (PR2) database for eukaryotes , applying a minimum similarity threshold of 80% to the closest reference.

Alpha diversity indices, including OTU richness (S), Shannon (H′), Simpson (1 – D), Pielou's evenness (J′), and Chao1 richness estimator, were computed for each sample using the phyloseq package (McMurdie & Holmes, 2013) in R v4.4.1 (R Core Team, 2021). Calculations were based on untransformed read counts. The Simpson index is a measure of dominance and calculates the probability that two randomly selected reads will belong to the same OTU; Shannon is an indicator of evenness; and Pielou's index, calculated as the ratio of H to the natural logarithm of the OTUs richness (ln(S)), quantifies how evenly distributed OTUs are within a community. SChao1 estimates the expected richness in a sample by correcting for unseen taxa based on low-abundance OTUs. For each domain we converted counts to relative abundance per sample, and visualized composition with stacked bar charts. To improve readability, we retained Classes exceeding 1% for bacteria and 2% for eukaryotes in at least one sample. For a eukaryote sensitivity view focused on unicellular taxa, we repeated summaries after excluding exclusively multicellular Classes. We further produced heatmaps of Class-level relative abundance using the same 1–2% inclusion thresholds to emphasize dominant Classes. Community dissimilarity was computed using Bray–Curtis on relative abundances and visualized with NMDS.

Furthermore, in selected samples, metagenomic shotgun sequencing was performed and raw reads were processes with MEGAHIT (Li et al., 2015) and Prokka (Seemann, 2014) towards amino acid fasta generation that were used to retrieve KEGG mapped functional profiles with the GhostKoala blast function (Kanehisa et al., 2016).

Multiplexed fastq files were processed through the MJOLNIR pipeline (https://github.com/uit-metabarcoding/MJOLNIR) and the OBITools package (Boyer et al. 2016). Paired-end reads were aligned using *illuminapairedend*, retaining only those with quality scores above 40. Demultiplexing and primer-sequence removal were performed by *ngsfilter* and paired-end alignment was implemented by *illuminapairedend*. Filtering and dereplication of sequences were performed with *obigrep* and obiuniq, respectively, retaining sequences of 140–190 bp length. Singleton sequences and chimeric amplicons were eliminated using the *uchime_denovo* algorithm from VSEARCH (Rognes et al. 2016). Molecular operational taxonomic units (MOTUs) delimitation was implemented using the SWARM procedure (Mahe et al. 2015) and putative pseudogenes were removed using LULU (Froslev et al. 2017).

Multiple indices, namely Richness, Shannon and Simpson were computed with the package vegan in R (Oksanen et al. 2022) to evaluate fish diversity per sampling area and method. Non-metric multidimensional scaling (NMDS) ordinations, using the *metaMDS* function in the vegan package, provided a reduced-space graphical representation of the species composition per site. Distance matrices were calculated

using the Jaccard coefficient (function *vegdist*). Permutational analysis of variance (PERMANOVA) was performed to assess the influence of the sampling location, method, and depth on the composition of fish communities. PERMANOVA was conducted using the *adonis2* function in the *vegan* package and using presence/absence data with 10,000 permutations.

3. Results

3.1. Bacterial communities

3.1.1. Alpha diversity

In total, the bacterial communities exhibited a wide range of alpha diversity (Table 2, Figure 1) across different locations, sampling methods and depths. Observed OTU richness (the count of unique OTUs) ranged from as low as only 2 OTUs in one extreme case in Thermaikos (G_T3s) to as high as 1214 OTUs in an offshore sample (DK4s). Correspondingly, the Chao1 estimated richness generally mirrored these patterns, with the highest Chao1 (≈1302) in that highly diverse offshore sample and similarly low Chao1 in samples with very few observed OTUs. The considerable spread between observed richness and Chao1 in some samples (e.g., Filter samples from Kavala) suggests that certain communities had many taxa undetected in sequencing (i.e., sampling may have been incomplete in those cases).

Shannon diversity indices for bacteria were generally high across most samples, indicating substantial diversity and evenness. Shannon values ranged from 0.69 (in the extremely low-richness Gauze sample with 2 OTUs) up to 5.61 in the most diverse offshore sample. For context, Shannon values above ~4 are exceptionally high for 16S rRNA gene surveys, reflecting very complex communities (Nagai et al., 2024). Most samples had Shannon indices in the 3.0–4.5 range, and Simpson indices were likewise high in the majority of cases (Simpson's index >0.9 in many samples), meaning that communities were not dominated by only one or two taxa but rather had a more even distribution. Evenness (Pielou's evenness) values mostly fell between 0.60–0.75 for many bacterial samples, with a few notable extremes: the lowest evenness (~0.47) occurred in a Saronikos Gauze surface sample (G_S2s) that was relatively dominated by a few OTUs, whereas the highest evenness (~0.90) was observed in some low-richness samples (e.g. Gauze K4 surface) where even a small number of OTUs were present in nearly equal abundance.

Spatial differences among the three coastal areas (Kavala, Thermaikos, Saronikos) and the offshore station were also observed. Overall, Thermaikos Gulf showed the highest bacterial diversity, with consistently elevated richness, Shannon, and Simpson indices across all sampling methods and depths. Several Thermaikos samples presented exceptionally high richness (notably G_T4d (742 OTUs), G_T3d (621 OTUs) and G_T5d (486 OTUs)) coupled with high evenness (Evenness > 0.73) and very low dominance (Simpson > 0.95), indicating rich and well-balanced bacterial communities. Bacterial communities from Kavala generally showed the lowest diversity: Kavala samples had an average OTU richness around 240 and Shannon ~3.5,

and none of the Kavala samples reached the extreme high diversity seen elsewhere. In contrast, Thermaikos and Saronikos had higher diversity on average. Saronikos showed somewhat intermediate-to-high diversity (average Shannon ~3.7), though it was marked by heterogeneity. Some Saronikos samples were very diverse (e.g. S_S1d, Shannon ~4.02; G_S6d, Shannon ~4.81), while others were relatively lower (S_S2d, Shannon ~3.45). The most striking result spatially was that the offshore samples exhibited exceptionally high diversity. Despite only a few offshore stations sampled (A1, DK3s, DK4s, DS3s), all done with the Filter method, they contributed the highest richness (over 1000 OTUs) and high average Shannon (~3.98) among all regions. This suggests that the open-water offshore environment, while oligotrophic, harbors a very broad diversity of bacteria.

Depth-related patterns in bacterial alpha diversity were present but not uniformly strong. In many station pairs, the deep samples had equal or higher diversity than the surface samples from the same location. For instance, in Thermaikos, deep Sterivex samples often had higher Shannon diversity than their surface counterparts (e.g. S_T1d Shannon 4.26 vs S_T1s 3.90; S_T3d 4.53 vs S_T3s 3.65). Similarly, deep Gauze samples at some stations had more OTUs and higher diversity (G_T2d had ~536 OTUs, Shannon 4.70, compared to 481 OTUs, Shannon 4.38 in G_T2s). These suggest that the deeper water in certain locations harbored a slightly more diverse or even bacterial community, possibly due to more stable conditions or different nutrient profiles. However, there were also cases where surface diversity was on par with or higher than deep. In Kavala Sterivex samples, for example, surface and deep Shannon indices were quite close (e.g. S_K5s 3.53 vs S_K5d 3.78), and at one Thermaikos site (station T2) the surface Sterivex S_T2s, actually had much higher richness (285 OTUs) than the deep S_T2d (116 OTUs), though the deep sample's evenness was higher. On average, deep samples showed slightly higher Shannon (mean ~3.77) than surface (mean ~3.59), but the overlap was considerable. This indicates that depth alone was a minor factor compared to method and site-specific conditions in shaping bacterial alpha diversity.

Table 2. Alpha diversity indices of the bacterial communities in each sample. See Table 1 for sampling code annotations.

	Bacteria				
Sampling points	**OTUs richness**	**SChao1**	**SHANNON (H)**	**Simpson (1-D)**	**Pielou's J**
S_K1S	162	176.286	3.457	0.942	0.679
S_K1D	263	270.692	3.567	0.925	0.640
S_K2S	179	200.120	3.475	0.942	0.670
S_K2D	304	314.328	3.386	0.840	0.592
S_K3S	166	176.500	3.492	0.945	0.683
S_K3D	230	235.385	3.546	0.935	0.652
S_K4S	168	178.000	3.460	0.941	0.675
S_K4D	203	219.917	3.542	0.939	0.667
S_K5S	204	212.784	3.526	0.936	0.663
S_K5D	319	337.638	3.779	0.949	0.655
S_K6S	122	131.500	3.224	0.913	0.671
S_K6D	278	309.098	3.627	0.933	0.645
S_T1S	361	413.560	3.900	0.955	0.662
S_T1D	345	368.263	4.263	0.973	0.729
S_T2S	285	309.522	3.514	0.912	0.622
S_T2D	116	149.158	3.193	0.921	0.672
S_T3S	316	353.078	3.654	0.921	0.635
S_T3D	379	412.889	4.528	0.978	0.763
S_T4S	360	388.049	3.899	0.953	0.662
S_T4D	353	387.263	3.999	0.959	0.682
S_T5S	303	331.923	3.857	0.940	0.675
S_T5D	349	381.571	4.323	0.977	0.738
S_T6S	229	241.353	3.700	0.955	0.681
S_T6D	314	373.167	3.781	0.930	0.658
S_S1S	179	214.000	3.601	0.954	0.694
S_S1D	334	345.679	4.021	0.968	0.692
S_S2S	234	268.167	3.707	0.945	0.679
S_S2D	264	283.568	3.447	0.928	0.618
S_S3S	233	247.167	3.557	0.930	0.652
S_S3D	269	294.738	3.469	0.921	0.620
S_S4S	249	265.528	3.693	0.943	0.669
S_S4D	197	217.647	3.518	0.890	0.666
S_S5S	295	321.102	3.584	0.934	0.630
S_S5D	272	301.167	3.702	0.948	0.660
S_S6S	274	291.727	3.499	0.941	0.623
S_S6D	230	247.000	3.459	0.943	0.636
G_K1S	56	62.563	3.461	0.955	0.860
G_K1D	160	218.800	4.365	0.981	0.860
G_K2S	207	234.029	4.122	0.971	0.773
G_K2D	354	401.878	4.051	0.954	0.690
G_K3S	208	232.474	4.126	0.970	0.773
G_K3D	261	304.063	4.101	0.964	0.737

G_K4S	34	36.625	3.163	0.946	0.897
G_K4D	19	30.250	2.396	0.874	0.814
G_K5S	373	399.233	4.528	0.978	0.765
G_K5D	415	438.558	4.538	0.974	0.753
G_K6S	354	367.722	4.217	0.952	0.718
G_K6D	386	414.493	4.549	0.978	0.764
G_T1S	504	536.530	4.184	0.958	0.672
G_T1D	565	585.613	4.377	0.946	0.691
G_T2S	481	498.755	4.382	0.971	0.710
G_T2D	536	564.030	4.699	0.976	0.748
G_T3S	2	3.000	0.693	0.500	1.000
G_T3D	621	645.797	4.803	0.982	0.747
G_T4S	510	525.697	4.240	0.962	0.680
G_T4D	742	761.689	5.145	0.982	0.778
G_T5S	31	38.000	2.921	0.925	0.851
G_T5D	486	523.979	4.268	0.954	0.690
G_T6S	419	433.412	4.405	0.970	0.730
G_T6D	490	511.093	4.303	0.945	0.695
G_S1S	355	374.500	4.350	0.973	0.741
G_S1D	382	404.479	4.344	0.966	0.731
G_S2S	248	260.750	2.611	0.728	0.474
G_S2D	232	256.600	2.881	0.777	0.529
G_S3S	588	609.122	4.076	0.943	0.639
G_S3D	160	175.030	3.844	0.949	0.757
G_S4S	321	337.533	4.180	0.957	0.724
G_S4D	344	383.487	4.291	0.964	0.735
G_S5S	1947	1967.448	4.971	0.981	0.656
G_S5D	131	146.000	4.154	0.975	0.852
G_S6S	321	339.192	4.202	0.969	0.728
G_S6D	1027	1039.917	4.808	0.984	0.693
G2_K1S	272	284.000	3.507	0.921	0.626
G2_K1D	297	311.875	3.748	0.919	0.658
G2_K4S	261	272.667	3.498	0.922	0.629
G2_K4D	174	181.906	3.353	0.908	0.650
F_K1D	316	371.018	3.126	0.881	0.543
F_K1S	211	257.941	2.986	0.886	0.558
F_K2D	316	350.597	2.648	0.809	0.460
F_K2S	225	292.385	2.871	0.881	0.530
F_K3D	273	326.333	3.188	0.896	0.568
F_K3S	210	277.031	2.855	0.881	0.534
F_K4D	267	344.083	3.153	0.891	0.564
F_K4S	200	259.000	2.974	0.896	0.561
F_K5D	315	387.818	3.147	0.896	0.547
F_K5S	229	288.094	3.026	0.908	0.557
F_K6D	297	345.226	3.233	0.910	0.568
F_K6S	253	334.162	2.872	0.886	0.519
F_S1D	367	476.019	3.945	0.964	0.668
F_S1S	271	334.088	3.534	0.947	0.631
F_S2D	303	380.233	3.164	0.889	0.554

F_S2S	304	395.935	3.347	0.903	0.586
F_S3D	293	334.553	3.122	0.888	0.550
F_S3S	318	379.620	3.359	0.904	0.583
F_S4D	308	374.071	3.300	0.896	0.576
F_S4S	307	378.250	3.607	0.936	0.630
F_S5D	325	377.684	3.232	0.902	0.559
F_S5S	296	361.395	3.591	0.941	0.631
F_S6D	293	350.791	3.525	0.943	0.621
F_S6S	256	312.163	3.392	0.937	0.612
F_T1D	357	444.208	3.987	0.958	0.678
F_T1S	330	441.625	3.282	0.914	0.566
F_T2D	391	472.182	3.979	0.938	0.667
F_T2S	336	396.857	3.784	0.946	0.650
F_T3D	358	416.882	3.834	0.924	0.652
F_T3S	333	388.208	3.804	0.955	0.655
F_T4D	334	408.754	3.587	0.942	0.617
F_T4S	300	359.702	3.262	0.911	0.572
F_T5D	353	450.500	3.712	0.923	0.633
F_T5S	299	356.000	3.248	0.917	0.570
F_T6D	317	372.350	3.141	0.892	0.545
F_T6S	223	268.024	3.154	0.910	0.583
A1	196	287.833	3.424	0.937	0.649
DK3S	332	446.231	3.408	0.900	0.587
DK4S	1214	1302.440	5.615	0.984	0.791
DS3S	346	496.955	3.466	0.927	0.593

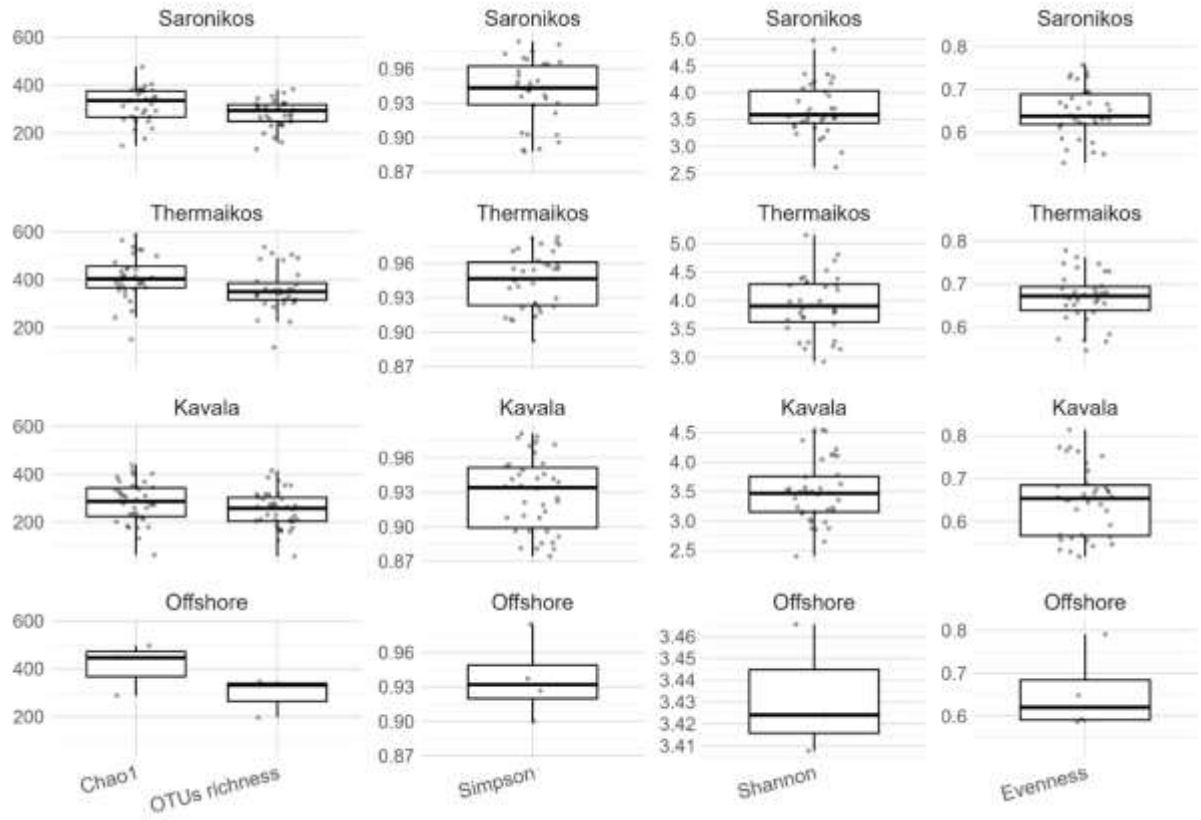

Figure 1. Boxplots of the distribution of the alpha-diversity indices of bacterial communities in the examined locations. The horizontal black bars depict the median values, and the dots depict individual samples. Outliers were filtered per-index using 1.5xIQR thresholds.

Methodological differences had a clear impact on bacterial alpha diversity. The Sterivex, Gauze, and Filter methods yielded different diversity outcomes. The Sterivex samples generally showed moderate to high diversity with observed OTU richness typically in the low-to-mid hundreds (e.g., 160–360 OTUs in coastal samples). Shannon indices for Sterivex samples were consistently around 3.2–4.3, and Simpson indices were high (often >0.93), indicating very diverse and fairly even communities. Sterivex tended to give reliable and intermediate richness values without the extreme lows or highs seen in other methods. For example, Sterivex samples from Thermaikos had Shannon ~3.9–4.3, higher than many Filter samples from the same area, but Sterivex never reached the absolute maximum richness observed by Gauze or Filter in certain cases.

The Gauze samples showed the highest variability. In some cases, Gauze captured extraordinarily high richness. For instance, a Saronikos surface Gauze sample (G_S5s) contained ~1947 observed OTUs with Shannon ~4.97, one of the richest bacterial communities observed. Several other Gauze samples, especially from Thermaikos and Saronikos, had Shannon diversity well above 4, indicating that when Gauze filtration was effective, it retained a very broad spectrum of bacteria. However, Gauze also yielded the lowest values in the dataset for a few samples. Notably, one Thermaikos surface Gauze sample (G_T3s) recovered only 2 OTUs (Shannon 0.69, Simpson 0.5), pointing to a potential sampling anomaly or extremely low biomass in that particular

sample. Similarly, initial Gauze trials at some Kavala stations (e.g. K1 and K4) had surprisingly low richness (34 OTUs in G_K4s), which prompted duplicate extraction (denoted as "G2" codes) that then yielded more typical richness values (e.g. 261 OTUs in G2_K4s). On average, Gauze filtration tended to detect slightly more OTUs than Sterivex, but with less consistency. Gauze samples often had very high evenness (up to ~0.86) in the successful samples, contributing to high Shannon scores, whereas failed/low-yield Gauze samples showed artificially high evenness values of ~1.0 (because only one or two taxa were present in equal abundance).

Filter samples generally showed similar or slightly higher OTU richness and Chao1 estimates compared with Sterivex across most stations. This indicates that the active filtration method was efficient in retaining a broad range of bacterial taxa, including rarer lineages. For the diversity indices (Shannon, Simpson, Evenness), Filter values were overall comparable to Sterivex, though often marginally lower. These modest differences suggest that Filter samples tended to include a larger number of taxa but with a slightly more uneven distribution of abundances, where a few OTUs contributed more strongly to the community profile. Importantly, these differences were not systematic or large, and substantial overlap between methods was observed across many stations.

In summary, bacterial alpha diversity was influenced by a combination of methodology, environmental site, and depth. Spatially, the coastal sites Thermaikos and Saronikos showed higher diversity than Kavala, and offshore waters were remarkably diverse. Depth differences were modest, with a tendency for slightly higher diversity in deeper waters. The Gauze method yielded the highest richness in successful samples (likely due to filtering larger volumes), but also suffered occasional under-sampling. Sterivex and Filter approaches produced largely comparable a-diversity patterns for bacterial communities. Filters often recovered equal or higher richness, while Sterivex samples tended to display slightly higher evenness, but the magnitude of these differences was generally small. Both methods performed consistently across gulfs and depths, suggesting that they capture the underlying bacterial taxonomic structure in similar ways, with Filter showing less dispersion overall. Nonetheless, across all samples the high Simpson indices (mostly 0.92–0.98) indicate that most bacterial communities were relatively even and species-rich, reflecting the complex microbial ecosystems present in these Mediterranean water bodies.

3.1.2. Relative abundance trends

Across all sampled locations, Alphaproteobacteria was the most dominant bacterial class. It accounted for roughly one-third of the community on average and frequently exceeded 50% of the sequences in individual samples. By comparison, the next most abundant groups overall were Cyanobacteriia and Gammaproteobacteria, though their contributions varied markedly with location (Figure 2).

In the Kavala samples, Alphaproteobacteria consistently dominated the community. On average this class comprised ~37–38% of the bacterial sequences. In Kavala for example, a surface water sample F_K5s contained about 42.6% Alphaproteobacteria. Gammaproteobacteria and Cyanobacteriia were the next most abundant classes in Kavala, each contributing roughly 20% on average. Notably, Gammaproteobacteria showed high variability among Kavala samples: in most cases it was a secondary class

(~10–20%), but in at least one deep sample (G_K2d) it surged to ~56.5% of the community. Cyanobacteriia generally made up around 15–25% in Kavala (e.g. ~19% in F_K5s) and tended to co-occur as a substantial fraction alongside the proteobacteria. Bacteroidia were present at more moderate levels (~5–10% in Kavala). Meanwhile, members of the phylum Actinobacteriota were relatively more prominent in Kavala than elsewhere – for instance, the class Acidimicrobiia reached ~22% in one surface sample (F_K6s), whereas actinobacterial classes were near or below 2% in the other regions. Overall, Kavala's bacterial profile was heavily dominated by Alphaproteobacteria, with Gammaproteobacteria and Cyanobacteriia as important sub-dominant groups.

The Thermaikos Gulf samples showed a similar pattern by Alphaproteobacteria, though to a slightly lesser degree than Kavala. Alphaproteobacteria made up roughly one-third of the community on average in Thermaikos (~30–32%). For example, in a deep sample F_T2d, Alphaproteobacteria constituted ~42.5%. Cyanobacteriia were the second-major group in Thermaikos (around 20% on average), followed by Bacteroidia and Gammaproteobacteria which each contributed 15% of sequences. Verrucomicrobiae also formed a noticeable portion (~5–10%) in Thermaikos, comparable to their levels in Saronikos. The Thermaikos community thus featured a prominent Alphaproteobacteria dominance with Cyanobacteria, Bacteroidia, and Gammaproteobacteria each making meaningful secondary contributions. As in Kavala, Alphaproteobacteria sometimes reached very high proportions (>50% in certain samples), whereas Gammaproteobacteria and Bacteroidia generally remained in the tens of percent. For instance, Bacteroidia occasionally contributed to ~16–24% in individual Thermaikos samples (e.g. sample T1d about 24% Bacteroidia) but were usually lower than the Cyanobacteriia fraction.

The Saronikos Gulf exhibited a more balanced community among the top classes. Alphaproteobacteria was still the single most abundant class in Saronikos, but at a lower relative abundance (~28–29% on average) compared to the other areas. Notably, Cyanobacteriia and Bacteroidia were highly represented in Saronikos, often approaching or even exceeding the Alphaproteobacteria in individual samples. For example, a deep Saronikos sample (F_S5d) contained ~36.1% Cyanobacteriia alongside ~40.2% Alphaproteobacteria. In another Saronikos sample (F_S1s), Bacteroidia reached ~26.0% of the community while Alphaproteobacteria were ~37–38%. Overall, Saronikos had the largest contributions of Bacteroidia (averaging ~18%) among the regions. Gammaproteobacteria were somewhat lower in Saronikos (around 10–14% on average), and Verrucomicrobiae were on par with Thermaikos at ~10%. This made the Saronikos bacterial community composition more evenly spread across Alphaproteobacteria, Cyanobacteriia, and Bacteroidia, rather than being overwhelmingly dominated by a single class.

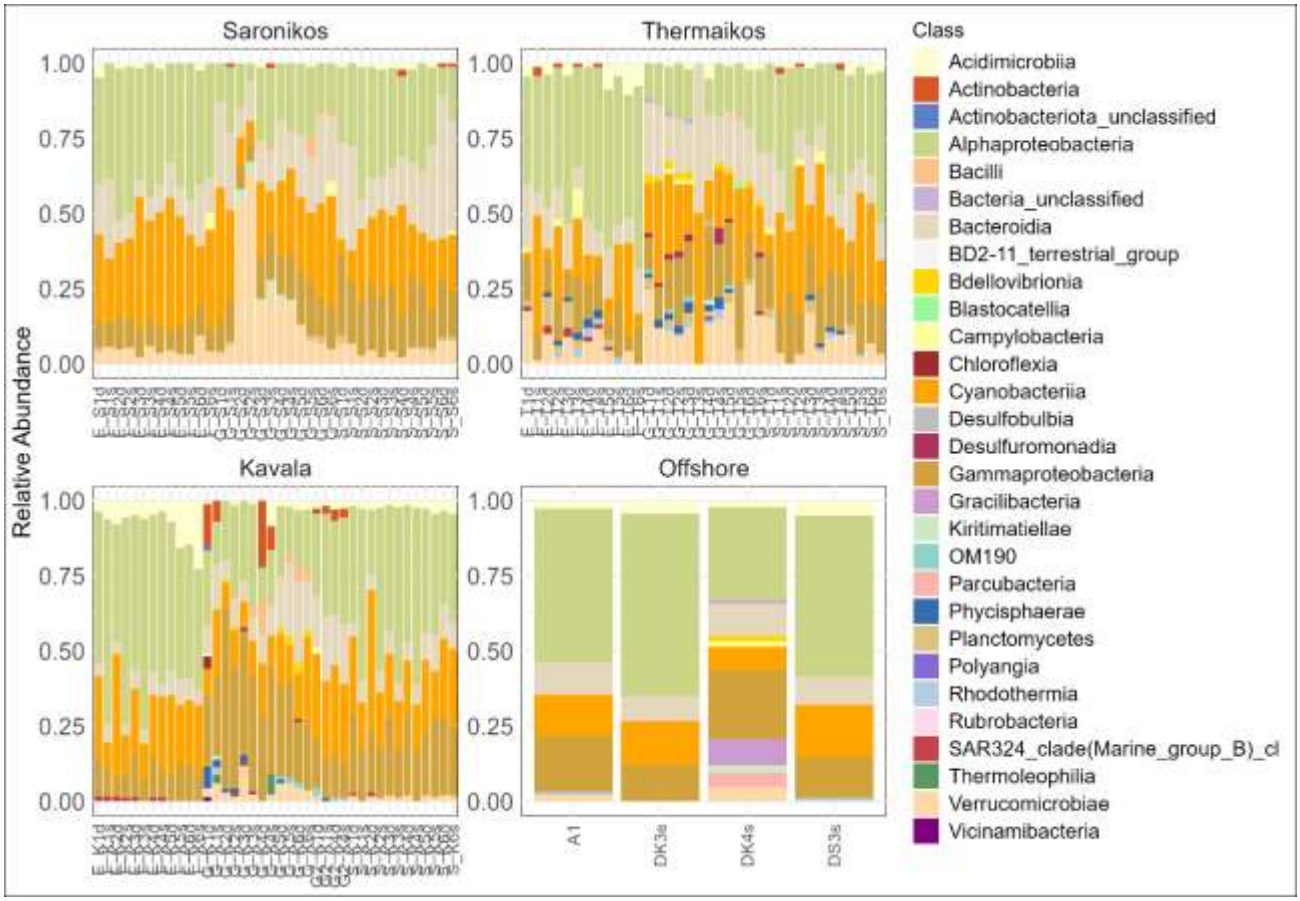

Figure 2. Bar plots showing the relative abundance of higher taxonomic bacterial groups per Location. The taxonomic groups contributing >1 % of the total number of reads in the sample are plotted.

The offshore Aegean samples were characterized by an extreme dominance of Alphaproteobacteria. On average, Alphaproteobacteria comprised nearly half of the total bacterial community offshore (~48% mean). Many offshore samples showed Alphaproteobacteria in the 50–60% range. For instance, a surface offshore sample (DS3s) was ~52% Alphaproteobacteria. The next most abundant classes offshore were the Gammaproteobacteria and Cyanobacteriia, but at much lower levels (generally ~10–20% each). A deep offshore sample A1, for example, had about 18.2% Gammaproteobacteria and ~13.6% Cyanobacteriia. Bacteroidia were present at ~5–10% in the offshore stations, similar to their levels in Kavala. One notable feature of the offshore bacterial communities was the presence of Gracilibacteria, formed a measurable fraction offshore (for example, ~8.3% in sample DK4s) whereas this group was virtually absent in Kavala, Thermaikos, and Saronikos samples. Overall, the offshore sites had the simplest community profile, being dominated by Alphaproteobacteria with moderate contributions from Gamma-proteobacteria and Cyanobacteria, and with unique minor groups like Gracilibacteria appearing only in these open-sea samples.

Despite Alphaproteobacteria dominance in all areas, there were clear differences in community structure between the coastal gulfs and the open-sea stations. Alphaproteobacteria maintained the highest relative abundance in every location but reached its peak in the offshore waters (nearly 50% on average, vs. ~30–38% in the

coastal locations). Coastal sites, especially Saronikos and Thermaikos, showed elevated levels of Cyanobacteriia and Bacteroidia compared to the offshore. Saronikos and Thermaikos each had roughly 15–25% contributions from these groups, whereas offshore waters had lower Cyanobacteria (~13%) and Bacteroidia (~9%) proportions. Similarly, Verrucomicrobiae were consistently present at around 10% in Saronikos/Thermaikos, but were only minor components in Kavala and essentially negligible offshore. Kavala stood out for a higher presence of Actinobacteria relative to the other coastal sites, suggesting some unique inputs or conditions there. In contrast, the offshore microbiome contained unique taxa like Gracilibacteria, not seen in the gulfs, indicating a distinct community in open waters. In summary, while all locations shared the feature of Alphaproteobacteria being the dominant class, the secondary bacterial assemblages differed. Thermaikos and Saronikos harbored greater proportions of Cyanobacteria and Bacteroidetes, Kavala had a noticeable Actinobacteria presence, and the offshore stations featured a simplified community with few Bacteroidetes but some unique pelagic bacteria (e.g. Patescibacteria).

Taken together, these region-specific signatures are consistent with broader ecological patterns in coastal marine systems. Alphaproteobacteria dominate across all environments but coastal gulfs exhibit enhanced contributions from Gammaproteobacteria, Cyanobacteria, and Bacteroidia, groups well known to respond to nutrient enrichment, phytoplankton blooms, and particle-rich conditions typical of eutrophic waters (Parulekar et al., 2017; Lo et al., 2022; Yang et al., 2024).

Across all four regions, the bacterial communities share a common dominant core composed of Cyanobacteria, Gammaproteobacteria, Alphaproteobacteria, and Bacteroidia, which consistently appear in every sampling point and represent the most abundant classes throughout the study area (Figure 3). These taxa form the fundamental core of the coastal microbiome. In addition to this highly persistent Classes, several other groups such as Verrucomicrobiae, Acidimicrobia, Actinobacteria, Rhodothermia, Planctomycetes, and Phycisphaerae are present across almost all stations, though generally at lower abundance, shaping the secondary layer of the community structure.

Despite these shared patterns, each gulf exhibits distinct compositional features. In Saronikos, the core assemblage dominated by Cyanobacteria, Gamma- and Alphaproteobacteria, Bacteroidia, and consistently abundant Verrucomicrobiae is clearly visible. Acidimicrobia, Actinobacteria, Rhodothermia, Planctomycetes, and Phycisphaerae are also present in all samples but at more moderate levels, completing the main structure of the community. A few groups appear only sporadically, including Blastocatellia in samples G_S2s, G_S3s, and S_S4d, and Vicinamibacteria in G_S5s, S_S2s, and S_S4d, while classes such as Chloroflexia, Parcubacteria, Gracilibacteria, and Polyangia appear in only a handful of stations. Unclassified bacterial reads remain low overall, although they show elevated values in a restricted number of samples such as G_S2s, G_S2d, S_S1s, S_S2s, S_S2d, and S_S5d. Overall, the Saronikos community is highly coherent across stations, with strong dominance of the main coastal classes and occasional appearances of rare candidate phyla.

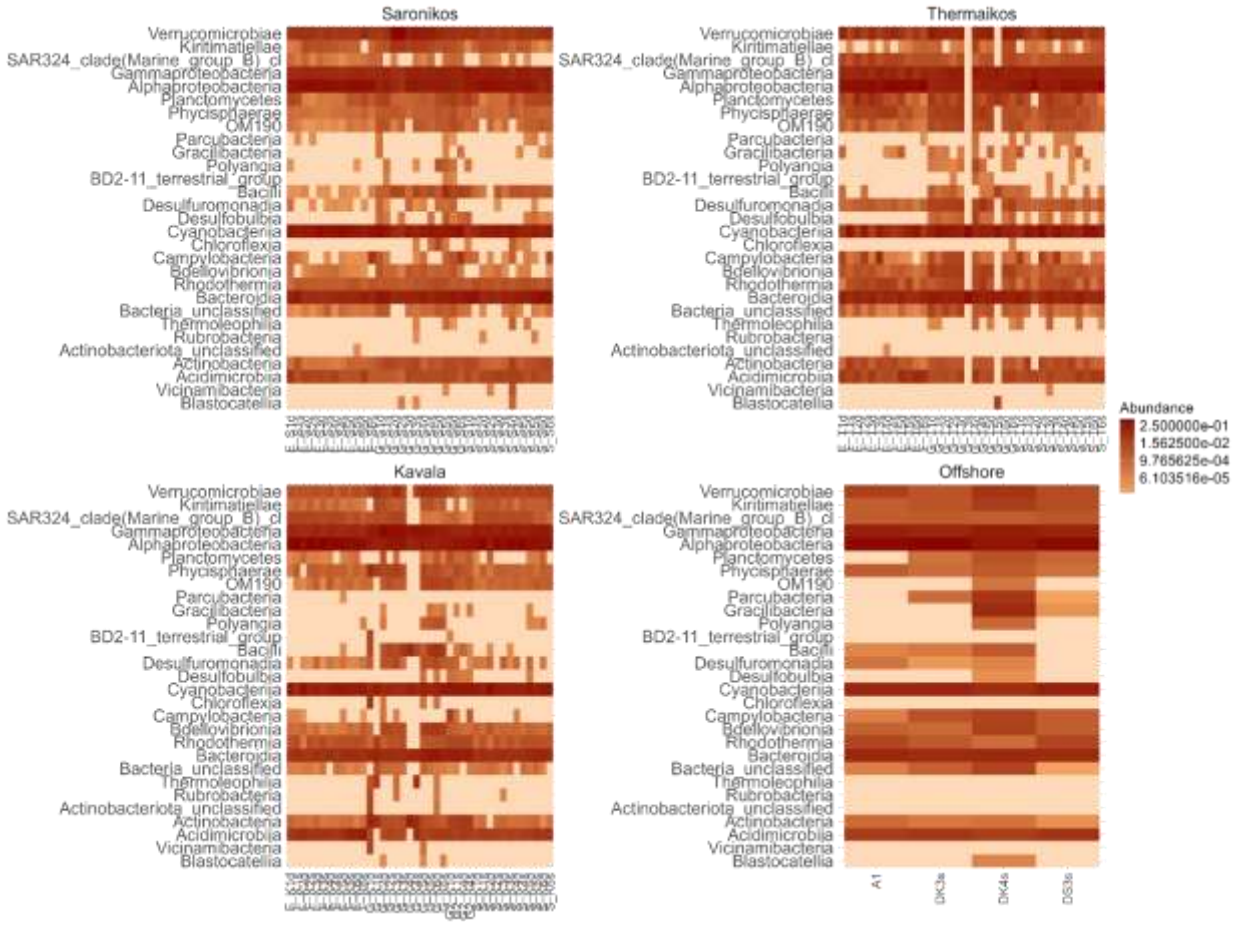

Figure 3. Heatmap of bacterial Class-level, log4 transformed relative abundance, showing only Classes >1% in at least one sample.

Thermaikos exhibits a more heterogeneous profile, characterized by the same dominant core but a broader spread of other Classes. Parcubacteria, Polyangia, and Gracilibacteria appear more frequently and at higher abundances compared to Saronikos. Bacilli are present in fewer stations, while Thermoleophilia appear in both regions but are detected in a larger number of stations in Thermaikos. Rubrobacteria, which occur in five Saronikos samples, are entirely absent from Thermaikos. Planctomycetes and Phycisphaerae contribute strongly to the Thermaikos community and are more abundant and consistent there. Planctomycetes are absent from only one sample (S_T2d). Verrucomicrobiae are widespread but missing from G_T5s. Rhodothermia, though absent from two samples (G_T3s and G_T5s), shows higher average abundance in Thermaikos compared to Saronikos. Chloroflexia is represented only in G_T6s. The more frequent appearance of candidate phyla and particle-associated lineages in Thermaikos suggests a more diverse and possibly more dynamic environmental setting than in Saronikos.

Kavala maintains the same dominant taxa but with alterations in distribution and abundance across Classes. Actinobacteria are absent from only one sample (S_K2d), whereas Acidimicrobia are present in all samples and often at relatively high abundance, making them more prominent contributors than in Saronikos or Thermaikos. Thermoleophilia appear only in a few stations (G_K1d, G_K1s, G_K3d, G_K4s), and Rubrobacteria appear in G_K3d, G_K1d, and G_K6d. Planctomycetes

and Phycisphaerae are present across many stations, with Phycisphaerae showing broader coverage. Verrucomicrobiae are abundant in every Kavala sample, reinforcing their status as a core taxon in this gulf. Rhodothermia are also present in every sample, which distinguishes Kavala from the other gulfs. Chloroflexia appear in several deep samples such as G_K1d, G_K2d, G_K5d, and G_K6d, being particularly abundant in G_K1d. Altogether, Kavala exhibits a distinct signature enriched in Acidimicrobia, Actinobacteria, and Rhodothermia, differentiating the region from Saronikos and Thermaikos.

Offshore samples share the same core structure dominated by Alphaproteobacteria, Gammaproteobacteria, Cyanobacteria, Bacteroidia, and Acidimicrobia. A variety of other classes including Verrucomicrobiae, Phycisphaerae, Rhodothermia, unclassified bacteria, Actinobacteria, Campylobacteria, and Bdellovibrionia are present in all samples, though with lower abundance. Some classes show specific absences, such as Bacilli and Desulfuromonadia in DS3s. Gracilibacteria, Polyangia, and Parcubacteria occur with very high abundance in DK4s, whereas Gracilibacteria and Parcubacteria appear at much lower levels in DS3s. Parcubacteria also appear in DK3s. The offshore communities resemble Saronikos and Kavala more than Thermaikos but exhibit occasional enrichment in candidate phyla, reflecting the more open-water conditions.

Across all regions, several classes show consistent geographical gradients. Bdellovibrionia reach their highest levels in Thermaikos, followed by Kavala and then Saronikos. Campylobacteria are more widespread and persistent in Thermaikos, moderately represented in Saronikos but often missing from gauze-collected samples, and least represented in Kavala. Although the dominant taxa remain stable across locations, these finer-scale differences in rare and supporting taxa reveal subtle but ecologically meaningful distinctions between the gulfs.

3.1.3. Beta diversity

The NMDS ordination of bacterial communities reveals a clear location-driven structuring of beta diversity across samples (Figure 4). The four geographic groups, Kavala, Saronikos, Thermaikos, and Offshore, form well-defined color-coded clusters, indicating pronounced differences in community composition among gulfs. This spatial separation is the strongest signal in the ordination and dominates the overall pattern.

Within these location-based clusters, a secondary but weaker structuring effect is associated with sampling method. Filter and Sterivex samples tend to group more closely, forming overlapping sub-clusters within each gulf. This pattern is consistent with their similar capture characteristics and with earlier observations showing that these two methods recover broadly comparable bacterial communities. In contrast, Gauze samples (triangular symbols) are consistently positioned farther from the main method-consistent clusters in all three gulfs. This dispersion likely reflects their more variable performance, which in several cases generated unusually high or low values for specific taxa, thereby causing strong divergence in beta diversity space.

When focusing exclusively on Filter and Sterivex samples, the tightest geographic grouping appears in Kavala, where most samples cluster very closely regardless of

method, suggesting relatively homogeneous bacterial communities within this gulf. Saronikos also shows a compact core cluster, more spread compared to Kavala, but it also includes several peripheral samples from both methods that deviate from the main grouping, indicating localized variability among sampling points. Thermaikos, in contrast, forms a more dispersed cluster overall, with Gauze samples diverging the most strongly, reinforcing the method-driven spread within this gulf.

The Offshore samples, all collected with the Filter method, fall within the broader Filter/Sterivex grouping but clearly associate more strongly with the Kavala cluster than with Saronikos or Thermaikos. One notable exception is sample DK4, which is positioned at the extreme edge of the ordination. Its distinct placement suggests that its bacterial community differs substantially from the other offshore samples, likely due to unique local environmental conditions or compositional shifts not captured elsewhere.

Overall, the NMDS results indicate that location is the primary determinant of bacterial beta diversity, with sampling method, especially Gauze, introducing additional within-gulf variation. The consistency between Filter and Sterivex samples strengthens confidence in cross-site comparisons, while the broader dispersion of Gauze samples highlights the need for careful interpretation when integrating results across methods.

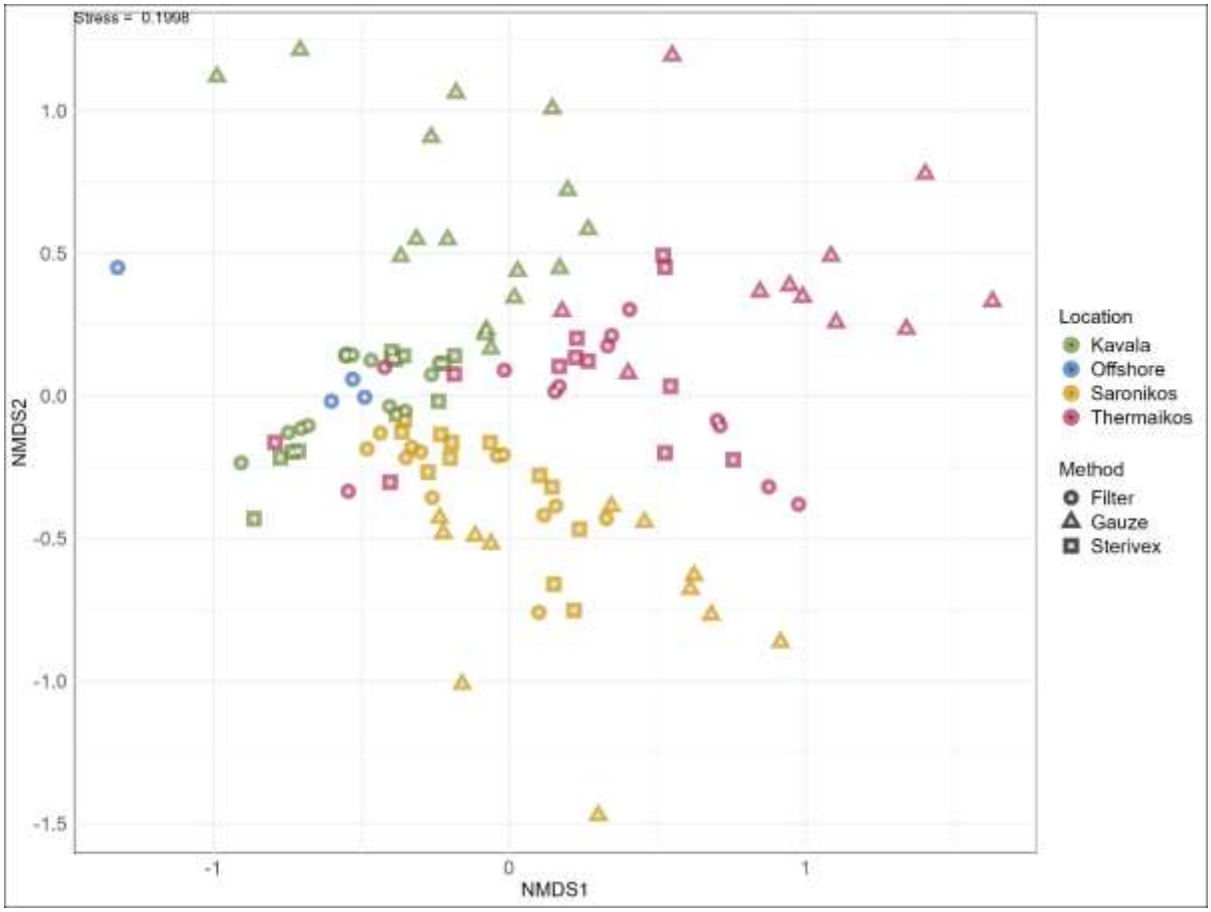

Figure 4. Non-metric Multidimensional Scaling (nMDS) ordination plot of bacterial OTUs relative abundances showing the relatedness of sampling points based on Bray-Curtis dissimilarities. Different colors indicate the locations from which the samples were collected and different shapes the different sampling method.

3.2. Unicellular eukaryotes

3.2.1. Alpha diversity

Alpha diversity of microbial eukaryotic communities (Table 3, Figure 5) was generally lower than that of bacteria, as expected, but it also showed significant variation among samples. Observed OTU richness for eukaryotes ranged from just 7–10 OTUs in the most depauperate samples up to nearly 500 OTUs in the richest samples. Several coastal samples had extremely low eukaryotic richness: for example, a deep Saronikos Gauze sample (G_S2d) recovered only 8 OTUs, and a surface Saronikos Sterivex (S_S1s) had just 19 OTUs. On the high end, a few samples approached or exceeded 400 OTUs (e.g. Sterivex Kavala S_K5d had 494 OTUs, and an offshore Filter sample DS3s had 484 OTUs). Chao1 estimates again followed suit, with some samples showing much higher Chao1 than observed (indicating many singletons). For instance, Sterivex S_K3d (Kavala deep) had 268 observed OTUs but Chao1 ~364, suggesting a lot of taxa possibly missed. Similarly, many Filter samples had Chao1 estimates 1.3–1.5 times their observed OTUs, reflecting under sampling. On the other hand, in cases of low richness (like G_S2d), Chao1 was not much higher than observed, implying that those communities truly were dominated by very few taxa.

The Shannon diversity indices for eukaryotes were lower on average than those for bacteria. Many eukaryote samples had Shannon values in the range of 1 to 3, and a number of samples even had Shannon < 1.0, indicating extreme dominance by one or few taxa. The lowest Shannon values occurred in Saronikos and Thermaikos filters: for example, a Thermaikos surface Filter sample (F_T5s) had Shannon ~0.62, and a Saronikos deep Filter (F_S1d) had Shannon only 0.57. These values mean that essentially one species comprised the vast majority of the community in those samples (Simpson indices for those were very low, ~0.17–0.27). Indeed, in G_S3s, a Saronikos surface Gauze sample with Shannon 0.49, Simpson's index was just 0.15, implying over 80% of reads belonged to a single OTU. Such cases point to bloom events or strong dominance, perhaps one eukaryotic plankton (like a particular diatom or dinoflagellate) bloomed and overwhelmed other taxa in that sample. Conversely, the highest eukaryotic Shannon observed was about 4.20 (Filter offshore DS3s), followed by ~4.18 in Sterivex Kavala S_K5d. Shannon values above 4 for microbial eukaryotes are quite high, signifying a very diverse community with no extremely dominant species (Haegeman et al., 2013; Bukin et al., 2019). Only a handful of samples (DS3s, S_K6d, S_K5d, S_K4s) reached this level, mostly from offshore or certain Kavala stations. The majority of samples fell in between these extremes, with Shannon around 2–3, indicating moderate diversity.

Evenness in eukaryotic communities varied widely, reflecting the dominance patterns. In low-Shannon samples, Pielou's evenness was extremely low (often 0.2–0.4), confirming that a few OTUs constituted most of the community. For example, F_T5s (Shannon 0.62) had evenness ~0.14, and G_S2d (Shannon 0.47) had evenness ~0.23, some of the lowest evenness values recorded. On the other hand, a few samples with modest richness had surprisingly high evenness – for instance, Sterivex S_S2s had only 61 OTUs but an evenness of ~0.83, yielding a relatively high Shannon (~3.42) despite low richness. This indicates those 61 OTUs were nearly equally abundant. Overall, many eukaryote samples had evenness in the 0.5–0.7 range and evenness above 0.75

mainly in samples with very few OTUs, where by definition the limited taxa present can be evenly abundant.

Spatial differences in eukaryotic alpha diversity were more pronounced, especially due to the occurrence of localized dominance events. Saronikos consistently exhibited the lowest diversity, irrespective of sampling method. Richness values in Saronikos were frequently below 50 OTUs – for example, Sterivex samples ranged from 19–91 OTUs, and Gauze samples included the lowest richness observed in the entire dataset (8 OTUs in G_S2d). Even when richness increased under the Filter method (e.g., 126–318 OTUs), values remained lower than those recorded in Kavala, Thermaikos, or the Offshore station, which commonly exceeded 300–400 OTUs. In contrast, Thermaikos and Kavala showed substantially higher and more stable richness. Kavala's eukaryote communities were moderate in diversity (average Shannon ~2.5) without too many severe dominance cases, except Gauze issues and Filter samples reached up to 391 OTUs (F_S3s). Kavala Sterivex samples even included one of the highest diversity observations (S_K5d Shannon 4.18). Thermaikos had a mix: some stations in Thermaikos had moderate to high diversity (e.g. S_T6s Shannon ~3.98, G_T1s Shannon 3.46), but others saw different patterns (S_T3s Shannon 1.19, F_T4d Shannon 0.96). Thus, Thermaikos' average Shannon (~2.33) was a bit lower than Kavala's, due to a couple of low-diversity incidents. Finally, the offshore eukaryotic samples were notably diverse. The offshore station samples, all collected by Filter, had high OTU counts (171–247 OTUs for A1, DK3s, DK4s and 484 OTUs for DS3s) and correspondingly high diversity metrics. The offshore surface sample DS3s, in particular, was among the most diverse (Shannon 4.20). The two "DK" offshore surface samples also had Shannons around 3.5–3.7, significantly above the coastal average. This indicates that the open sea environment contained a rich variety of microbial eukaryotes. These results mirror the bacterial pattern, reinforcing that oligotrophic offshore waters can support a high alpha diversity of microbial eukaryotes.

Looking at depth differences for eukaryotes, the pattern was not uniform. Sometimes the surface had higher diversity, and other times the deep did. For example, in Saronikos Sterivex, the deep sample at station S4 (S_S4d) had 103 OTUs and Shannon 3.21, higher than the surface S_S4s (27 OTUs, Shannon 2.18). Similarly, in Kavala Sterivex K6, the deep (272 OTUs, Shannon 4.02) exceeded the surface (239 OTUs, Shannon 3.15). These cases suggest deeper waters can harbor more diverse or at least more even eukaryote communities, potentially due to reduced impact of surface blooms. On the other hand, we saw instances where surface water was more diverse: Thermaikos Sterivex S_T2 (surface Shannon 1.50 vs deep 3.53) here the deep was much more diverse, likely because surface had a dominant organism, whereas deep was diverse. Another example: Saronikos Sterivex S_S2s vs S_S2d both had low richness (~61 vs 46 OTUs) but were very even, giving relatively high Shannon (~3.42 and 3.19), with the surface slightly higher in that case. If we aggregate all data, surface eukaryote samples showed a marginally higher mean Shannon (~2.46) than deep (~2.34), but this average conceals the site-specific alterations. Essentially, when a surface sample coincided with a phytoplankton bloom or macroorganism DNA pulse, its diversity dropped sharply compared to the deep sample, whereas in absence of this condition, surface and deep diversities were comparable or surface slightly higher. Thus, depth-related diversity shifts for eukaryotes seemed to be driven by sporadic

dominance events in the surface layer rather than a consistent richness gradient. The deep waters, being more stable, rarely had extremely low diversity events (e.g., no deep sample had Shannon <0.5, whereas several surface ones did), so in that sense the deep samples in our study presented more even communities at times.

Table 3. Alpha diversity indices of the unicellular eukaryotic communities in each sample. See Table 1 for sampling code annotations.

	Unicellular eukaryotes				
Sampling points	**OTUs richness**	**SChao1**	**SHANNON (H)**	**Simpson (1-D)**	**Pielou's J**
S_K1S	102	147.048	3.175	0.881	0.686
S_K1D	101	144.556	3.256	0.869	0.705
S_K2S	119	206.353	3.571	0.905	0.747
S_K2D	164	221.600	3.278	0.884	0.643
S_K3S	128	199.292	3.752	0.942	0.773
S_K3D	268	363.923	2.936	0.816	0.525
S_K4S	375	425.571	4.171	0.946	0.704
S_K4D	216	330.429	3.544	0.878	0.659
S_K5S	264	310.490	3.804	0.896	0.682
S_K5D	494	596.065	4.183	0.908	0.674
S_K6S	239	296.303	3.148	0.792	0.575
S_K6D	272	361.146	4.024	0.938	0.718
S_T1S	342	404.341	3.350	0.879	0.574
S_T1D	180	216.692	2.292	0.782	0.441
S_T2S	174	201.824	1.503	0.494	0.291
S_T2D	356	424.467	3.530	0.903	0.601
S_T3S	154	253.750	1.187	0.402	0.236
S_T3D	29	42.200	2.042	0.769	0.606
S_T4S	179	230.484	2.127	0.594	0.410
S_T4D	122	182.273	2.059	0.607	0.429
S_T5S	373	426.927	2.306	0.632	0.389
S_T5D	238	328.618	3.150	0.861	0.576
S_T6S	416	515.113	3.983	0.918	0.660
S_T6D	401	502.063	3.811	0.935	0.636
S_S1S	19	34.000	1.778	0.746	0.604
S_S1D	26	45.500	1.134	0.398	0.348
S_S2S	61	136.250	3.422	0.933	0.832
S_S2D	46	73.600	3.191	0.927	0.834
S_S3S	28	41.000	2.032	0.796	0.610
S_S3D	111	180.391	3.046	0.887	0.647
S_S4S	27	36.000	2.178	0.828	0.661
S_S4D	103	138.053	3.214	0.923	0.694
S_S5S	119	164.150	2.875	0.859	0.602
S_S5D	40	71.625	2.199	0.776	0.596
S_S6S	91	103.833	1.794	0.669	0.398

S_S6D	30	36.429	1.613	0.667	0.474
G_K1S					
G_K1D					
G_K2S	42	92.000	2.376	0.827	0.636
G_K2D	29	59.600	1.652	0.707	0.490
G_K3S	15	33.333	1.882	0.745	0.695
G_K3D	30	60.000	2.477	0.852	0.728
G_K4S					
G_K4D					
G_K5S	63	117.167	1.225	0.394	0.296
G_K5D	10	13.333	1.845	0.760	0.801
G_K6S	19	37.333	1.184	0.471	0.402
G_K6D	16	19.000	2.236	0.846	0.806
G_T1S	250	334.774	3.456	0.907	0.626
G_T1D	284	377.366	2.799	0.871	0.495
G_T2S	242	272.625	2.854	0.879	0.520
G_T2D	287	342.275	2.768	0.837	0.489
G_T3S	166	208.000	2.861	0.886	0.560
G_T3D	159	247.750	3.048	0.905	0.601
G_T4S	69	139.000	2.493	0.812	0.589
G_T4D	148	187.808	2.793	0.862	0.559
G_T5S					
G_T5D	109	184.250	2.294	0.819	0.489
G_T6S	54	77.000	2.077	0.749	0.521
G_T6D	104	147.043	2.887	0.907	0.622
G_S1S	19	30.000	2.461	0.878	0.836
G_S1D	65	127.000	3.072	0.924	0.736
G_S2S	37	62.500	2.035	0.723	0.564
G_S2D	8	13.000	0.473	0.178	0.228
G_S3S	74	138.500	0.486	0.149	0.113
G_S3D	113	166.118	2.613	0.793	0.553
G_S4S	54	75.083	2.982	0.891	0.748
G_S4D	130	165.357	2.508	0.767	0.515
G_S5S	148	177.292	3.174	0.833	0.635
G_S5D	17	72.000	0.951	0.395	0.336
G_S6S	109	142.056	1.777	0.702	0.379
G_S6D	145	207.667	2.687	0.817	0.540
G2_K1S	4	7.000	1.242	0.667	0.896
G2_K1D	7	7.333	0.598	0.240	0.307
G2_K4S					
G2_K4D					
F_K1D	183	355.118	1.937	0.647	0.372
F_K1S	163	246.182	2.137	0.704	0.419
F_K2D	206	306.500	1.912	0.570	0.359
F_K2S	208	339.667	2.507	0.742	0.470
F_K3D	201	294.857	2.147	0.688	0.405

F_K3S	184	293.773	2.553	0.772	0.490
F_K4D	144	197.455	2.251	0.806	0.453
F_K4S	141	204.840	2.402	0.811	0.485
F_K5D	200	345.037	1.168	0.556	0.220
F_K5S	184	264.161	1.784	0.577	0.342
F_K6D	223	391.889	2.492	0.803	0.461
F_K6S	166	245.565	2.083	0.713	0.408
F_S1D	107	163.000	0.575	0.173	0.123
F_S1S	126	190.688	2.326	0.758	0.481
F_S2D	194	253.036	2.372	0.792	0.450
F_S2S	232	396.538	3.077	0.826	0.565
F_S3D	270	440.625	2.525	0.718	0.451
F_S3S	194	326.143	2.743	0.799	0.521
F_S4D	217	344.600	2.663	0.840	0.495
F_S4S	207	309.700	3.158	0.901	0.592
F_S5D	175	275.625	2.714	0.861	0.526
F_S5S	198	315.435	2.651	0.843	0.501
F_S6D	115	176.895	0.783	0.278	0.165
F_S6S	151	208.955	2.294	0.800	0.457
F_T1D	159	297.188	1.085	0.331	0.214
F_T1S	185	287.000	1.560	0.631	0.299
F_T2D	142	237.056	1.681	0.630	0.339
F_T2S	157	206.214	1.765	0.629	0.349
F_T3D	217	347.885	2.481	0.798	0.461
F_T3S	128	238.000	1.506	0.691	0.310
F_T4D	162	371.000	0.965	0.322	0.190
F_T4S	184	314.565	1.835	0.647	0.352
F_T5D	140	243.313	1.986	0.660	0.402
F_T5S	97	177.571	0.625	0.271	0.137
F_T6D	233	350.794	2.056	0.595	0.377
F_T6S	212	294.833	2.328	0.703	0.435
A1	171	239.440	2.061	0.693	0.401
DK3S	247	262.600	3.678	0.908	0.668
DK4S	193	203.500	3.534	0.917	0.672
DS3S	484	501.105	4.205	0.938	0.680

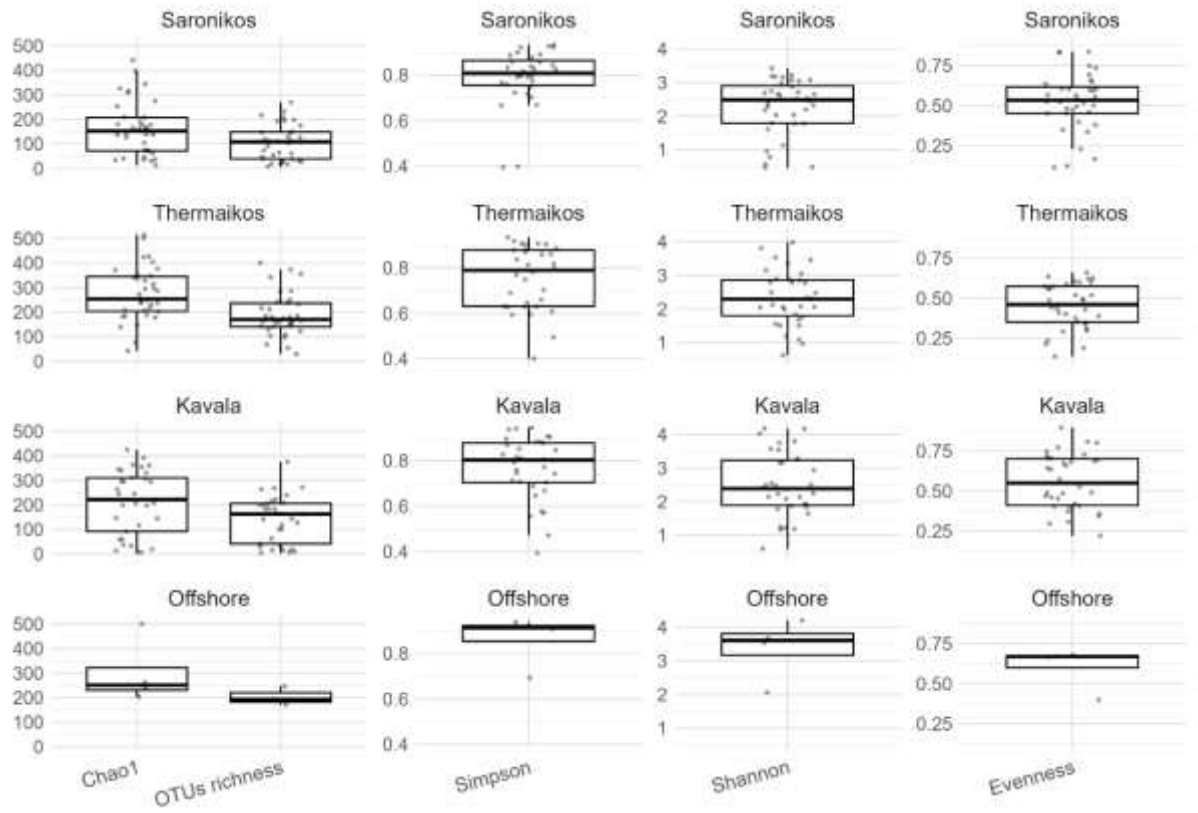

Figure 5. Boxplots of the distribution of the alpha-diversity indices of eukaryotic communities in the examined locations. The horizontal black bars depict the median values, and the dots depict individual samples. Outliers were filtered per-index using 1.5xIQR thresholds.

Examining method effects on eukaryotic a-diversity revealed clear differences among sampling approaches, though none performed uniformly across all conditions. Sterivex samples (S_ codes) generally recovered moderate to high diversity in most coastal stations, with Shannon values often >3 and richness typically in the range of 100–270 OTUs, and they included some of the most diverse coastal samples in the dataset, for example, the surface Kavala sample S_K4s (375 OTUs, Shannon ~4.17) and the deep sample S_K5d (494 OTUs, Shannon ~4.18). These results suggest that Sterivex filters efficiently capture a broad proportion of micro-eukaryotic diversity under typical environmental conditions. However, Sterivex was not uniformly superior: in bloom-dominated or strongly skewed communities, such as S_T3s in Thermaikos (Shannon 1.19), Sterivex diversity was as low as that seen in Gauze or Filter outliers, indicating that the method is equally sensitive to strong dominance patterns. Gauze samples displayed the highest variability and the lowest mean richness, performing particularly poorly in Kavala where several Gauze samples contained only 15–63 OTUs with Shannon 1.2–2.5. Yet, Gauze was not uniformly low-performing: in Thermaikos, for example, several Gauze samples reached moderate richness (150–287 OTUs) and Shannon around 2.5–3.0, and some Saronikos samples (e.g., G_S1d, G_S5s) also showed relatively high evenness (~3.0). Standard Filters produced site-dependent outcomes: in coastal areas with high biomass, Filter samples tended to show low

Shannon values (<2) despite moderate richness, implying dominance by a few taxa, potentially due to clogging or selective retention of larger phytoplankton. In contrast, in offshore oligotrophic waters, the Filter method produced some of the highest eukaryotic diversities observed, including DS3s (484 OTUs, Shannon 4.20) and DK3s/DK4s (Shannon ~3.5–3.7). Thus, while Sterivex was generally the most stable across coastal sites, Filters performed were also efficient especially in offshore areas, and Gauze was consistently the least reliable. Overall, the dataset shows that no single method captures eukaryotic diversity equally across all environments, and differences between methods partly reflect how each interacts with water clarity, biomass load, and community dominance structure.

In summary, eukaryotic alpha diversity in our samples was highly variable and often dictated by whether a single taxon dominated the community. Spatially, Saronikos were less diverse and more unevenly distributed, in contrast to the richer and more balanced communities observed in the other regions. Kavala and offshore stations displayed the highest richness and diversity estimates, while Thermaikos showed intermediate but generally elevated values. The choice of sampling method had substantial impact on the detected diversity. Sterivex filters provided consistently higher and more even diversity measurements for eukaryotes, followed by equally well-performed Filter whereas Gauze frequently under-sampled or lost eukaryotic diversity. These findings suggest that methodological biases must be accounted for, especially for eukaryotic surveys, to avoid underestimating community diversity academic.oup.com. Despite these challenges, our data clearly show that the Aegean planktonic eukaryote communities can be quite diverse (Shannon 3–4), though still generally less diverse than the co-occurring bacterial communities. Each region's alpha diversity profile thus provides insight into its ecological state: high evenness and richness offshore hint at a complex stable community, while the low diversity events in coastal gulfs signal episodic dominance by opportunistic eukaryotic species (e.g. phytoplankton blooms) that significantly reduce community evenness. This dynamic range of alpha diversity underscores the importance of broad spatial sampling and method cross-validation in accurately characterizing planktonic ecosystem diversity.

3.2.2. Relative abundance trends

Across all sampling sites, the full eukaryotic dataset shows that multicellular organisms substantially influence community profiles, with frequent dominance of metazoans such as cnidarians, urochordates and copepods. Dinophyceae, however, still represent the largest single group (~44% of all reads), followed by diatom classes such as Mediophyceae and Bacillariophyceae, fungi (primarily Ascomycota), and several green algal and parasitic lineages (Figure 6a.). These multicellular taxa can mask the structure of the underlying microbial eukaryotic community. When restricting the analysis to unicellular taxa (Figure 6b.), a clearer picture emerges for phytoplankton communities. Dinophyceae rise to roughly 57% of all unicellular reads, becoming the dominant phytoplankton group overall, followed by diatoms, haptophytes and various small green algae, along with widespread Syndiniales parasites.

Focusing on specific locations, In Saronikos Gulf, unicellular assemblages are strongly dominated by Dinophyceae across a wide range of samples, methods and depths ~0.59% on average). Some samples reach near-total dominance, such as F_S1d at 96.1% F_S6d reaching 98.7%, as well as S_S1 both in surface and deep station, where Dinophyceae approach 98% relative abundance, indicating the near-exclusion of other phototrophic groups.

Within Saronikos, however, there is noteworthy heterogeneity between sampling points. For instance, at station S4 in the surface layer (F_S4s), there is a marked co-dominance of small green algae, with Mamiellophyceae contributing approximately 25% relative abundance, accompanied by chlorophytes and haptophytes. At the same station in the deep layer (F_S4d), the community shifts back toward dinoflagellate dominance (around 48%). Saronikos also provides a clear internal comparison of methodological effects at a single station. At S5 surface, all methods detect a dinoflagellate-dominated community, but with differing emphasis on additional groups. In F_S5s, Dinophyceae represent about 63% of the community, whereas in S_S5s they reach approximately 69% and are accompanied by stronger signals from picoeukaryotic groups such as Picozoa_XX. Meanwhile, G_S5s presents a somewhat more particle-associated assemblage, with Prymnesiophyceae and other nanoplanktonic taxa contributing more prominently. Although these differences highlight method-specific biases, the underlying community structure remains consistent: Saronikos, especially at the surface, is a dinoflagellate-rich system with variable contributions from small green algae, haptophytes, and heterotrophs depending on the station.

Kavala Gulf presents patterns that resemble Saronikos in terms of strong dinoflagellate dominance but also includes distinctive components not as pronounced in the other gulfs. Dinophyceae frequently exceed 75% of the relative abundance in numerous samples, with examples such as F_K1s, F_K4s and F_K6s highlighting this trend. One of the most extreme cases is G2_K4s, where Dinophyceae reach a relative abundance of 100%, indicating complete dominance in that sample, may be driven by a problematic sampling attempt. The consistency of high Dinophyceae values across Filter, Sterivex and Gauze samples suggests a stable system dominated by dinoflagellate representatives, similar in this respect to Saronikos.

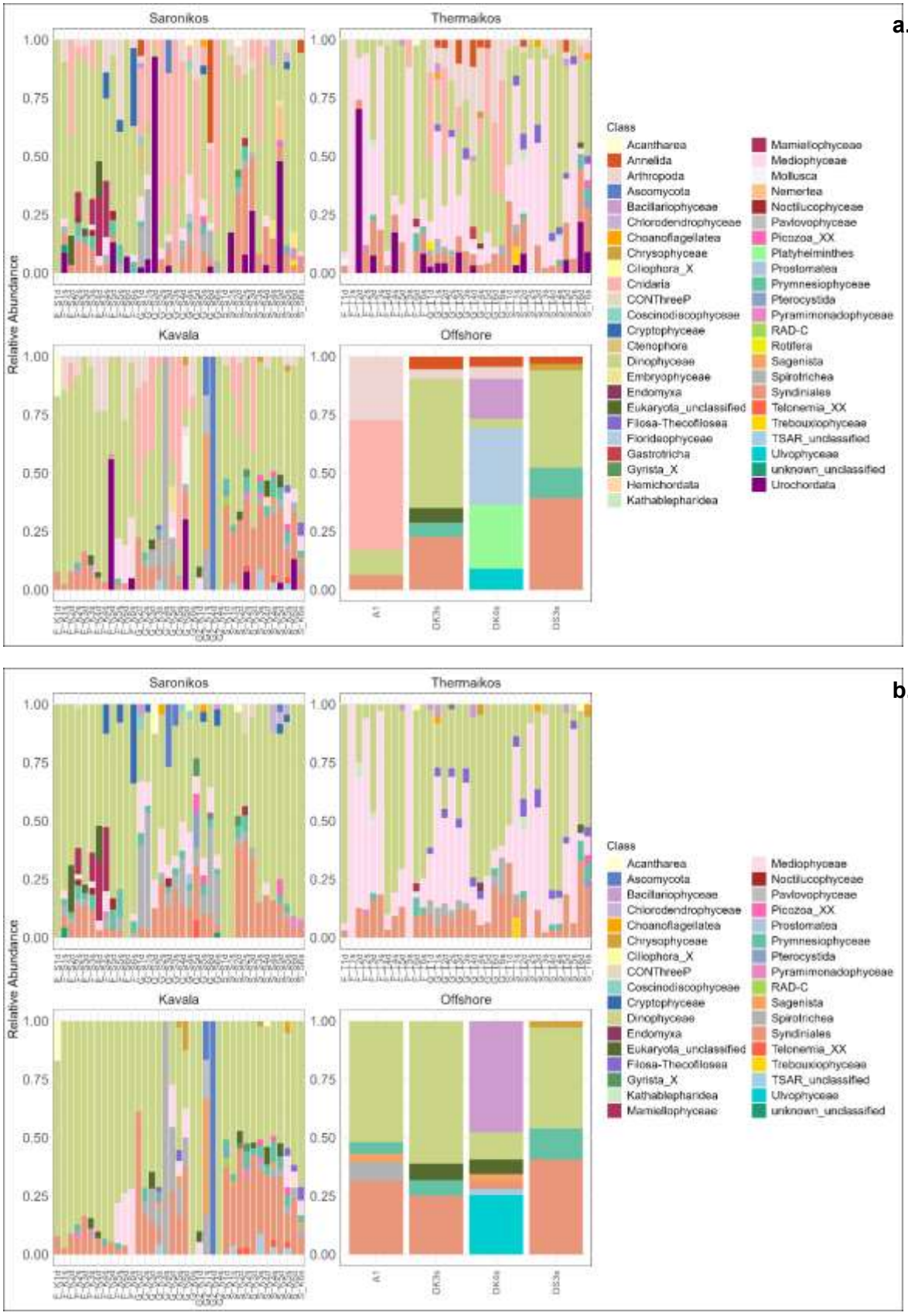

Figure 6. Bar plots showing the relative abundance of higher taxonomic eukaryotic groups per Location (a). The taxonomic groups contributing >2 % of the total number of reads in the sample are plotted (b).

What distinguishes Kavala is the representation of diatoms such as Mediophyceae (~8% on average) and parasitic Syndiniales (18% on average), which contribute less than dinoflagellates and fungi but still form distinct peaks in specific samples, indicating that Kavala hosts a mixed assemblage in which several functional types (photosynthetic, parasitic, heterotrophic) can locally dominate. Deep samples such as S_K3d, S_K4d and S_K5d show Syndiniales at relative abundances above 28–39%, while Gauze samples such as G_K2d reach approximately 56% for the same group. This indicates that Kavala, like Saronikos, hosts a dynamic system of dinoflagellate–syndiniale interactions. Kavala also contains a richer contribution from fungal and radiolarian groups. In the extreme case of Ascomycota reaching full dominance 100% in G2_K4d, this might be due to sampling failing instead of actual representation of this Class in the community. Radiolarian taxa such as Acantharea appear prominently in F_K1d (~16%), pointing to local environmental or hydrographic conditions promoting the retention or deposition of particle-associated taxa.

In Thermaikos, the unicellular community is characterised by a more classical diatom–dinoflagellate co-dominance. Dinophyceae have a mean relative abundance of ~44%, while Mediophyceae reach ~34% on average. Many Thermaikos samples are dominated by diatoms at the surface, especially in Filter and Sterivex collections. For example, surface Filter samples F_T5s and F_T1s reach ~98% and ~95% Mediophyceae, respectively, while surface Sterivex sample S_T3s reaches ~88% Mediophyceae. At the same time, some deep Filter and Sterivex samples show strong dinoflagellate signals: deep Filter F_T1d has 92% Dinophyceae, and deep Filter F_T2d has ~51% Mediophyceae together with 16% Dinophyceae. Across the dataset, parasitic Syndiniales contribute a mean of ~12%, with values such as ~18% in surface Filter F_T3s and ~15% in deep Filter F_T3d, indicating a substantial parasitic component associated with the diatom–dinoflagellate assemblage. Small green algae like Trebouxiophyceae and Kathablepharidea are generally present at lower abundances but occasionally peak in individual samples (e.g. S_T1s ~7% Trebouxiophyceae; F_T2d ~6% Kathablepharidea), further diversifying the unicellular community. Overall, Thermaikos stands out as the gulf where diatoms and dinoflagellates jointly structure the unicellular eukaryotic assemblage, with parasitic and minor phytoplankton groups layered on top.

The offshore station shows a somewhat different picture, with strong diatom dominance and a large contribution from macroalgal or larger green-algal lineages. Dinophyceae is still the dominant Class across most samples (mean ~36%) apart from sample DK4, where Bacillariophyceae reach a mean abundance of ~40%, slightly exceeding Dinophyceae. In sample DK4s, Ulvophyceae (green algae) are also very prominent (~21%). Syndiniales also contribute strongly offshore (~22% on average). Other notable components include Sagenista (~2%) and Chrysophyceae (~2%). Thus, offshore waters are dominated by diatoms and dinoflagellates like the gulfs, but with a relatively higher proportion of benthic or macroalgal lineages (e.g. Ulvophyceae) and unclassified eukaryotes, likely reflecting different environmental conditions and source inputs compared to the semi-enclosed gulfs.

Methodological effects are visible but clearly secondary to location-driven patterns. Within each gulf, the three sampling approaches, broadly recover the same dominant unicellular classes, but they differ somewhat in relative emphasis. In Kavala, Filter samples tend to emphasise Dinophyceae and Mediophyceae, whereas Gauze samples show very strong Ascomycota (~76% in G2_K4d) and high Spirotrichea, highlighting epipelagic ciliates and fungi more strongly. Sterivex in Kavala sits in between, with Dinophyceae and Syndiniales as major components. In Saronikos, all three methods recover high Dinophyceae, but Filter samples particularly capture peaks of small phytoplankton such as Mamiellophyceae (e.g. F_S4s and F_S4d) and Cryptophyceae (F_S6s, F_S4s), whereas Gauze samples tend to highlight Spirotrichea and Pterocystida. In Thermaikos, Filter and Sterivex both recover the diatom–dinoflagellate co-dominance, but Filter shows especially high Mediophyceae at the surface (e.g. F_T5s, F_T1s), while Gauze samples more clearly reveal the Syndiniales and dinoflagellate components in some deep and surface layers.

Comparing gulfs, all locations share a strong unicellular dinoflagellate signature but diverge in their secondary components and in the balance between dinoflagellates and diatoms. Kavala is characterised by a very mixed unicellular community where Dinophyceae is highly abundant on average, and where individual stations were dominated by Spirotrichea, Ascomycota, or dinoflagellates. Saronikos is also dinoflagellate-dominated but stands out for its rich assemblage of pico- and nanophytoplankton (Mamiellophyceae, Prymnesiophyceae, Cryptophyceae) and heterotrophic protists (Pterocystida, Spirotrichea), which are consistently present at relatively high average abundances and reach strong peaks in specific samples (e.g. F_S4s, F_S6s, G_S5d). Thermaikos, by contrast, is the gulf where diatoms (Mediophyceae) are co-dominant with Dinophyceae across many samples, with parasite-rich Syndiniales communities overlaying this phytoplankton backbone. The offshore station resembles the gulfs in that dinoflagellates are abundant but also include high contributions from Ulvophyceae and Chrysophyceae. Taken together, the unicellular eukaryotic communities across these locations suggest a common regional background of dinoflagellate dominance, modulated locally by varying contributions of diatoms, small green algae, haptophytes, ciliates, parasites, and fungi, with location.

In general, the eukaryotic community patterns observed across the Aegean align with broad trends reported for coastal and open marine systems. Dinoflagellates, diatoms, small green algae, haptophytes, and Syndiniales typically make up the dominant fractions of marine microbial eukaryotes, and these groups were also consistently represented in our dataset. The strong dinoflagellate dominance in several coastal stations, and the frequent co-occurrence of diatoms and Syndiniales, fall within commonly described Mediterranean plankton dynamics, although the exact proportions varied by gulf. At the same time, each region displayed distinct secondary components and variability among sampling methods, reflecting local environmental conditions and trophic states. Overall, our results generally agree with established patterns of marine eukaryotic community composition, while highlighting site-specific

differences across the four regions (Kalu et al., 2023; Mordret et al., 2023; Miao et al., 2024).

Across all locations, the eukaryotic community is structured around a small group of consistently dominant Classes. Dinophyceae and Syndiniales represent the most ubiquitous and abundant groups in the dataset, appearing in nearly every sample and forming the core of the assemblage (Figure 7). Mediophyceae are also widely present across regions, but their distribution varies geographically: they are prominent and consistently abundant in Thermaikos and Offshore samples, whereas in Kavala they are absent from a substantial number of stations, including F_K1d, F_K2d, F_K3d, G_K2d, G_K3s, G_K4s, G_K5d, and G_K4d. This uneven representation suggests that diatom dominance, although general across the system, is strongly location-dependent, being weakest in Kavala or underrepresented through the sampling methods.

Other protistan groups display regionally distinctive patterns. Choanoflagellatea are especially consistent in Thermaikos, where they occur in nearly all stations except G_T6s. Their presence is more patchy in Saronikos, missing from several sites, and even more irregular in Kavala, where they are largely absent from gauze-collected samples and from two deep filter samples (F_K1d and F_K4d). In contrast, sterivex sampling captures Choanoflagellatea in all Kavala stations. Offshore waters host Choanoflagellatea in moderate abundance across all stations except DK4s.

Telonemia shows a similarly structured but location-sensitive pattern. In Saronikos, Telonemia is missing from several stations (notably gauze G_S1 and G_S2, F_S1d, S_S1d, S_S1s), whereas in Thermaikos it is consistently present with stable relative abundance, absent only from S_T3d. The offshore sites show high Telonemia contribution except for DK4s, mirroring the pattern of Choanoflagellatea. In Kavala, Telonemia is reliably detected in all sterivex and filter samples, but frequently absent from gauze samples, indicating a method-specific capture pattern.

Gyrista are markedly more persistent and abundant in Thermaikos than in Saronikos, both in the number of positive stations and their mean relative abundance. In Saronikos, Gyrista appear at high levels in specific stations such as G_S5d and S_S2d but are less uniformly distributed. Offshore samples show a Gyrista distribution similar to Telonemia, with consistent representation except at DK4s. In Kavala, Gyrista are captured in all sterivex and filter samples but rarely detected in gauze samples, again illustrating method-linked differences in recovery.

Some classes exhibit highly specific spatial signatures. Noctilucophyceae contribute strongly to many filter-collected samples in Saronikos, appearing in nearly all filter stations except S_S1 and F_S2d, and also in G_S3d, S_S3s, S_S4d, and S_S5s. Their presence in Thermaikos is restricted to only three stations (F_T1d, F_T5d, S_T5d) and they are entirely absent from Kavala and Offshore.

Cryptophyceae also show location-dependent variation. They are more consistently detected and often more abundant in Thermaikos, missing from only six samples,

compared to Kavala, where they are missing from nine. Despite this, Saronikos exhibits slightly higher average cryptophyte abundance across the samples in which they occur. Offshore waters contain Cryptophyceae at moderate levels across all stations except DK3s. In Kavala, Cryptophyceae are generally absent from gauze samples and mainly detected in filter and sterivex collections.

Sagenista and Bacillariophyceae appear across all regions but are especially abundant and consistent in Thermaikos, where they contribute significantly in most samples. In Saronikos, however, both groups are absent from several stations, yielding a more heterogeneous distribution. Their near-complete absence from Kavala gauze samples again underscores the sampling-method effect already observed for several other protist groups.

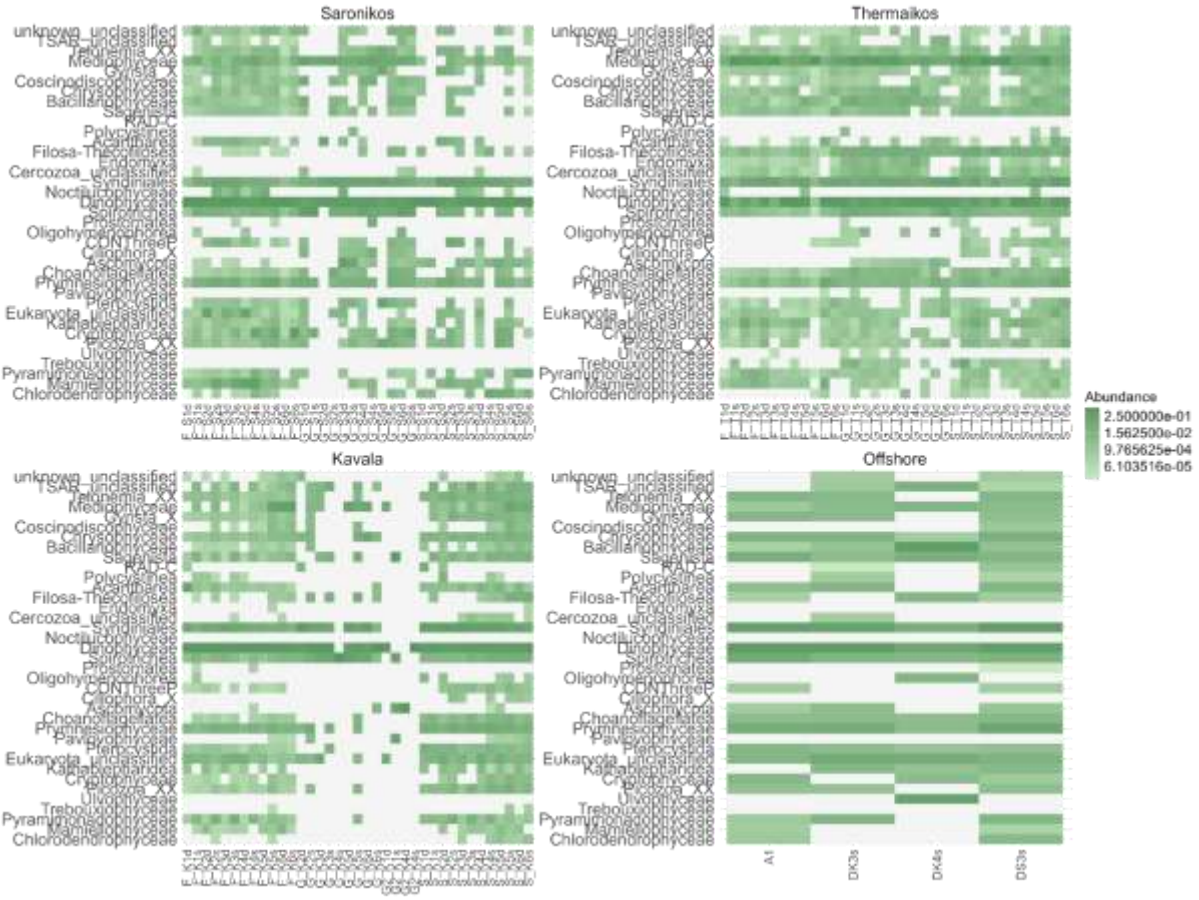

Figure 7. Heatmaps of eukaryotic Class-level, log4 transformed relative abundance, showing only Classes >2% in at least one sample.

Overall, these patterns highlight that all four regions share a core community dominated by Dinophyceae, Syndiniales, and Mediophyceae, but exhibit distinct secondary signatures. Thermaikos stands out for its high consistency across stations and greater representation of Choanoflagellatea, Telonemia, Gyrista, while Saronikos shows stronger contributions from classes like Noctilucophyceae and displays notable method-linked variability. Kavala is more method-sensitive, with gauze samples consistently lacking several protist groups, whereas Offshore stations show

streamlined, open-water patterns dominated by the major phytoplankton classes. The overall result is a structured but location-specific eukaryotic landscape with strong phytoplanktonic consistency and clear regional differentiation.

3.2.3. Beta diversity

The beta diversity patterns of eukaryotic communities reveal markedly different structuring compared to bacteria, both in the integrated dataset and when focusing exclusively on unicellular taxa. In the full dataset, which includes multicellular groups such as Cnidaria, Arthropoda, and Urochordata, the NMDS ordination does not show a strong or clear geographic separation among gulfs (Figure 8 a.). Instead, most samples converge toward a single broad central cluster. This reduced spatial structuring contrasts with the bacterial NMDS and reflects the inherently more heterogeneous and patchy distribution of eukaryotes across marine environments—especially when large metazoans, gelatinous taxa, and macrozooplankton are included.

Within the integrated ordination, Gauze samples stand out as the main outliers. In both Saronikos and Kavala, many Gauze samples diverge substantially from the main cluster, dispersing broadly across the biplot. This pattern is consistent with the method's tendency to retain larger particles and macro-organisms, which dominate many Gauze profiles in the full dataset. In Thermaikos, Gauze samples also deviate, but the effect is less pronounced than in the other gulfs. Interestingly, in this gulf it is some Filter and Sterivex samples that show larger departures from the central cluster, likely reflecting the strong diatom and dinoflagellate blooms observed in the relative abundance data (e.g., extreme dominance of Mediophyceae in F_T1s and F_T5s).

The Offshore sites also show mixed behaviour: two samples cluster within the central group, whereas two others are placed remotely on the ordination. The distant samples correspond to cases where dinoflagellates or unclassified eukaryotes strongly dominated the composition, and where alpha diversity values were unusually low—conditions that typically pull points away due to their highly uneven community structure.

After the exclusion of multicellular classes, the NMDS stress decreases, reflecting an improved fit to the reduced dataset (Figure 8b.). However, the overall ordination pattern remains broadly similar: most samples still converge into a large central cluster, with a second, slightly looser group formed primarily by Thermaikos samples, consistent with the previously identified distinct phytoplankton structure in this gulf. What changes most is the behaviour of the previously scattered Gauze samples. Once multicellular OTUs are removed, the Gauze outliers largely disappear, and these samples become "absorbed" into the main cloud. This shift strongly supports the interpretation that the initial Gauze dispersion was driven by their overrepresentation of metazoan taxa.

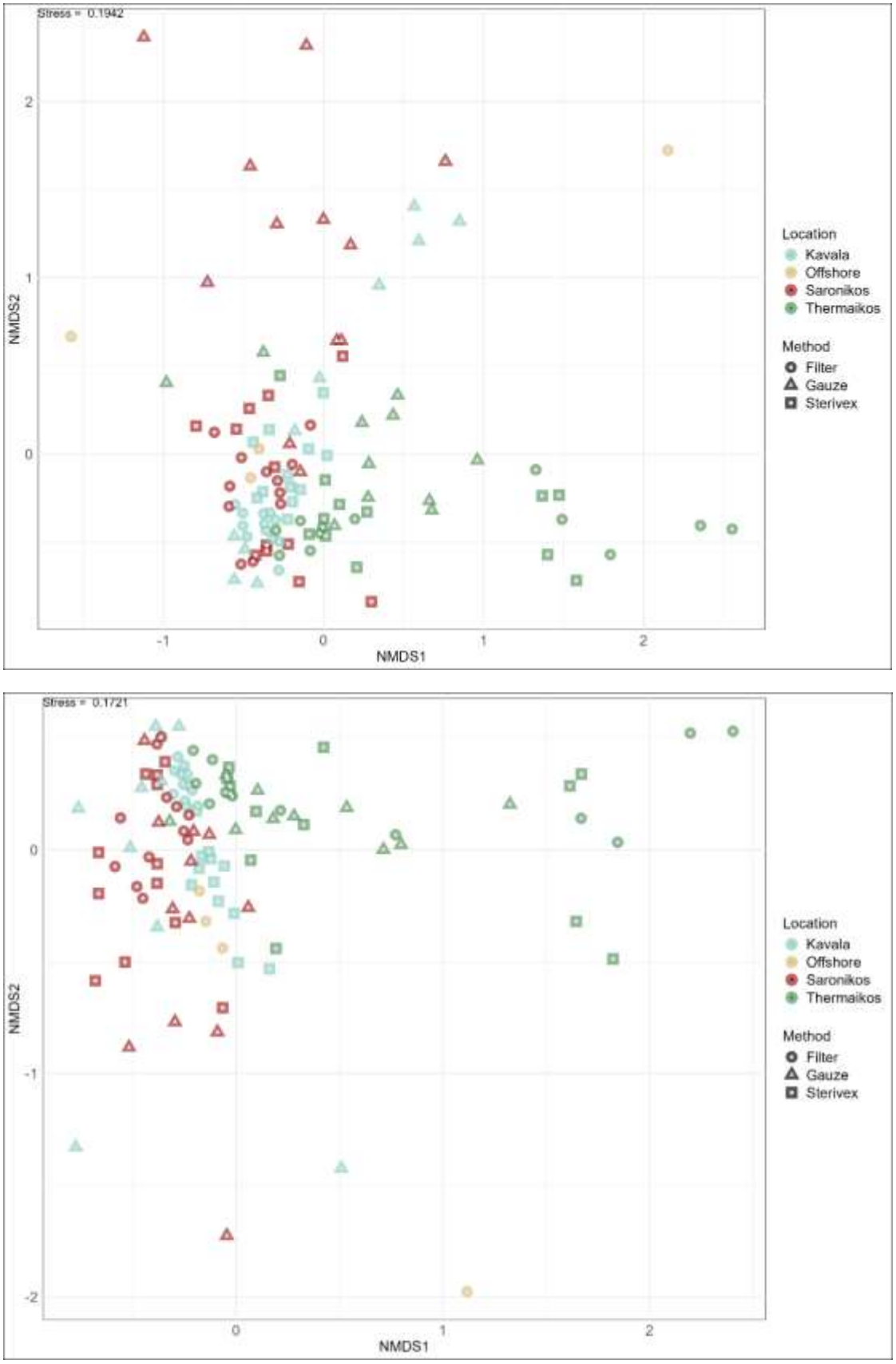

Figure 8. Non-metric Multidimensional Scaling (nMDS) ordination plot of eukaryotic OTUs relative abundances showing the relatedness of sampling points based on Bray-Curtis dissimilarities. Different colors indicate the Locations from which the samples

were collected and different shapes the different sampling method (a). Eukaryotic NMDS after removing multicellular Classes (b).

The two Offshore samples that appeared as clear outliers in the integrated dataset also collapse toward the main group after filtering. This behaviour is consistent with their full profiles, where multicellular taxa contributed substantially to their divergence. Once these groups are removed, their unicellular composition resembles that of other marine pelagic samples across gulfs.

In contrast to bacteria, eukaryotic beta diversity shows weaker geographic structuring and a dominant "core cluster" reflecting the overall similarity in unicellular phytoplankton among gulfs. The strongest source of variation in the integrated dataset is methodological, specifically, the disproportionate capture of metazoans by Gauze filters. After excluding multicellular taxa, this methodological signal collapses, while a subtler ecological gradient appears, mainly distinguishing Thermaikos due to its distinct diatom–dinoflagellate balance. Overall, the NMDS patterns align closely with the relative abundance and alpha diversity features characteristic of each gulf.

3.3. Fish

A total of 12,297,457 reads were grouped into MOTUS, which were assigned to 22 fish species (Figure 9). The anchovy, *Engraulis encrasicolus*, was the most frequently detected species and was present at all stations, except Station 2 in Kavala. The presence of all species is known in Greek seas, except for the Atlantic salmon *Salmo salar*, which is not found in Mediterranean Sea. Additionally, the bluefin tuna *Thunnus thynnus* was detected at Station 4 in the Saronic Gulf. Eleven species were found in all areas, whereas six and five were exclusively found in Saronikos and Thermaikos Gulfs, respectively (Figure 10).

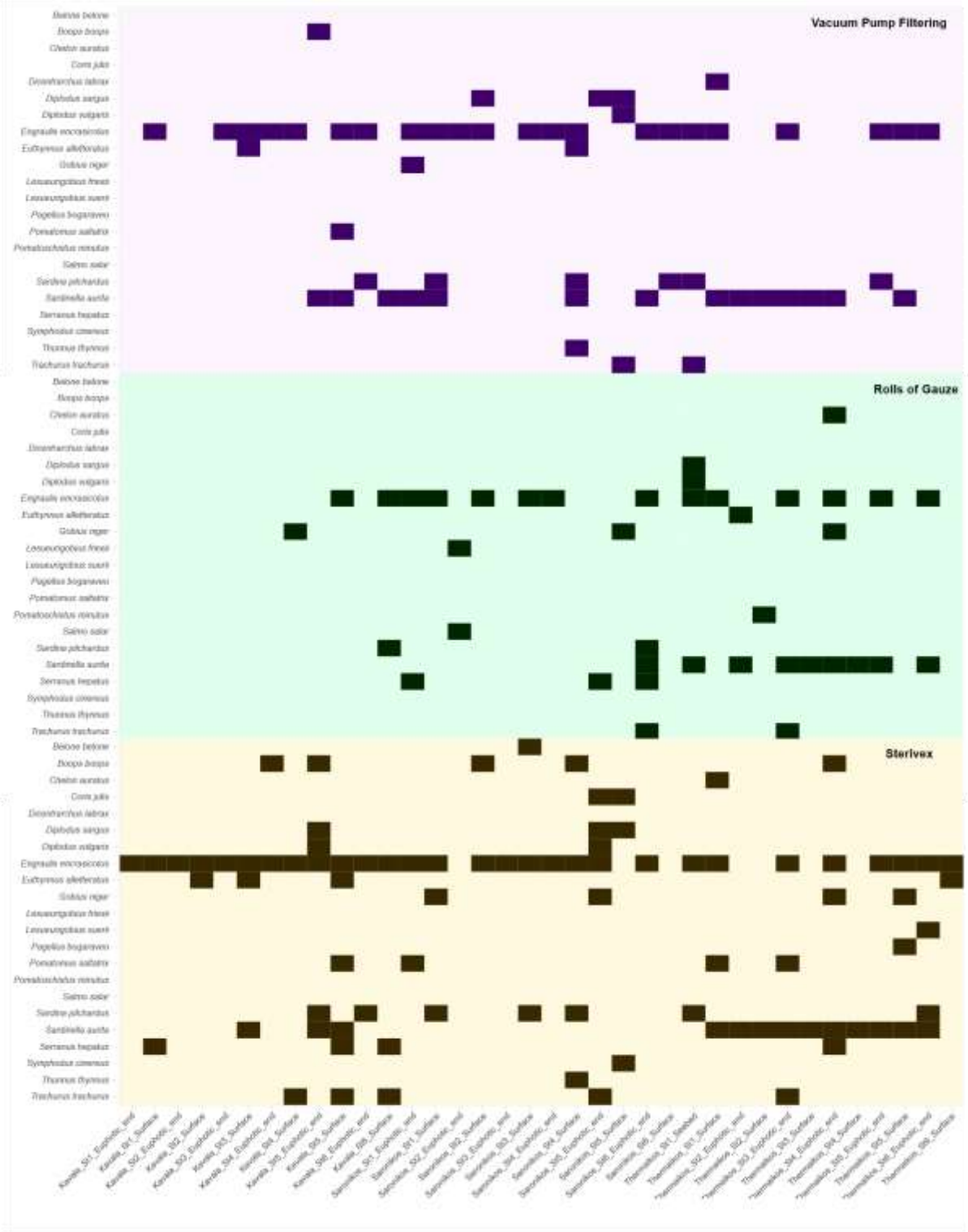

Figure 9. Species detected using three sampling techniques [a) vacuum pump filtering, b) passive sampling in rolls of gauze, and c) Sterivex], sampling station (Kavala, Saronikos, and Thermaikos Gulfs) and depth (sea surface and end of euphotic zone).

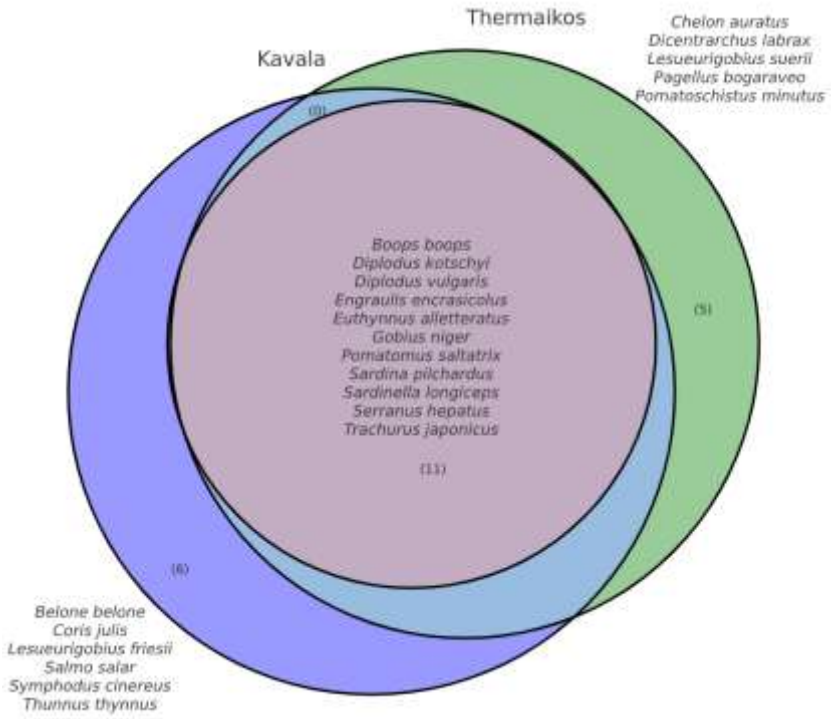

Figure 10. Species detected across all sampling areas.

3.3.1. Alpha diversity

In total, 17 species were found across all stations and with all methods. Thermaikos Gulf had the highest number of species (17), followed by Saronikos Gulf (16), and then Kavala (11) (Figure 9). The fish communities exhibited a wide range of alpha diversity (Table 4, Figures 11-12) across different locations, sampling methods and depths. The species detected by eDNA metabarcoding ranged from 0 to 7. The Shannon and Simpson indices from all stations ranged from 0.000 to 1.697 and 0.000 to 1.000, respectively (Table 4, Figure 11). Overall, the species Richness and the Shannon index were lower in Kavala, whereas a wider range in the Simpson index was shown in the same location (Figure 11). Overall, Thermaikos Gulf exhibited the highest fish diversity, with consistently elevated richness, Shannon, and Simpson indices across all sampling methods and depths. On average, samples from the sea surface showed slightly lower Shannon (mean 4.17 vs 3.56) and Simpson (mean 0.42 vs 0.40) values than the end of euphotic zone.

Table 4. Alpha diversity indices of fish communities at each sampling point. See Table 1 for sampling code annotations.

Sampling Points	Fish		
	Species Richness	Shannon	Simpson
S_K1D	1	0.000	0.000
S_K1S	2	0.000	0.000

S_K2D	1	0.000	0.000
S_K2S	2	0.317	0.174
S_K3D	1	0.000	0.000
S_K3S	3	0.665	0.386
S_K4D	2	0.108	0.044
S_K4S	2	0.140	0.061
S_K5D	6	1.231	0.577
S_K5S	7	0.931	0.436
S_K6D	2	0.282	0.149
S_K6S	3	0.396	0.234
S_S1D	2	0.004	0.001
S_S1S	3	0.852	0.489
S_S2D	0	0.000	1.000
S_S2S	3	0.004	0.001
S_S3D	1	0.000	0.000
S_S3S	3	0.668	0.473
S_S4D	1	0.000	0.000
S_S4S	4	1.284	0.702
S_S5D	6	1.697	0.801
S_S5S	3	1.098	0.667
S_S6D	1	0.000	0.000
S_S6S	0	0.000	1.000
S_T1D	2	0.566	0.378
S_T1S	5	0.701	0.366
S_T2D	1	0.000	0.000
S_T2S	1	0.000	0.000
S_T3D	4	1.294	0.702
S_T3S	1	0.000	0.000
S_T4D	6	1.486	0.732
S_T4S	1	0.000	0.000
S_T5D	3	0.822	0.485
S_T5S	4	1.160	0.646
S_T6D	4	0.784	0.387
S_T6S	3	1.074	0.649
G_K1D	0	0.000	1.000
G_K1S	0	0.000	1.000
G_K2D	0	0.000	1.000
G_K2S	0	0.000	1.000
G_K3D	0	0.000	1.000
G_K3S	0	0.000	1.000
G_K4D	0	0.000	1.000
G_K4S	1	0.000	0.000
G_K5D	0	0.000	1.000
G_K5S	1	0.000	0.000

G_K6D	2	0.692	0.499
G_K6S	3	1.079	0.653
G_S1S	1	0.000	0.000
G_S1D	2	0.544	0.358
G_S2D	2	0.683	0.490
G_S2S	1	0.000	0.000
G_S3D	0	0.000	1.000
G_S3S	1	0.000	0.000
G_S4D	1	0.000	0.000
G_S4S	0	0.000	1.000
G_S5D	1	0.000	0.000
G_S5S	1	0.000	0.000
G_S6D	5	1.377	0.700
G_S6S	0	0.000	1.000
G_T1D	5	0.868	0.539
G_T1S	2	0.669	0.476
G_T2D	3	1.020	0.617
G_T2S	1	0.000	0.000
G_T3D	3	0.982	0.584
G_T3S	2	0.384	0.224
G_T4D	6	1.461	0.688
G_T4S	1	0.000	0.000
G_T5D	2	0.641	0.449
G_T5S	0	0.000	1.000
G_T6D	3	0.399	0.234
G_T6S	0	0.000	1.000
F_K1D	0	0.000	1.000
F_K1S	1	0.000	0.000
F_K2D	0	0.000	1.000
F_K2S	0	0.000	1.000
F_K3D	1	0.000	0.000
F_K3S	2	0.493	0.314
F_K4D	1	0.000	0.000
F_K4S	1	0.000	0.000
F_K5D	3	0.903	0.527
F_K5S	3	0.995	0.593
F_K6D	2	0.361	0.207
F_K6S	1	0.000	0.000
F_S1D	3	1.041	0.629
F_S1S	3	0.814	0.473
F_S2D	1	0.000	0.000
F_S2S	2	0.566	0.378
F_S3D	0	0.000	1.000
F_S3S	1	0.000	0.000

F_S4D	1	0.000	0.000
F_S4S	5	1.522	0.771
F_S5D	1	0.000	0.000
F_S5S	3	1.066	0.646
F_S6D	2	0.664	0.471
F_S6S	2	0.693	0.500
F_T1D	3	1.038	0.629
F_T1S	3	1.005	0.612
F_T2D	1	0.000	0.000
F_T2S	1	0.000	0.000
F_T3D	2	0.429	0.260
F_T3S	1	0.000	0.000
F_T4D	1	0.000	0.000
F_T4S	0	0.000	1.000
F_T5D	2	0.628	0.436
F_T5S	2	0.650	0.457
F_T6D	2	0.509	0.328
F_T6S	1	0.000	0.000

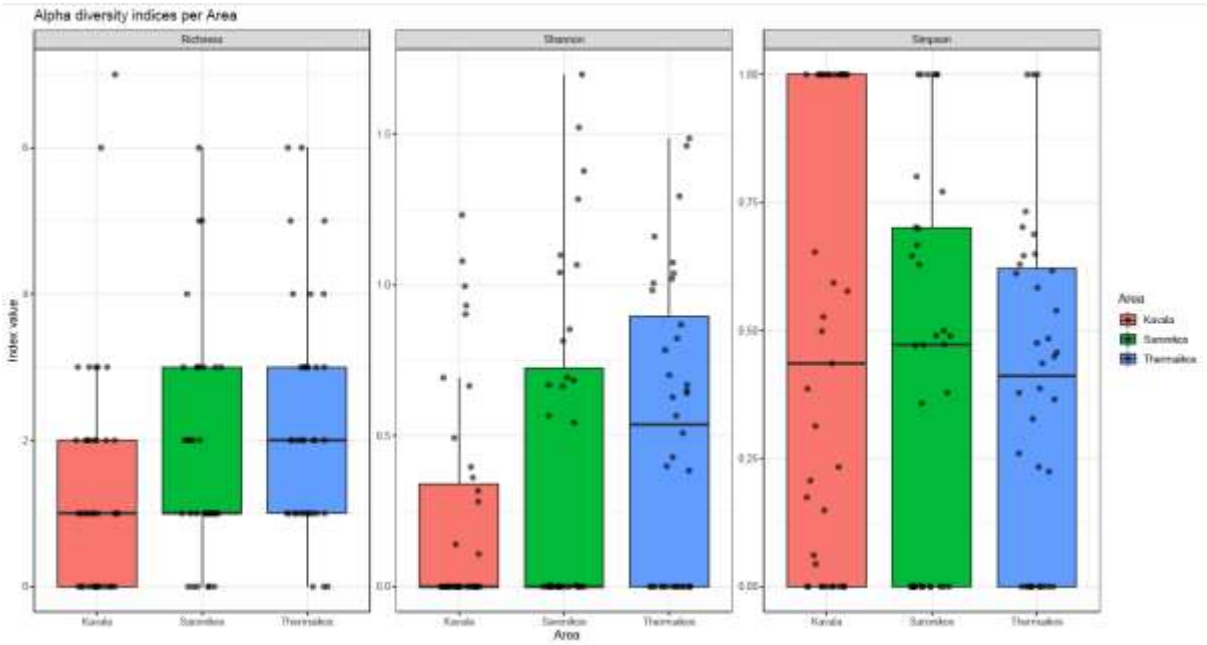

Figure 11. Alpha-diversity indices of fish communities across sampling locations.

Differences in alpha diversity of fish communities were detected amongst the three methods used. The Sterivex, Gauze, and Filter methods yielded different diversity outcomes. The Sterivex samples generally showed the highest values amongst the methods, followed by the Filter, whereas the Gauze yielded the lowest values (Table

4, Figure 12). The passive collection of environmental samples with gauzes produced the lowest species-level records compared to active filtration methods (Figure 12).

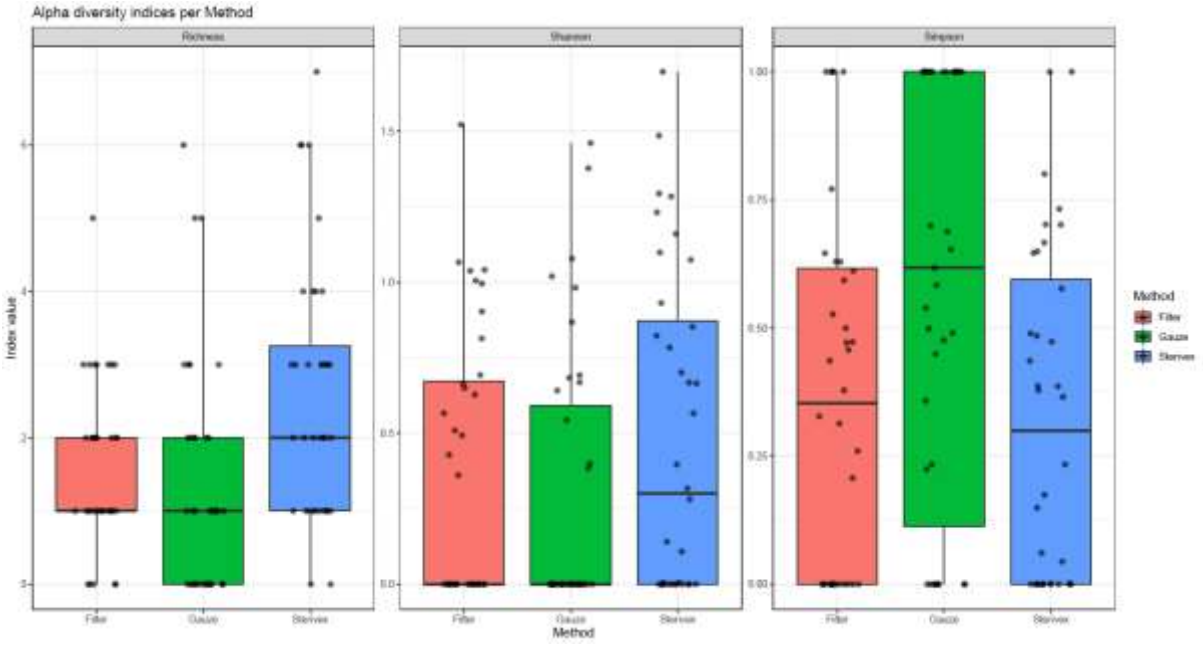

Figure 12. Alpha-diversity indices of fish communities by sampling method.

The NMDS ordination of fish communities revealed distinction among the three sampling areas (Figure 13), with Kavala showing comparable results to Thermaikos Gulf, whereas Saronikos was a bit more distinct from the other two Gulfs.

PERMANOVA analysis confirmed the homogeneity between the stations at each region, and that the fish communities detected in the samples from the surface and the end of the euphotic zone did not differ significantly ($p > 0.05$). Furthermore, fish communities differ both geographically amongst the three gulfs ($R2=0.093$, $p < 0.001$), and the sampling methods ($R2=0.068$, $p < 0.001$).

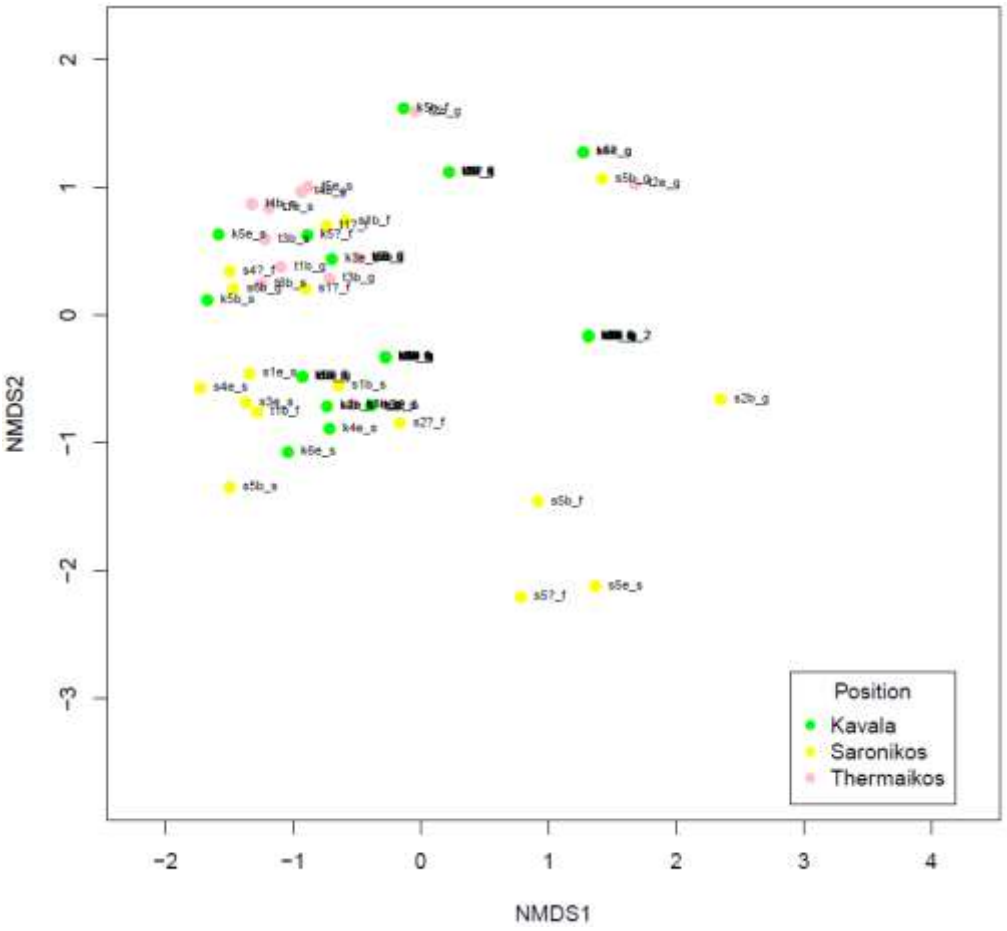

Figure 13. Non-metric multidimensional scaling (NMDS) plot of fish communities detected with eDNA metabarcoding across the three sampling areas.

3.4. Functional profiles

Shotgun metagenomic sequencing was performed on a selected subset of samples across the surveyed regions to characterize the functional potential and to evaluate whether functional signatures exhibited spatial structuring analogous to the taxonomic patterns observed in the amplicon-based analyses. The functional classification of recovered Clusters of Orthologous Genes (COGs) yielded pathway-level, subsystem-level, and module-level profiles. Collectively, these results provide a large-scale functional overview of the three examined gulfs using an eDNA framework.

A striking result emerging from the dataset is the high degree of functional similarity of the global pathways across stations despite marked taxonomic differences (Figure 14). Whereas the taxonomic composition of the detected assemblages varied significantly among sampling locations the functional profiles were comparatively stable. This pattern suggests substantial functional redundancy across the sampled gradients. In other words, while the identity of dominant taxa may shift between regions, key metabolic capabilities remain consistently represented. This

phenomenon has been observed in other coastal systems and is generally interpreted as an indication that metabolic functions are buffered against environmental variability through overlapping functional roles contributed by phylogenetically distinct taxa (Louca et al., 2016).

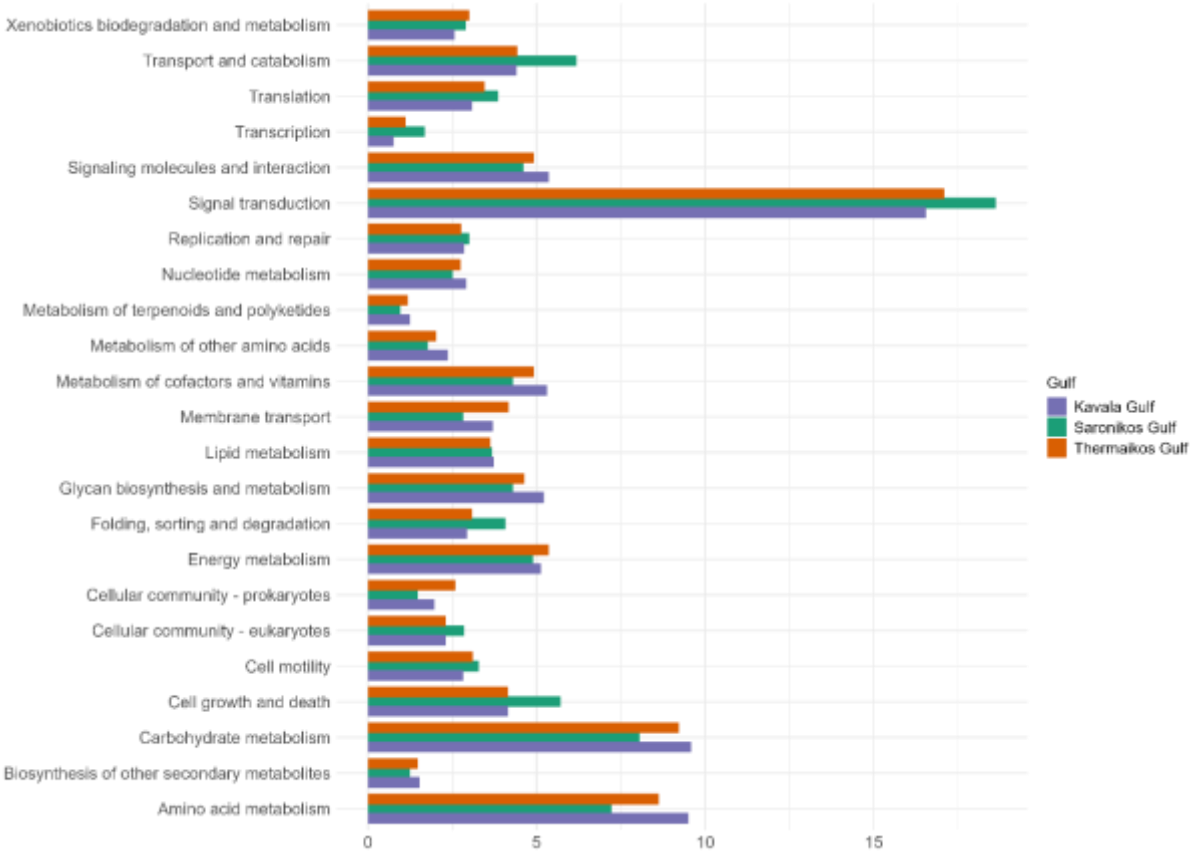

Figure 14. Relative abundance of KEGG global pathway categories across sampling regions based on shotgun metagenomic annotation. Pathway groups represent aggregated gene counts normalized to total annotated functional assignments. Colors correspond to the different examined gulfs.

Across all samples, genes associated with carbohydrate metabolism, amino acid metabolism, energy production and conversion, and lipid metabolism formed the core functional repertoire. Genes involved in cellular processes and signalling, including replication, transcription, translation, and repair pathways, represented the second most abundant cluster. Environmental processing functions, such as membrane transport, xenobiotic degradation, and biogeochemical transformations of nitrogen and phosphorus, were also consistently detected, although at lower relative abundance compared to core metabolic pathways.

Spatial comparisons indicate that while overall proportions remain similar, subtle yet ecologically meaningful differences exist among regions. Samples from Thermaikos and, to a lesser extent, Saronikos displayed elevated signals in categories associated with organic matter degradation, including carbohydrate-active enzyme systems

and polymer breakdown pathways. This aligns well with known biogeochemical characteristics of these gulfs, where elevated anthropogenic nutrient input and particulate organic matter availability promote heterotrophic and particle-attached microbial lifestyles. Despite these regional trends, pathway diversity remained relatively consistent across sampling locations, indicating that eDNA approaches successfully captured the functional repertoire of the present communities (Sunagawa et al., 2015). Compared with taxonomic data, functional category structure showed functional redundancy in the system.

At a finer module-level resolution, several functional features emerge that reflect ecological differences across the sampling sites. Genes related to nitrogen cycling, including nitrification, denitrification, and ammonification pathways, were present across all regions but exhibited more pronounced presence in Thermaikos Gulf. This enrichment mirrors increased detection of nitrifying bacteria and dinoflagellate-associated microbial consortia observed in amplicon-based results from this region. Notably, some samples also contained detectable signatures of nitrogen fixation. Genes associated with sulfur cycling, specifically pathways linked to dimethylsulfoniopropionate (DMSP) degradation and sulfate reduction, were present at low but measurable levels and appeared strongly associated with coastal rather than offshore samples. DMSP in particular is heavily produced by phytoplankton groups such as dinoflagellates and prymnesiophytes, both of which formed significant components of the eukaryotic dataset in Saronikos and Thermaikos Gulf. The presence of functional categories related to secondary metabolite synthesis, including terpenoid biosynthesis and antibiotic compound production, further supports the presence of complex ecological interactions and microbial competition within coastal communities. Environmental processing modules, such as ABC transporters, TonB-dependent receptors, and phosphotransferase systems (PTS), were consistently detected in all samples. These systems likely enable microbes to acquire nutrients efficiently under fluctuating resource regimes, especially in coastal zones where particulate organic matter, micronutrients, and dissolved organic carbon vary strongly across small spatial scales (Tully et al., 2018).

KEGG module completeness was used as a complementary approach to examine the coherence of functional pathways at the genome-equivalent level. Several modules were consistently complete across samples, including those involved in core oxidative phosphorylation reactions, ribosome assembly and transcription machinery, fatty acid biosynthesis and central sugar metabolism (glycolysis and pentose phosphate cycle) (Figure 15). The stability of these modules across geographically and taxonomically distinct communities indicates a conserved baseline of essential biochemical potential in the system. In contrast, modules related to specialized biosynthetic pathways, including toxin metabolism, light-harvesting complexes, and specific carbohydrate transport systems, exhibited region-specific patterns. Thermaikos Gulf in particular, exhibited elevated completeness across modules related to xenobiotic degradation and molecular response to environmental stressors. These included pathways for polycyclic aromatic hydrocarbon degradation, metal

resistance, and oxidative stress response systems (e.g., superoxide dismutase and peroxidase modules). The presence and completeness of these modules strongly align with known anthropogenic pressures such as maritime traffic, fisheries, and land-derived pollutants, suggesting microbial communities in impacted zones may act as functional sentinels of ecosystem stress.

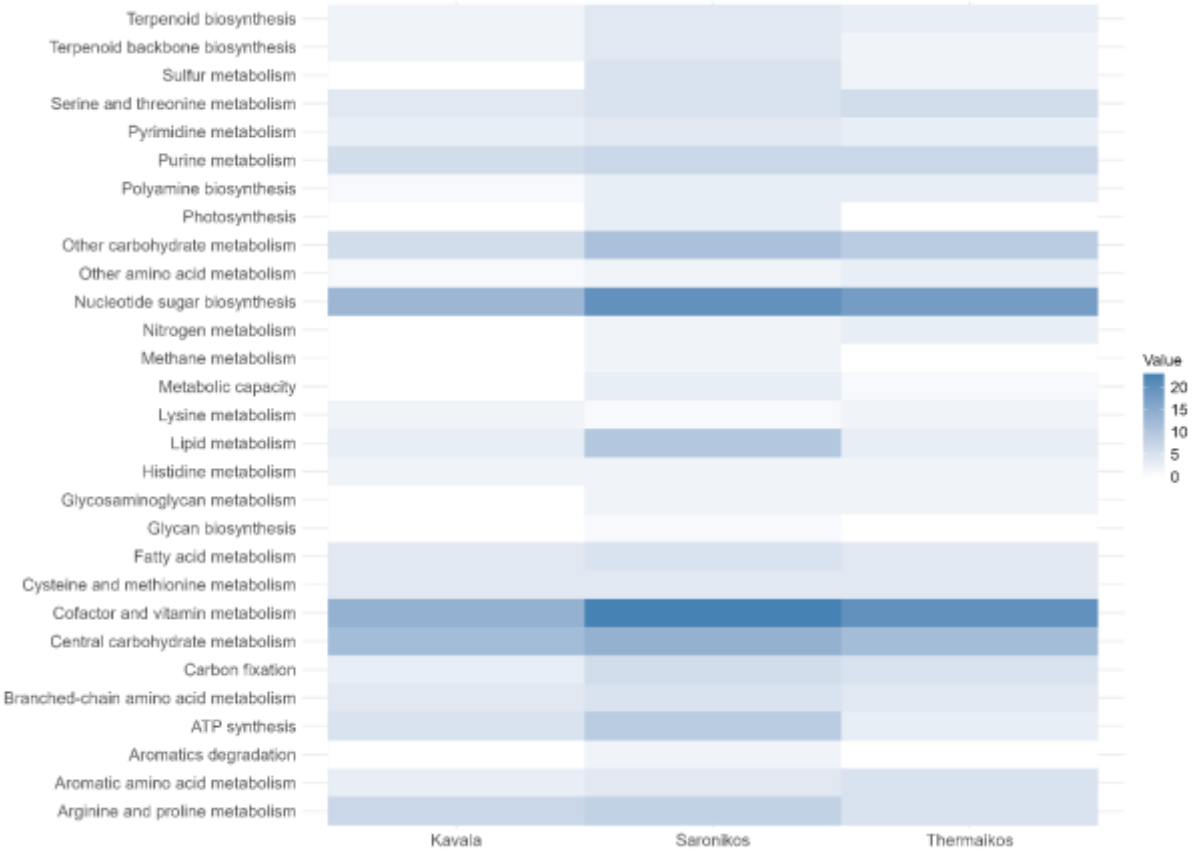

Figure 15. Heatmap showing the relative abundance of selected KEGG complete modules across the examined gulfs. Warmer colors indicate higher number of complete modules in the category.

Together, these metagenomic functional profiles demonstrate that while the examined coastal areas exhibit geographically structured communities, their functional capacity remains highly conserved. The coastal communities revealed enrichment in pathways associated with organic matter processing, stress adaptation, and nutrient cycling. Functional redundancy across sampling stations underscores the resilience of these systems and supports their role in maintaining stable ecosystem processes despite spatiotemporal variation and anthropogenic pressures.

4. Conclusions

This deliverable provides an integrated environmental DNA–based assessment of microbial and fish biodiversity and associated functional capacity across three ecologically and socio-economically important coastal basins in Greece and the adjacent offshore waters of the Aegean Sea. By combining multi-marker metabarcoding and metagenomic functional profiling with three complementary sampling approaches, this study demonstrates that eDNA can deliver a comprehensive and spatially resolved picture of marine biological communities at multiple trophic levels. The findings highlight clear biogeographic structuring across locations, particularly reflecting known gradients such as coastal eutrophication intensity, human activity, and natural environmental variability. Across all sampled regions, Alphaproteobacteria and key planktonic eukaryotic groups such as dinoflagellates, diatoms, and picoalgae dominated microbial community structure, while fish assemblages reflected spatial gradients with expected coastal residents and more pelagic signatures offshore. Moreover, no differences in fish communities were detected between the sea surface and the end of the euphotic zone in coastal zones. Spatial patterns were consistent and robust, demonstrating that eDNA captures ecologically meaningful variation aligned with local oceanographic and environmental dynamics.

A central outcome of this work is the demonstration that although taxonomic composition varied strongly across space, the functional profiles derived from shotgun metagenomics were comparatively consistent. These results suggest a high degree of functional redundancy within marine communities, where distinct taxonomic assemblages support similar metabolic potential. Carbon metabolism pathways, biosynthesis functions, and nutrient processing modules were pervasive across all regions, indicating that core ecosystem functions remain conserved despite local changes in species composition. Local enrichment in specific pathways, such as nitrogen cycling in Thermaikos or stress response and organic matter degradation functions near more impacted coastal areas, supports the value of eDNA-derived functional datasets as sensitive indicators of environmental condition and biogeochemical processes.

From a methodological standpoint, comparison of sampling techniques revealed that active filtration and Sterivex approaches produced more consistent biodiversity results across domains, while gauze sampling performed unevenly and occasionally under-represented biological diversity. Despite this variability, all methods detected major spatial gradients, reinforcing the flexibility and applicability of eDNA-based marine surveys.

Overall, the results underscore the utility of eDNA-based monitoring as a scalable, sensitive, and comprehensive approach for biodiversity mapping and ecological assessment in marine systems. Beyond establishing a high-resolution ecological baseline for coastal areas of Greece, this dataset provides a reference for future longitudinal monitoring within the NEMO-Tools framework and supports the transition

toward next-generation, non-invasive approaches to marine ecosystem observation. Continued integration of eDNA, environmental metadata, and long-term time-series assessments will enhance the capacity to detect change, assess ecological status, and support evidence-based marine management and policy implementation in both Greek and Mediterranean waters.

5. References

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). "obitools: A unix-Inspired Software Package for DNA Metabarcoding." Molecular Ecology Resources 16, no. 1: 176–182.

Bukin, Y. S., Galachyants, Y. P., Morozov, I. V., Bukin, S. V., Zakharenko, A. S., & Zemskaya, T. I. (2019). The effect of 16S rRNA region choice on bacterial community metabarcoding results. Scientific Data, 6(1), 190007. https://doi.org/10.1038/sdata.2019.7

Comeau, A. M., Li, W. K. W., Tremblay, J.-É., Carmack, E. C., & Lovejoy, C. (2011). Arctic Ocean Microbial Community Structure before and after the 2007 Record Sea Ice Minimum. PLOS ONE, 6(11), e27492. https://doi.org/10.1371/journal.pone.0027492

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 26(19), 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Froslev, T. G., Kjoller, R., Bruun, H. H., et al. (2017). "Algorithm for Post-Clustering Curation of DNA Amplicon Data Yields Reliable Biodiversity Estimates." *Nature Communications* 8: 1188. https://doi.org/10.1038/s41467-017-01312-x.

Haegeman, B., Hamelin, J., Moriarty, J., Neal, P., Dushoff, J., & Weitz, J. S. (2013). Robust estimation of microbial diversity in theory and in practice. The ISME Journal, 7(6), 1092–1101. https://doi.org/10.1038/ismej.2013.10

Kalu, E. I., Reyes-Prieto, A., & Barbeau, M. A. (2023). Community dynamics of microbial eukaryotes in intertidal mudflats in the hypertidal Bay of Fundy. ISME Communications, 3(1), 21. https://doi.org/10.1038/s43705-023-00226-8

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. Journal of Molecular Biology, 428, 726-731. DOI: 10.1016/j.jmb.2015.11.006.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glöckner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Research, 41(1), e1. https://doi.org/10.1093/nar/gks808

Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: An ultra-fast single-nod solution for large and complex metagenomics assembly via succinc de Brujin graph. Bioinformatics, 31, 1674-1676. DOI: 10.1093/bioinformatics/btv033.

Lo, L. S. H., Xu, Z., Lee, S. S., Lau, W. K., Qiu, J.-W., Liu, H., Qian, P.-Y., & Cheng, J. (2022). How elevated nitrogen load affects bacterial community structure and nitrogen cycling services in coastal water. Frontiers in Microbiology, 13. https://doi.org/10.3389/fmicb.2022.1062029

Louca, S., Parfrey, L. W., & Doeveli, M. (2016). Decoupling function and taxonomy in the global ocean microbiome. Science, 353, 1272-1277. https://doi.org/10.1126/science.aaf4507.

Mahe, F., Rognes, T., Quince, C., De Vargas, C., & Dunthorn, M. (2015). "Swarm v2: Highly-Scalable and High-Resolution Amplicon Clustering." PeerJ 3: e1420. https://doi.org/10.7717/peerj.1420.

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PloS One, 8(4), e61217. https://doi.org/10.1371/journal.pone.0061217

Miao, W., Wang, S., Lin, T., Yan, Y., Bao, Z., Zhang, D., Jiang, Z., & Zhang, H. (2024). Interaction patterns and assembly mechanisms of dinoflagellates and diatoms in a coastal bay suffering from long-term eutrophication. mSphere, 9(7), e00366-24. https://doi.org/10.1128/msphere.00366-24

Miya, M., Sato, Y., Fukunaga, T., et al. (2015). "MiFish, a Set of Universal PCR Primers for Metabarcoding Environmental DNA From Fishes: Detection of More Than 230 Subtropical Marine Species." Royal Society Open Science 2, no. 7: 150088. https://doi.org/10.1098/rsos.150088.

Mordret, S., Piredda, R., Zampicinini, G., Kooistra, W. H. C. F., Zingone, A., Montresor, M., & Sarno, D. (2023). Metabarcoding reveals marked seasonality and a distinctive winter assemblage of dinoflagellates at a coastal LTER site in the Gulf of Naples. Marine Ecology, 44(3), e12758. https://doi.org/10.1111/maec.12758

Nagai, T., Shiba, T., Komatsu, K., Watanabe, T., Nemoto, T., Maekawa, S., Kobayashi, R., Matsumura, S., Ohsugi, Y., Katagiri, S., Takeuchi, Y., & Iwata, T. (2024). Optimal 16S rRNA gene amplicon sequencing analysis for oral microbiota to avoid the potential bias introduced by trimming length, primer, and database. Microbiology Spectrum, 12(12), e03512-23. https://doi.org/10.1128/spectrum.03512-23

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P. et al. (2022). Vegan: community ecology package. R package version 2.6-2. http://CRAN.R-project.org/package=vegan

Parulekar, N. N., Kolekar, P., Jenkins, A., Kleiven, S., Utkilen, H., Johansen, A., Sawant, S., Kulkarni-Kale, U., Kale, M., & Sæbø, M. (2017). Characterization of bacterial community associated with phytoplankton bloom in a eutrophic lake in South Norway

using 16S rRNA gene amplicon sequence analysis. PLOS ONE, 12(3), e0173408. https://doi.org/10.1371/journal.pone.0173408

Pruesse, E., Peplies, J., & Glöckner, F. O. (2012). SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics, 28(14), 1823–1829. https://doi.org/10.1093/bioinformatics/bts252

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Research, 41(Database issue), D590-596. https://doi.org/10.1093/nar/gks1219

R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). "VSEARCH: A Versatile Open Source Tool for Metagenomics." PeerJ 4: e2584. https:// doi. org/ 10. 7717/ peerj. 2584.

Schloss, P. D., Gevers, D., & Westcott, S. L. (2011). Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies. PLOS ONE, 6(12), e27310. https://doi.org/10.1371/journal.pone.0027310

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., & Weber, C. F. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Applied and Environmental Microbiology, 75(23), 7537–7541. https://doi.org/10.1128/AEM.01541-09

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics, 30, 2068-2069. https://doi.org/10.1093/bioinformatics/btu153

Sunagawa, S., Coelho, L. P., Chaffron, S., et al. (2015). Structure and function of the global ocean microbiome. Science, 348, 1261359. https://doi.org/10.1126/science.1261359.

Tully, B. J., Graham, E. D., & Heidelberg, J. F. (2018). The reconstruction of 2631 draft metagenome-assembled genomes from the global oceans. Scientific Data, 5, 170203. https://doi.org/10.1038/sdata.2017.203.

Yang, X., Liu, Z., Zhang, Y., Shi, X., & Wu, Z. (2024). Dinoflagellate–Bacteria Interactions: Physiology, Ecology, and Evolution. Biology, 13(8), 579. https://doi.org/10.3390/biology13080579