communications biology

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42003-026-10977-4

DNA-based monitoring uncovers contrasting salinity-driven diversity patterns of bacterial and protist plankton in the Baltic Sea area

Check for updates

Krzysztof T. Jurdzinski[1] ✉, Meike AC Latz[1,2], Anders Torstensson[3], Sonia Brugel[4,5], Mikael Hedblom[3], Yue O. O. Hu[1,6,7], Markus Lindh[3], Agneta Andersson[4,5], Bengt Karlson[3] & Anders F. Andersson[1] ✉

Salinity has been identified as a primary factor structuring community composition in global diversity assessments of free-living bacteria in non-extreme environments. It remains unknown whether the same is true for protists. Moreover, changes in alpha diversity of both bacteria and protists across the brackish salinity gradient remain elusive. Here, we analyzed protist and bacterial diversity using DNA metabarcoding across 398 surface water samples synchronized with national monitoring of the Baltic Sea and the Kattegat-Skagerrak. We show that salinity, unlike other environmental factors, had a stronger effect on bacterial than on protist community composition. Likewise, protist taxonomic units tended to have broader salinity niches than bacterial ones. In contrast, protist but not bacterial alpha diversity (richness) increased with salinity. Bacterial alpha diversity was similar across salinities despite drastic changes in community composition, as marine-associated taxa steadily replaced freshwater-associated ones. Conversely, marine-associated protists dominated across most of the salinity gradient, with fewer taxonomic units adapted to lower salinities. Overall, the diversity of protists but not bacteria followed patterns consistent with local adaptation as observed in other groups of organisms. Ultimately, the observed patterns form a baseline for environmental monitoring and suggest unexplored differences between bacterial and protist eco-evolutionary dynamics.

Since the earliest global surveys of bacterial diversity, salinity has emerged as the strongest factor structuring the community composition of free-living bacteria[1–4]. It is well established that freshwater-marine transitions have been infrequent events in microbial evolution, separating bacterial as well as protist communities[5–7]. However, the freshwater-marine distinction is a binary one, omitting the brackish salinities in between (0.5-30 PSU). Brackish waters comprise a volume at least comparable with that of global lakes[8–10], and are mostly concentrated in coastal zones and inland lakes, areas of markedly high importance for and pressure from human activities. Moreover, these ecosystems are highly dependent on microorganisms, which are responsible for most of the primary production and decomposition in these ecosystems[11]. Finally, brackish waters host a distinct microbiome, with a vast majority of bacterial species lacking representatives in freshwater and marine environments[8,12].

A number of studies have assessed the impact of salinity on bacterial[13–22] and protist[16,18,21–25] plankton communities in brackish waters. Generally, a pronounced shift in community composition along the salinity gradient has been observed for both bacteria and protists. Bacterial community composition changes more with salinity than with season[15,17], but it remains unclear whether the same is true for protists. A few studies have

[1]KTH Royal Institute of Technology, Department of Gene Technology, Science for Life Laboratory, Stockholm, Sweden. [2]University of Copenhagen, Department of Plant and Environmental Sciences, Frederiksberg C, Denmark. [3]Swedish Meteorological and Hydrological Institute, Västra Frölunda, Sweden. [4]Umeå University, Department of Ecology, Environment and Geoscience, Umeå, Sweden. [5]Umeå Marine Sciences Centre, Umeå University, Norrbyn, Sweden. [6]Present address: State Key Laboratory of Ecological Safety and Sustainable Development in Arid Lands, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, China. [7]Present address: University of Chinese Academy of Sciences, Beijing, China. ✉e-mail: krzysztof.jurdzins@scilifelab.se; anders.andersson@scilifelab.se

compared beta diversity patterns of bacterial and protist brackish plankton[16,18,21,22], yet it is not established whether the community composition of both microbial groups is similarly affected by abiotic factors. Results obtained across steep salinity gradients in Antarctica suggested stronger environmental filtering among planktonic bacteria than protists[16,26], while studies in other environmental contexts (depth and time) came to the opposite conclusion[27,28]. This indicates that salinity, unlike other factors, has a greater impact on bacterial than protist community composition.

Apart from changes in community composition (beta diversity), the patterns of species richness (alpha diversity) are of high interest for understanding and monitoring brackish ecosystems. A classical model from the 1930s, Remane's *Artenminimum* (species-minimum) concept, developed for macrozoobenthos, predicts a minimum in diversity across the salinity gradient at 5-8 PSU, i.e., at the "horohalinicum"[29]. However, studies on fish[30], and macroalgae[31] have shown a steady increase in richness with salinity, challenging the generalizability of Remane's model. For phytoplankton, patterns of both a maximum[32] and a minimum[33] of species richness within the mid-range brackish salinity spectrum, as well as an increase along the salinity gradient[34], have been reported based on microscopic assessments. In contrast, DNA metabarcoding transect surveys of plankton showed no clear patterns for bacteria[14,17,18] nor for major protist taxa[18]. At the same time, bacterial alpha diversity has been observed to be higher in winter than in summer in the Baltic Sea[17,35], consistent with patterns observed in marine environments[36–38]. Little is known about the factors driving the seasonal patterns in the Baltic Sea or how the seasonal patterns may confound the alpha diversity patterns along the brackish salinity gradient.

The Baltic Sea and the adjacent Kattegat-Skagerrak (the Baltic Sea area, Fig. 1a) are among the best-studied brackish waters and coastal areas[39]. Recent formation (8000 years ago[40]), pronounced environmental gradients, and narrow connecting straits make the Baltic Sea a useful model for studying eco-evolutionary processes driven by dispersal and adaptation[41]. In addition, its early exposure to anthropogenic pressures can foreshadow the processes awaiting other coastal regions[42]. The Baltic Sea area also has a long record of microbial research, with the earliest data coming from the 19th century[43]. Regular microscopy-based phytoplankton monitoring has been run by multiple countries in the region for decades[44] and since the 1980s in Sweden[45]. More recently, the Baltic Sea area was one of the first brackish waters surveyed using DNA metabarcoding[14,18,46] and metagenomics[12,47]. Nevertheless, most of the unknown diversity in the Baltic Sea area is believed to be microbial[48].

Here, we present the first in-depth comparative analysis of bacterial and protist diversity across the spatiotemporal scales in the Baltic Sea area and the corresponding environmental gradients. We use a recently published DNA metabarcoding dataset[49] and a new dataset released here, both originating from samples collected along the Swedish National Marine Monitoring Program. The datasets together span more than two complete calendar years and include 398 samples. We contrast the alpha and beta diversity patterns of bacteria and protists along brackish salinities and other environmental gradients of the Baltic Sea area. We link the differences between the bacterial and protist beta diversity patterns to systematic differences in salinity niche breadths. Finally, we connect our metabarcoding results with publicly available sequences from freshwater and marine samples, showing that distinct distributions of freshwater- and marine-associated taxonomic units underlie the contrasting bacterial and protist alpha diversity patterns across the brackish salinity gradient.

Results and discussion

Regionally distinct seasonal dynamics of major bacterial and protist taxa

We first analyzed a 16S and 18S rRNA gene metabarcoding dataset recently published by us, coming from over a year of biweekly sampling at multiple locations covering broad environmental gradients[49] (Fig. 1a, d). Amplicon sequence variants (ASVs) were clustered by simultaneous genetic and distribution similarity, using distribution-based operational taxonomic unit-calling[50] (we call these dbOTUs). We thus obtained 16,295 bacterial and 3935 protist dbOTUs. More than half of the dbOTUs were annotated to at least family level for both bacteria and protists (Supplementary Fig. 1c, d). Moreover, more than half of the reads represent dbOTUs annotated to at least genus or family level for bacteria and protists, respectively (Supplementary Fig. 1e, f). Because the samples also contained a known amount of synthetic spike-in DNA amplifiable by the metabarcoding primers, absolute abundances of individual dbOTUs could be estimated.

The major bacterial (Fig. 1b) and protist (Fig. 1c) taxa displayed distinct seasonal abundance dynamics. Both their abundances and seasonal patterns differed between the Kattegat/Skagerrak, and the three basins of the Baltic Sea included in this study (Baltic Proper, Bothnian Sea, and Bothnian Bay). Note that the abundances are given in rRNA gene copies, and as protists often have much higher rRNA gene copy numbers than bacteria[51], they appear similarly abundant despite being less numerous in reality. Indeed, 18S rRNA gene count-derived abundances of phytoplankton classes have been shown to correlate better with biomass than cell counts[52], and protists have higher biomass per cell and their total biomass in the Baltic Sea water column is comparable to bacteria[53].

We observed diatoms peaking and dinoflagellates starting to peak in abundance in March in the Baltic Proper (Fig. 1c), consistent with microscopy-based observations[54,55]. However, the dynamics and abundance of dinoflagellates and diatoms differed vastly between basins. Still, in all cases except for the Bothnian Bay, those groups increased in abundance in the first half of the year. Cyanobacteria increased in abundance as early as May in the Baltic Proper (Fig. 1b), as observed using microscopic methods[55]. The increase in cyanobacterial abundance was smaller and peaked later, moving northwards within the Baltic Sea. *Haptophyta* generally displayed increased abundance between May and August. In the Kattegat-Skagerrak, there was an additional dinoflagellate peak in September, corresponding to autumn blooms. *Cryptophyta* were abundant in the Baltic Sea, with increased abundance in the second half of the year.

We also observed geographic differences in abundance and seasonal dynamics of major heterotrophic bacteria. *Bacteroidia* (Fig. 1b) peaked in abundance in May across the Baltic Sea, consistent with previous reports[35,53], but displayed less clear seasonal dynamics in the Skagerrak and the Kattegat. *Actinomycetia* and *Acidimicrobiia* were abundant in the Baltic Sea but not in the Kattegat-Skagerrak, consistent with *Actinobacteria* abundance patterns from transect-based studies[14,18,47]. Additionally, *Acidimicrobiia* decreased in abundance between May and July, a dynamic that was more pronounced moving southwards in the Baltic Sea. Finally, contrary to previous results from a near-coast station[35,53], *Alphaproteobacteria*, and *Gammaproteobacteria* comprised a stable proportion of the whole community throughout the year in all basins but the Kattegat (Supplementary Fig. 2a). Nevertheless, consistent with metabarcoding[14] and metagenomic transects[47], *Alphaproteobacteria* and *Gammaproteobacteria* were more abundant in the Kattegat-Skagerrak than in the Baltic Sea.

Compared to relative abundance-based analysis, spike-in correction added ecologically relevant information. For example, the increase in dinoflagellate abundance in the Baltic Proper was observed in March with the spike-in approach (Fig. 1c), consistent with a microscopy-based report for the same year[55]. Meanwhile, their relative abundance increased only in April (Supplementary Fig. 2b), potentially being masked by a simultaneous increase in diatom abundance in March. Nevertheless, the spike-in approach appeared to generate substantial noise, best exemplified by the outlier results for total rDNA counts in some situations when the results were averaged across few samples (Fig. 1b, c), and relative abundances better captured gradual succession patterns, especially for bacteria (Supplementary Fig. 2a). Thus, spike-in correction is useful in cases where absolute abundances are crucial, e.g., bloom detection, but improvements are needed to reduce noise.

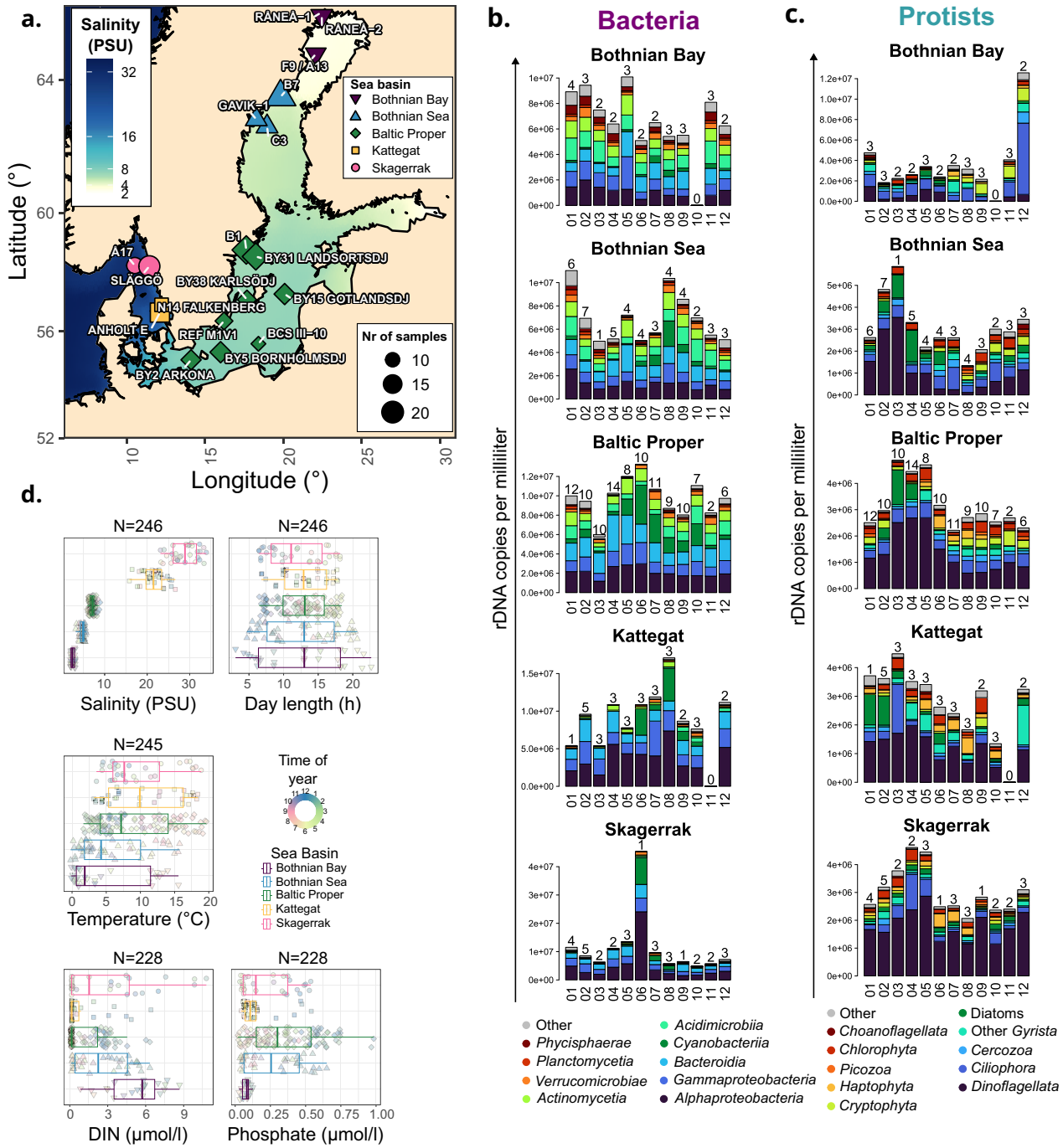


Fig. 1 | Spatiotemporal changes in abundance of major microbial taxa, 2019-2020. a A map of 18 sampling locations labeled by the name of the station and size-coded by the number of samples collected over a period of 13 months (January 2019 to February 2020). The map was made based on Bio-ORACLE[102] and GSHHG[103] data, and using the PBS mapping R package[104]. **b**, **c** The numbers of rRNA marker gene copies per liter annotated to (**b**) bacterial classes and (**c**) protist subdivisions averaged over each calendar month and Baltic Sea area basin (as marked in **a**). The values are based on the number of reads relative to spike-in reads and the amount of spike-in added per volume of sample. Only the classes/subdivisions which, on average, corresponded to >1% of reads across all samples are distinguished, and the rest is assigned to the group "Other". *Gyrista* has been divided into diatoms and other *Gyrista*. Above each bar, the number of samples collected in the respective month and basin is given. Note: Two samples (January in Bothnian Bay and September in Bothnian Sea) were removed for protists due to spike-in detection failure. **d** The distribution of measurements of major physicochemical and geographic parameters for each basin. The lines on the boxplot represent (from the middle) median, 75th percentile, and the whiskers extend to 1.5 times the interquartile range. The number of samples for which the relevant measurement was available is given above each plot. DIN - dissolved inorganic nitrogen.

Salinity has a dominant impact on bacterial but not protist community composition

The dissimilarities of bacterial and protist communities between the same pairs of samples were correlated ($R^2 = 0.58$). However, the protist communities were generally less similar to each other than bacterial ones (Supplementary Fig. 3a), as has been reported for picoeukaryotes and bacteria in ocean surface waters[56].

Principal Coordinate Analysis (PCoA) showed stronger geographic than seasonal structuring of bacterial communities, while for protists the effects of geography and season were comparable (Supplementary Fig. 4). To disentangle the environmental factors driving those structuring effects, we performed distance-based Redundancy Analysis (dbRDA). We focused on five fundamental environmental factors: salinity, temperature, day length (i.e., daytime length), dissolved inorganic nitrogen (DIN), and phosphate. These factors did not strongly correlate with each other ($R^2 < 0.4$, Supplementary Fig. 5) and were measured for the majority of the samples. While we do not expect day length to have strong direct impacts on most planktonic microbes, it has been repeatedly correlated with seasonal changes in bacterial alpha diversity[36,38]. We thus use it as a hypothetical proxy for commonly observed multifactorial seasonal changes.

The dbRDA-based variation partitioning showed that salinity could explain around half of the variation in bacterial community composition (Fig. 2a). The effect of salinity on protist community composition was much smaller (~ 13%), while temperature and nutrients were of greater relative importance than for bacteria (Fig. 2b). These results contradict the proposed rule that the distribution of smaller organisms is less limited by environmental factors than that of larger organisms[57], and in particular that bacteria are less environmentally filtered than protists[27]. Rather, our results are consistent with a previous report of a stronger selection of bacterial than protist communities across a salinity gradient[16,26]. However, we do not find support for the generalization of the stronger selection by other environmental factors[26].

Nevertheless, for both bacterial and protist communities, there was a pronounced shift coinciding with the salinity divide between the Baltic Sea and the Kattegat-Skagerrak and a more gradual change along the Baltic Sea salinity gradient (Fig. 2c, d). We identified microbial families whose abundance most strongly followed these changes in dbOTU-based beta diversity (Fig. 2c, d). For example, typically freshwater *Nanopelagicaceae* (acI *Actinobacteria*)[58] and *Limnocylindraceae* (*Chloroflexi*)[59] were associated with the lowest salinity samples, while commonly marine *Rhodobacteraceae*[60] and *Thioglobaceae*[61] were at the other end of the salinity spectrum (Fig. 2c). Regardless, the observed geographic patterns of beta diversity were also surely influenced by other abiotic and biotic factors such as climate, loads and composition of terrestrially derived organic matter, and particulate nutrient concentrations.

Both bacterial and protist communities underwent gradual community composition changes throughout the year in all the analyzed basins of the Baltic Sea area (Fig. 2e, f, Supplementary Fig. 6). Diverse groups significantly contributed to the seasonal patterns in beta diversity, ranging from known seasonally blooming diatoms *Thalassiosiraceae* and *Skeletonemaceae*[62] to multiple uncharacterized *Bacteroides* families. In all the basins, picocyanobacteria (*Cyanobiaceae*) were associated with high temperatures, while BACL11 (*Bacterioides*) were typical of winter conditions (Fig. 2e, Supplementary Fig. S6a).

Interannual recurrence of spatiotemporal patterns of microbial community composition

We further assessed whether the identified beta diversity patterns could be reproduced in an independent dataset from a different timeframe. We thus performed metabarcoding on stored DNA extracted from samples collected between 2015 and 2017 (Fig. 3a, Supplementary Fig. S7a, see Methods for details).

Overall, the community composition in the two datasets differed minimally, following the spatial and seasonal environmental gradients analogously (Fig. 3b, c). This apparent spatiotemporal recurrence of community composition is consistent with previous observations from the Baltic Sea[53,63] and other waters[28]. Moreover, the similarity of communities between the two datasets shows that the technical differences (see Methods for details) had little impact on the observed general community composition (Supplementary Fig. S7b-c). However, the datasets were separated by only two years, and the interannual stability is unlikely to hold in the longer timeframe as global change progresses. For example, there is strong evidence for a long-term change in diatom-dinoflagellate ratio during spring bloom in the Baltic Sea[62].

In contrast to community composition, alpha diversity differed significantly ($P < 10^{-14}$ in both cases) for both bacteria and protists (Fig. 3d, e). We doubt the increased richness in the 2015-2017 dataset, amounting to hundreds of dbOTUs on average, represented an actual biological change. Indeed, switching to downsampling raw reads instead of downsampling dbOTU counts removed nearly all of the difference in dbOTU richness between the 2015-2017 and 2019-2020 datasets (Fig. 3f). We also checked that downsampling raw reads consistently resulted in lower richness than downsampling the ASV counts (Supplementary Fig. S8). This is consistent with a previous report showing that downsampling before obtaining ASVs better removes sequencing-depth biases[64]. However, a small but significant difference in richness between datasets remained after downsampling reads (Fig. 3f), indicating that other technical or biological factors also contributed to the difference in alpha diversity between the datasets.

Seasonal vertical mixing brings deep-water bacteria to the surface and increases bacterial alpha diversity

Next, we analyzed the changes in alpha diversity along the salinity gradient. We applied the same statistical methods as were used in previous studies to argue for or against maximum[18,32] and minimum[33] protist diversity in mid-range brackish salinities (Fig. 4a, b). The horohalinicum, i.e., midrange salinities (5-8 PSU), displayed lower bacterial richness than in less (but not more) saline waters (Fig. 4a). The minimum in bacterial richness was at 15 PSU according to the generalized additive models. However, the minimum estimation was likely confounded by seasonal changes in salinity within Kattegat, as there was clear seasonal variation in bacterial richness overshadowing the geographic differences (Fig. 4a, c).

Across geographic regions, bacterial alpha diversity rapidly declined in spring and increased again in autumn (Fig. 4c). Out of the assessed environmental factors, DIN explained the largest part of the variation (Supplementary Fig. S9a).

Spring depletion of nitrogen in the Baltic Sea is closely connected to the emergence of the summer thermocline and thus reduced vertical mixing[65]. Winter vertical mixing has been linked to elevated bacterial alpha diversity and similarity in cytometric properties across the water column in a marine time series[66]. Additionally, some members of typically deepwater clades are rare taxa in surface waters with abundance maxima in winter[67]. The influx of deepwater bacteria has also been suggested to explain the higher diversity of bacteria in winter than in summer in the Baltic Sea, with winter communities being less structured by depth[17]. However, it remains unclear whether the influx of deepwater taxa or changes in environmental conditions concurrent with vertical mixing underlie the diversity pattern.

Therefore, we connected our results with a previously published metabarcoding dataset spanning different depths across a winter and a summer transect in the Baltic Sea[17], from here on referred to as the "transect-based dataset". We matched the sequences representative of the operational taxonomic units (OTUs) from the transect-based dataset to ASVs representative of our dbOTUs (see Methods for details).

We further chose from the transect-based dataset only the summer (July) samples. Among those, we categorized OTUs as found only in deep (> 10 m depth) or in surface (≤ 10 m depth) waters, or in both. The surface mixed layer in summer in the Baltic Sea is usually located above 10 m depth[65]. We assessed how many OTUs from each category matched the dbOTUs from winter (December–February) and summer (June–August) months (Fig. 5a). Among the deepwater OTUs, significantly more ($P < 10^{-73}$) were found in our surface water samples in winter than in summer, despite originating from summer samples.

The proportion of detected deepwater OTUs in a sample strongly correlated with dbOTU richness (Fig. 5b). In contrast, the proportion of detected non-deepwater-specific OTUs ('surface' and 'both' from Fig. 5a) showed a weaker correlation with richness (Supplementary Fig. S9c), and did not follow the proportion of deepwater OTUs matched (Supplementary Fig. S9d). Thus, the correlation between the proportion of deepwater OTUs

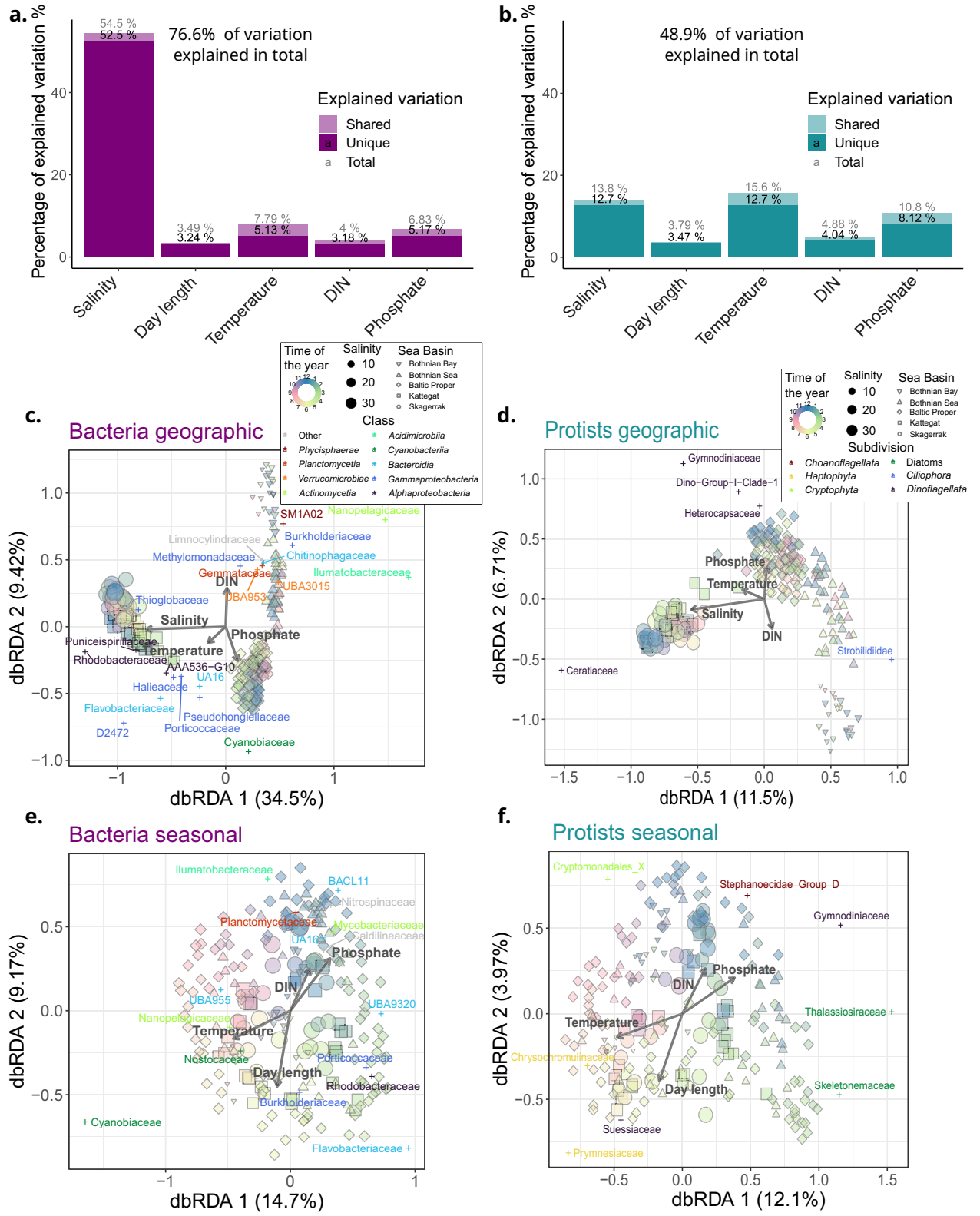


Fig. 2 | Environmental factors explaining geographic and seasonal community structuring and the families of bacteria and protists most strongly correlated with the beta diversity patterns. a, **b** Variance partitioning of bacterial (**a**) and protist (**b**) beta diversity, based on Bray-Curtis distances and distance-based Redundancy Analysis (dbRDA). **c**, **d** Partial dbRDA conditioned on seasonality (sine and cosine of the day of the year), highlighting bacterial (**c**) and protist (**d**) geographic community structuring. **e**, **f** Partial dbRDA conditioned on salinity and latitude, highlighting bacterial (**e**) and protist (**f**) seasonal community structuring. **c**–**f** Weighted averages of families significantly deviating from the general distribution of families across the axes (Bonferroni-corrected $P < 0.05$) were placed on the plot, with values rescaled by a factor of 0.5. The families were color-coded according to classes/subdivisions (low abundance classes are represented by "Other"). *Gyrista* was divided into diatoms and other *Gyrista*. Percentages of variation explained by each dbRDA component are given in brackets by the axes labels. The same number of samples ($N = 228$) was used for all the panels. DIN - dissolved inorganic nitrogen.

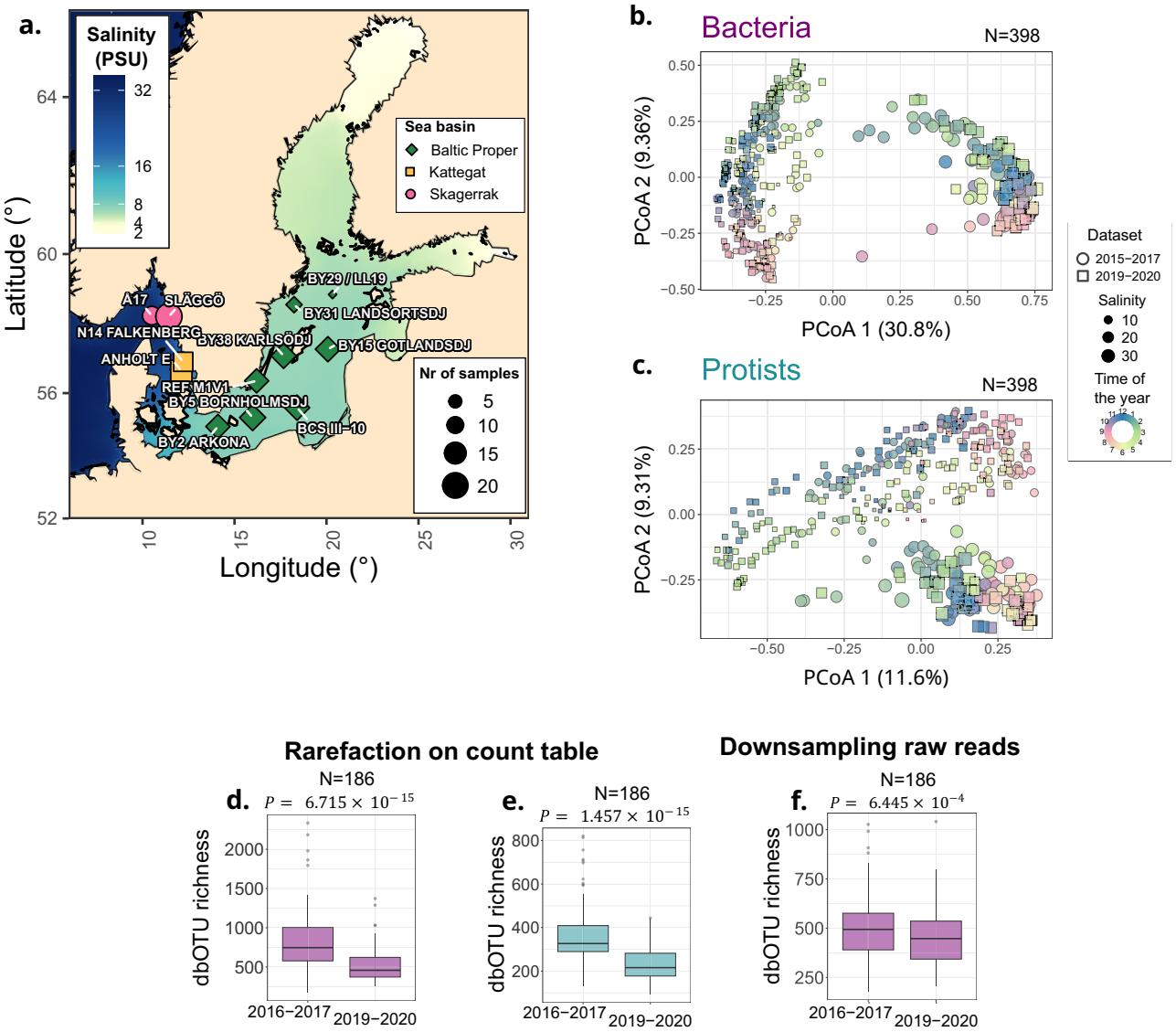


Fig. 3 | Bacterial and protist diversity across datasets collected in 2015–2017 and in 2019–2020. a A map of 12 sampling locations labeled by the number of samples collected between August 2015 and March 2017 (the 2015-2017 dataset), with all the 2015 samples coming from station Släggö in Skagerrak. The map was made based on Bio-ORACLE[102] and GSHHG[103] data, and using the PBS mapping R package[104]. **b**, **c** PCoA based on Bray-Curtis dissimilarities of bacterial (**b**) and protist (**c**) communities across the 2015–2017 and 2019–2020 datasets. **d**, **e** dbOTU richness of bacteria (**d**) and protists (**e**) in the two datasets. Samples of both datasets were rarefied to the same number of reads. Only dbOTU richness values for stations present in both datasets have been used. Moreover, samples were chosen in pairs, from the same station and the same month of the year but different datasets. Sample pairs with the lowest distance in terms of the day in a calendar year were chosen. *P*-values obtained using the pairwise, two-sided Wilcoxon (signed-rank) test are given above each plot. **f** Same as (**d**) but based on richness values obtained by down-sampling raw reads. The differences in the number of samples come from the availability of physicochemical parameters or (**d**–**f**) the interannual data balancing approach. The number of samples (N) used to create the plots is given above each panel. DIN - dissolved inorganic nitrogen, dbOTU - distribution-based operational taxonomic unit.

and richness did not originate from detection biases. Moreover, after partitioning out the correlations with deepwater OTUs, DIN no longer correlated with dbOTU richness (Fig. 5c). This relation was not bilateral; the proportion of deepwater OTUs still correlated with dbOTU richness after partitioning out DIN (Supplementary Fig. S9e). Therefore, seasonal convective water mixing increases winter alpha diversity by bringing deepwater bacteria to the surface and increasing the observed richness, consistent with previous reports[17,36,38,66,67].

In winter, deepwater taxa not only increased alpha diversity but also comprised a substantial part of the surface water bacterial communities, up to >30% in Skagerrak (Fig. 5e). With cumulative relative abundance commonly above 10% in winter, the deepwater taxa likely impact the surface water ecosystem, if not through metabolic activity, at least by providing alternative food sources for heterotrophs. A metatranscriptomic approach would be needed to determine whether the deepwater bacteria are active in surface waters in winter.

Ultimately, the higher mean bacterial richness in the low salinity area of the Bothnian Sea and the Bothnian Bay (Fig. 4a) can be explained by differences in seasonal dynamics, i.e., a longer period of vertical mixing due to colder climates[9]. In summer, when mixing ceased, richness dropped to similar levels across basins (Fig. 4c), consistent with previous findings reporting no salinity-dependent pattern of bacterial alpha diversity[14,17,18].

Higher protist alpha diversity in near-marine salinities

In contrast to the situation for bacteria, protist richness displayed a clear salinity-dependent pattern, monotonically increasing with salinity (Fig. 4b;

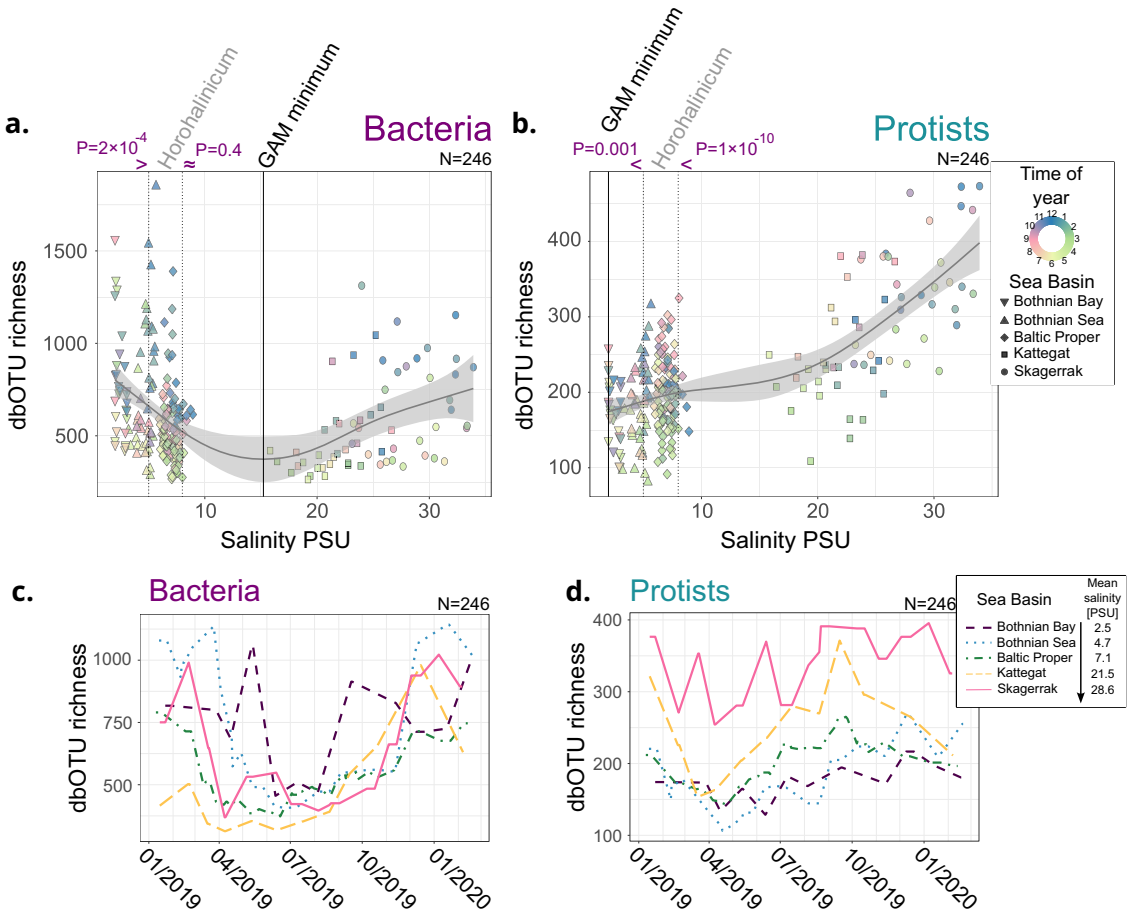


Fig. 4 | Bacterial and protist richness. Mean number of bacterial (**a**) and protist (**b**) dbOTUs found across 100 rarefaction iterations (dbOTU richness). The fit corresponds to smoothened generalized additive models (GAMs). The dotted vertical lines correspond to boundaries of horohalinicum (5-8 PSU). The colored (in)equality signs over the dotted lines mark the significant differences in dbOTU richness between horohalinicum and lower/higher salinities, and the *P*-values from two-sided Wilcoxon rank-sum tests are written above. The solid vertical lines mark salinities corresponding to minimum richness in the GAMs. The same legend applies to (**a**) and (**b**). The solid gray line represents the richness values predicted by the GAMs, and the gray area corresponds to ± standard error. **c**, **d** Rolling means of bacterial (**c**) and protist (**d**) richness for the distinguished Baltic Sea area basins. The richness values were averaged for measurements 2 weeks before and after each date. The same legend applies to (**c**) and (**d**). The number of samples (N) used to create the plots is given above each panel. The differences in the number of samples come from the availability of physicochemical parameters.

linear fit yields $R^2 = 0.44$). This is consistent with a recent microscopy-based study[34], while it contrasts with reports of other patterns[18,32,33]. However, unlike in the microscopy study by Olofsson et al.[34], richness in the near-marine Kattegat-Skagerrak was higher than in the Baltic Sea across the year (Fig. 4d).

Protist dbOTUs have broader salinity niches than bacterial dbOTUs

Salinity had a stronger impact on bacterial beta diversity (Fig. 2) but protist alpha diversity (Fig. 4). Alpha diversity is unaffected if one taxonomic unit is replaced with another, and a taxon's impact on beta diversity is proportional to its abundance. Thus, based on the diversity patterns, it is unclear whether all bacteria or only the abundant ones have narrower salinity niches than their protist counterparts. We therefore proceeded to more directly analyze the realized salinity niches of bacterial and protist dbOTUs.

First, we estimated the salinity niches of dbOTUs found only in salinities higher (> 15 PSU) or lower (< 9 PSU) than in the Danish Straits (Fig. 6a-d). We thus based each analysis on a unimodal distribution of salinity values. Among the Baltic Sea dbOTUs (lower salinities), bacterial ones had broader salinity niches (Fig. 6a), while in Kattegat/Skagerrak (higher salinities), it was protist dbOTUs that were usually found across a broader range of salinities (Fig. 6b). However, the differences between the groups were small compared to the spread of salinity niche breadths within the groups. It is expected that microbial niches are limited by multiple factors, which leads to salinity-independent limitations in dbOTUs detection range. To correct for the salinity-independent effects, we built a Bayesian model of bacterial and protist salinity niche breadths based on temperature, DIN, phosphate, and day length niche breadths. We then predicted and compared salinity niche breadths for hypothetical bacterial and protist dbOTUs with the same (mean) niche breadths across other factors (Fig. 6c, d). After correcting for other factors, the bacterial salinity niches in the Baltic Sea were not broader than for protists (Fig. 6c), while the trend towards broader salinity niches of protists than bacteria in Kattegat/Skagerrak became even stronger (Fig. 6d).

To estimate how commonly bacteria and protists have salinity niches spanning across the Danish Straits, we analyzed the proportions of dbOTUs found on both sides of the salinity barrier. To limit the effect of other factors on the analysis, we selected two stations on each side of the salinity barrier, keeping only near-marine stations (in Skagerrak) and stations in the Baltic Proper considerably far from the barrier itself, to lower the risk of detecting temporary barrier-crossing plankton due to storm events, transport with water currents, or other causes. We combined the 2019-2020 and 2015-2017 datasets and used samples as close in time as possible, resulting in 88 samples in total (22 per station, see Methods for details). A significantly higher proportion of protist dbOTUs occurred across the salinity barrier (Fig. 6e).

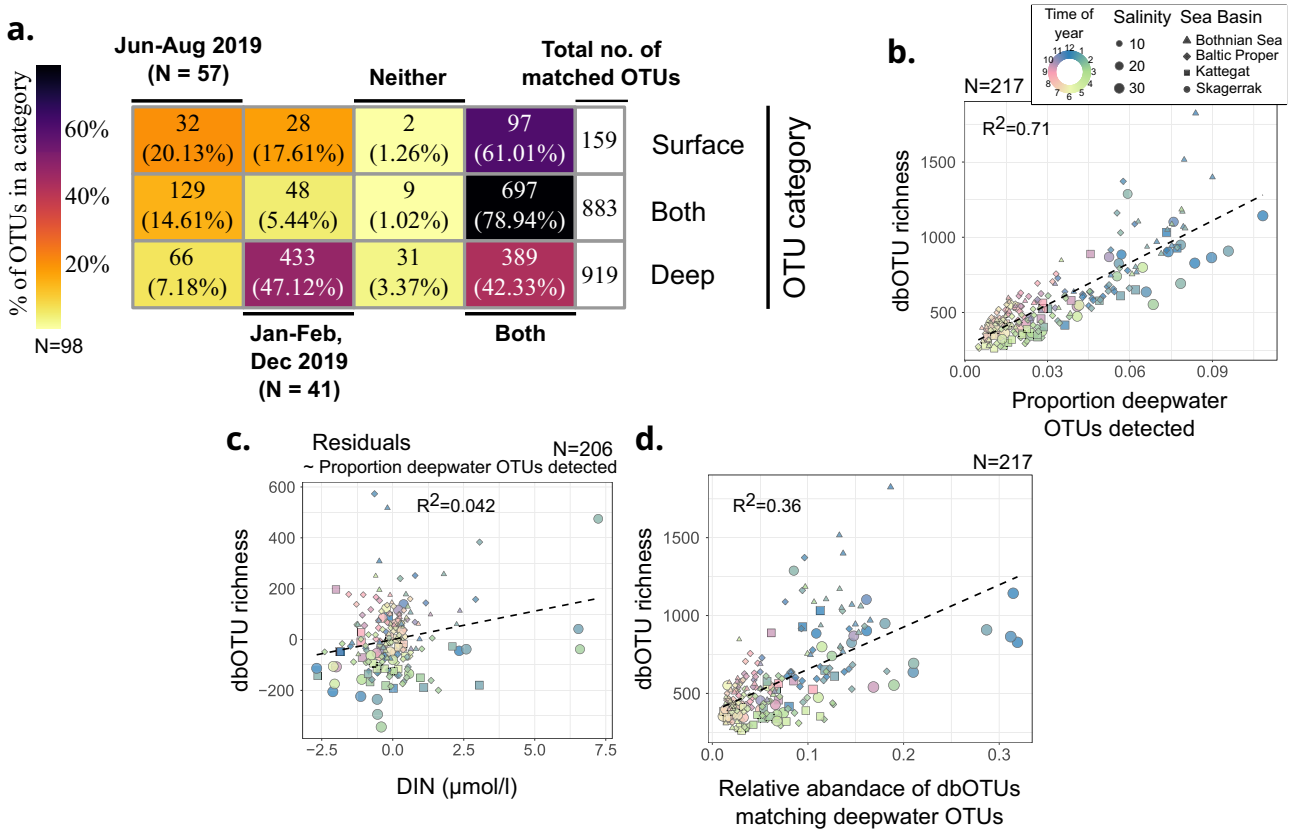


Fig. 5 | Bacterial alpha diversity and the proportion of deepwater OTUs detected. **a** Number of OTUs from a previous, transect-based study[17] matched to dbOTUs found in winter and summer samples or in neither or both seasons. The OTUs are categorized by being found exclusively in deep water (> 10 m depth) samples, exclusively in surface waters, or in both, in the original, transect-based study. Only OTUs matching any dbOTUs from this study are counted. *P*-values for the proportion of OTUs detected in summer being equal were 10^{-73} for deep, 10^{-8} for both, and 0.26 for surface. **b** Proportion of deepwater OTUs (including ones without a match in the 2019–2020 dataset) having a match to a dbOTU from a sample vs. bacterial dbOTU richness. **c** The residuals of DIN vs. dbOTU richness from models correlating the factors to the proportion of deepwater OTUs detected, corresponding to partitioning out the unvisualized parameter. **d** Relative abundance of all dbOTUs matching a deepwater OTU vs bacterial dbOTU richness. Bothnian Bay was excluded from the plots since there were no deepwater samples taken in this region in the transect-based study[17]. The number of samples (N) used to create each panel is given in each panel. The differences in the number of samples come from the availability of physicochemical parameters and data division strategies (**a**). The same set of samples from the transect-based study[17] (N = 106) was used for all the panels (samples not shown, see Methods for details). The same legend applies to (**b**)–(**d**). DIN dissolved inorganic nitrogen.

However, although we rarefied the samples to the same read depth, this simple analysis could be confounded by detection limits. Indeed, even on the preferred side of the salinity barrier, protist dbOTUs were on average found in a larger number of samples (Supplementary Fig. S10). We thus modeled the probability a dbOTU was found on both sides of the salinity barrier as a function of occupancy (i.e., the number of samples in which a dbOTU was found). This Bayesian model showed that protists were more likely to cross the barrier than bacteria across the occupancy values (Fig. 6f). For all the occupancy values, the probability of protists being more likely to cross the barrier was >99%.

Furthermore, among the dbOTUs that crossed the salinity barrier, protists tended to be present in more samples at the less-preferred side of the barrier (Fig. 6g). Thus, the barrier-crossing dbOTUs were more likely to grow on both sides of the barrier, as opposed to their genetic material being transported with currents.

Overall, protist dbOTUs had wider salinity niches than their bacterial counterparts, though this effect was not observed at the lower end of the brackish salinity gradient (Fig. 6). This suggests that the diversity patterns we observed present a general trend of different breadth of realized niches with respect to salinity across ecologically diverse bacteria and protists. Nevertheless, taxa or ecological groups that deviate from this trend could surely be found.

Among protists, but not bacteria, marine-associated lineages outnumber freshwater-associated ones across most of the salinity gradient

Similarly to our observations on protists, fish[30] and macroalgal[31] alpha diversity increases with salinity in the Baltic Sea area. The Baltic fauna and macroalgae stem mainly from a limited number of local adaptations to brackish conditions following the last glaciation period, with marine lineages outnumbering freshwater ones[31,41,48]. In contrast, for Baltic Sea bacteria, there is evidence against widespread occurrence of such local adaptation. Instead, the niches appear to be filled through colonization by a global brackish microbiota, consisting of species that are distinct from but often related to freshwater and marine lineages[8]. Consistently, in the current (Fig. 4a, c) and previous[14,17,18] studies, bacterial alpha diversity remained relatively stable across the brackish salinity gradient. A metabarcoding-based analysis by Herlemann et al. showed that in the Baltic Sea, freshwater-associated bacterial taxa are gradually replaced by marine-associated ones, keeping alpha diversity constant[14].

To test whether the difference in alpha diversity patterns between bacteria and protists can be explained by the distribution of marine- and freshwater-associated lineages, we matched our dbOTUs to surface water (< 10 m depth) freshwater and marine (> 34 PSU) ASVs from the 16S rRNA gene database MiCoDa[68] and the 18S rRNA gene database MetaPR^2[69]. The 16S rRNA gene ASVs in MiCoDa are 90bp-long and we

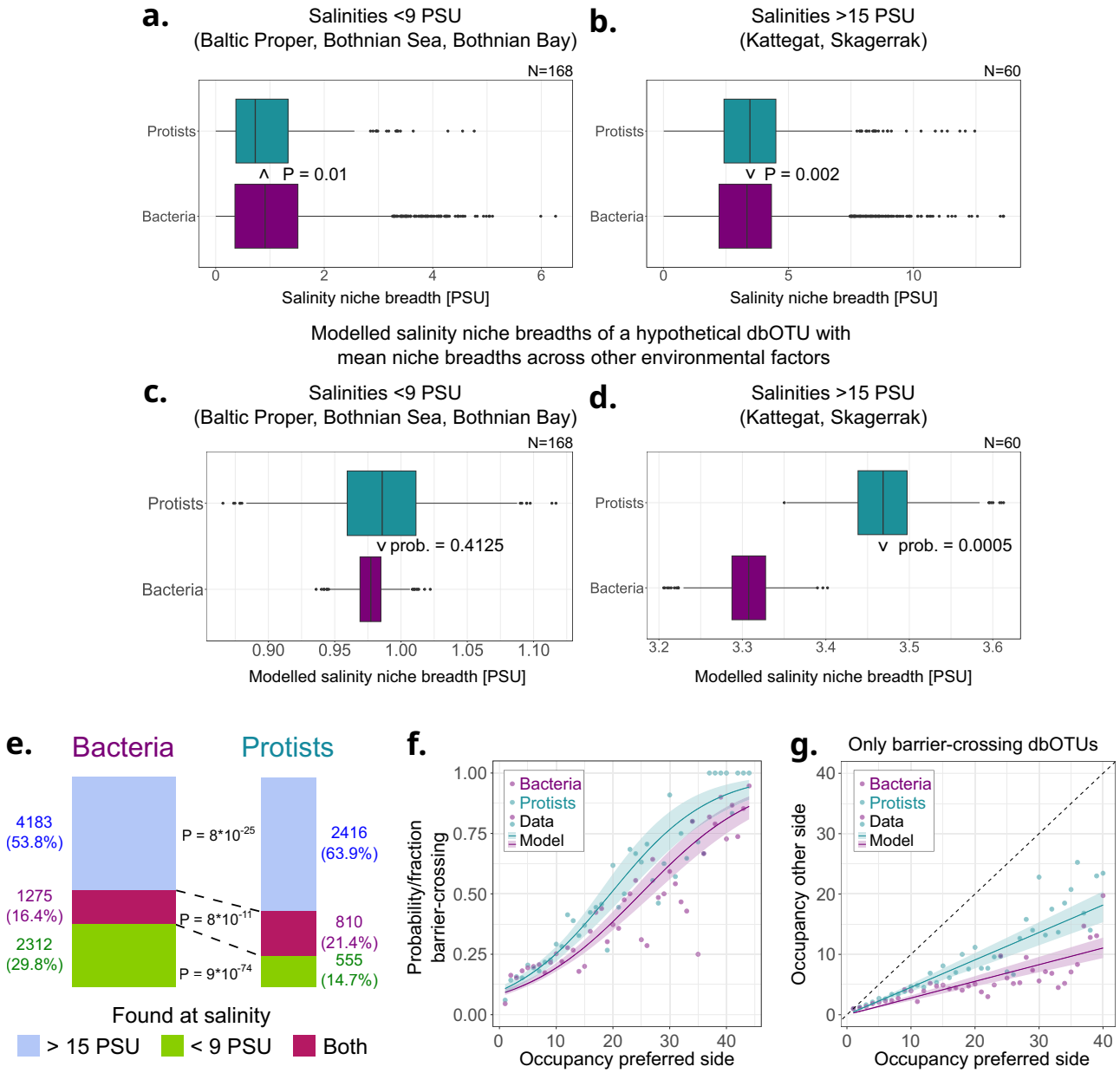


Fig. 6 | Salinity niche breadths and proportions of salinity barrier-crossing bacteria and protists. a, **b** The estimated salinity niche breadths of dbOTUs found only in salinities below 9 PSU (**a**) or above 15 PSU (**b**). **c**, **d** Results of Bayesian modeling. Predicted salinity niches for a hypothetical dbOTU with average temperature, DIN, phosphate, and day length niche breadths, depending on whether it would be a bacterium or a protist. The modeling has been done using dbOTUs found only in salinities below 9 PSU (**c**) or above 15 PSU (**d**). In (**a**)–(**d**), N represents the number of samples with salinities below 9 PSU or above 15 PSU. The sign between boxplots indicates which group had broader salinity niches, and *P*-values (**a**, **b**) or probabilities of obtaining an opposite trend in the Bayesian model (**c**, **d**) are given. The lines on the boxplot represent (from the middle) median, 75th percentile, and the whiskers extend to 1.5 times the interquartile range. The significance values in (**a**, **b**) come from two-sided Wilcoxon rank-sum test and the probabilites in (**c**, **d**) from the results of the Bayesian modeling (proportion of MCMC iterations yeilding a difference with a different direction to the difference between the medians). **e** The number and proportion of bacterial and protist dbOTUs found in high (> 15 PSU) and low (< 9 PSU) brackish salinities, and across the salinity barrier. The significance values comes from two-sided Fisher's exact test. **f**, **g** Results from Bayesian modeling. **f** The proportion of salinity barrier-crossing dbOTUs (found on both sides of the Danish Straits; points) and the modeled probabilities of a dbOTU with a given occupancy crossing the salinity barrier. **g** The occupancy on the less occupied side of the salinity barrier as a function of the occupancy on the more occupied side. In (**f**) and (**g**), the points correspond to the means of the observed values. The lines correspond to the values predicted from the model with 95% confidence intervals in shaded colors. In (**g**), the dashed black line is the identity line. For (**e**)–(**g**), only the two stations with the highest salinities (21-34 PSU), and two stations with salinities around 6-7 PSU were included. The same number of samples (n = 22) were included for each station, as close in time as possible between stations (see Methods for details). "Occupancy preferred side" and "occupancy other side" refer to the occupancy at the side of the salinity barrier with higher and lower occupancy, respectively.

trimmed the MetaPR^2 ASVs to the same length for comparability (see Methods for details).

Across the salinity gradient, bacterial dbOTUs gradually transitioned from those matching freshwater database ASVs to those with marine database matches (Fig. 7a), consistent with the Herlemann study[14]. There were more freshwater- or marine-associated dbOTUs, respectively, on the lower and higher salinity sides of the Danish Straits. In contrast, for protists, while the number of marine-associated dbOTUs increases steadily with salinity, freshwater-associated dbOTUs outnumber them only in near-freshwater Bothnian Bay (Fig. 7b). These

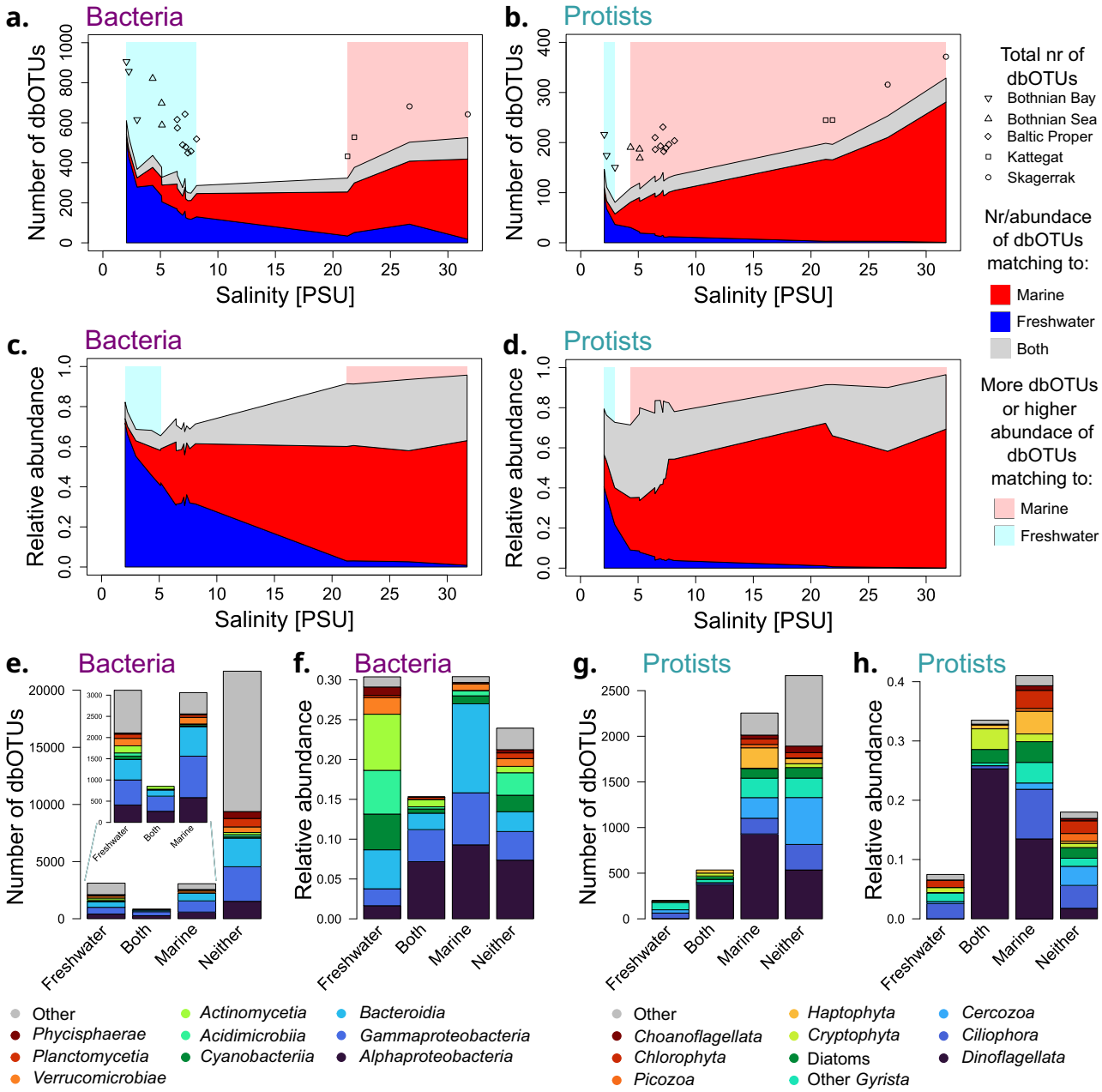


Fig. 7 | dbOTUs matching database marine and/or freshwater sequences. a, **b** The mean number of bacterial (**a**) and protist (**b**) dbOTUs matching marine or freshwater database sequences, or both. Mean total numbers of dbOTUs per station, based on the same rarefied count table, are labeled with sea-basin-coded shapes. The background color indicates salinity ranges at which all the stations had more dbOTUs matching freshwater or marine database sequences. **c**, **d** Relative abundance of bacterial (**c**) and protist (**d**) dbOTUs matching marine or freshwater database sequences, or both. Common legend for (**a**)–(**d**) is on the right-hand side of the panels. **e**-**f**. The number (**e**, **g**) and mean (across all samples) relative abundance (**f**, **h**) of bacterial (**e**, **f**) and protist (**g**, **h**) dbOTUs matching marine or freshwater database sequences, or both, or lacking a database match ("Neither"). The dbOTUs have been divided into bacterial classes and protist subdivisions. Only the classes/subdivisions which, on average, corresponded to >0.01 of reads across samples are shown. *Gyrista* has been divided into diatoms and other *Gyrista*. The legend for (**e**-**h**) are given in the bottom-most part of the figure. The marine and freshwater sequences were obtained from MiCoDa[68] for bacteria and metaPR^2[69] for protists. Note: The sequences of ASVs comprising dbOTUs (259-537 bp for bacteria, 265-544 bp for protists) were compared to 90 bp database sequences. This was the length of available sequences in the 16S rRNA gene fragment database[68], and we trimmed the 18S rRNA gene database sequences to the same length for comparability.

patterns held when the relative abundance, instead of the number of dbOTUs, was plotted (Fig. 7c, d).

Despite balancing the number and distribution of freshwater and marine reference samples (see Methods for details), we obtained more than three times as many marine (6867) as freshwater (1907) database protist ASVs. While a higher diversity of marine than freshwater protists has been reported, the true ratios are unlikely to be that high[70]. We thus downsampled the marine protist reference sequences to match the number of freshwater ASVs and repeated the analysis. However, the obtained pattern was congruent with the imbalanced analysis (Supplementary Fig. S11b, d). The number of freshwater and marine database ASVs for bacteria was very similar (12973 vs 11532), and downsampling to the same number of ASVs yielded near-identical results as the original analysis (Supplementary Fig. S11a, c).

The number and (mean across samples) relative abundance of bacterial freshwater- and marine-associated dbOTUs were nearly identical (Fig. 7e,

f.), while marine-associated protists were much more numerous and abundant than freshwater-associated ones (Fig. 7g, h.). The proportion of freshwater- and marine-associated dbOTUs also differed between broad taxonomic groups of bacteria and protists. For example, only four dinoflagellate dbOTUs were freshwater-associated, compared to almost 1000 marine-associated ones (Fig. 7f–h). Consistently, most described *Dinoflagellata* species are marine[71,72]. The only protist subdivision with more freshwater- than marine-associated dbOTUs (12 vs 5) was *Cryptophyta*. Similarly, phylogenetic analyses have shown that cryptomonads more often transitioned from freshwater to marine environments than the other way round[73]. Meanwhile, freshwater- or marine-associated dbOTUs dominated different bacterial classes (Fig. 7e, f), in accordance with their spatial distribution within the Baltic Sea area (Fig. 1b, Supplementary Fig. S2a).

A substantial number of dbOTUs were associated with both freshwater and marine database ASVs, indicating overlap between these two biomes. We attribute this overlap to the short sequence length of the database fragments used (90 bp). For protists, where full 18S rRNA gene V4 fragment sequences were available in the database, we re-ran the analysis on the ASVs with lengths similar to those used in this study. While the pattern of freshwater-to-marine-associated dbOTUs remained the same, nearly none of the dbOTUs matched to both biomes (Supplementary Fig. S11e-f.). The freshwater-marine divide is even better documented for bacteria than for protists[1–7]. Thus, using longer 16S rRNA gene fragments would very likely also eliminate the matching to both freshwater and marine ASVs for bacteria.

For protists, the results presented here resemble the well-studied fish[30] and macroalgal[31] patterns[48]. Thus, they are highly suggestive of an underlying mechanism of local adaptation[31,41,48]. This would stand in stark contrast to the colonization of brackish-specialist bacteria after the Baltic Sea was formed[8]. However, the taxonomic resolution of metabarcoding hinders conclusions about local adaptation of protists. Follow-up studies using genomic analyses can test this hypothesis.

Conclusions

We show that salinity has a stronger impact on bacterial than protist beta diversity, but that the reverse is true for alpha diversity. Alpha diversity (richness) of bacteria is maintained at similar levels across the brackish salinity gradient through a steady replacement of freshwater-associated lineages with marine-associated ones as salinity increases. Meanwhile, we show that marine-associated protists dominate in the Baltic Sea. As fewer and fewer of them can be found with decreasing salinity, alpha diversity drops. We also show that protist taxonomic units tend to have wider salinity niches, corroborating the beta diversity pattern.

These patterns are consistent with bacterial niches being dominated by brackish-water specialists that have dispersed from other brackish regions[8,12], whereas the protists primarily represent locally adapted marine populations, analogous to the prevailing situation in Baltic Sea fish and macroalgae[31,41,48]. Colonization of the Baltic Sea by brackish specialist bacteria most likely required dispersal across hundreds or thousands of kilometers. Evidence from global oceans suggests that bacteria disperse more easily than protists across such long distances[74]. Similarly, multiple studies have shown more limited dispersal of aquatic protists than bacteria across shorter distances[75–77]. Given this, our findings that bacterial communities change more than protist ones across the Danish Straits (Fig. 2c, d), and that protist taxonomic units are more prone to occupy both sides of the straits (Fig. 6f, g), may appear counterintuitive. However, it is consistent with bacteria being less limited in long-range dispersal, allowing brackish specialists to colonize newly formed brackish basins and thus preventing local adaptation of populations from the neighboring waters from filling the niches. The idea of brackish-specialist bacteria colonizing newly formed brackish basins is supported by genomic analyses[8,12]. However, environment-coupled genomic data for brackish protists is too scarce[41] to reach conclusions on the prevalence of local adaptation vs. long-distance recruitment. Thus, future environment-coupled genomic studies of diverse protists will challenge our hypothesis on differences in adaptation underlying the diversity patterns observed here.

Metabarcoding and metagenomics have shifted recent advances in understanding of aquatic microbial communities to focus much more on bacteria than on protists[78]. Our results strengthen the notion that conceptual frameworks developed for bacteria do not always translate into protists[78]. Salinity has been repeatedly described as the strongest factor affecting bacterial diversity, surpassed only by the separation of free-living and host-associated bacteria[1–4]. We show that this is unlikely to translate into protist diversity patterns, at the very least not within the brackish salinity spectrum. On the other hand, studies trying to assess whether bacteria or protist communities change more with the environment can overlook that the impact depends on the nature of the environmental change[16,26–28]. Here, we show that salinity has a stronger impact on bacterial than protist beta diversity, but the opposite is true for other factors. Ultimately, these patterns not only differ between bacteria and protists, but also within the groups. Therefore, we highlighted taxa, from families to subdivisions or classes, which most strongly contribute to the observed diversity patterns.

The spatiotemporal resolution of the study is a step forward a synthesis of knowledge from transect- and time-series-based diversity patterns. For example, we show that major bacterial and protist clades change not only in abundance but also in their seasonal dynamics across the Baltic Sea area. We also show that accounting for seasonal changes in bacterial alpha diversity is necessary to properly interpret the pattern across the salinity gradient. Alpha diversity increased in winter due to convective mixing, as observed before[17,36,38,66,67]. However, time-series-based studies consistently correlated the seasonal patterns of alpha diversity to day length[36,38,66,67], and thus the importance of irradiation and other cryptic effects on the pattern is still being debated[38]. In our results, this correlation was poor, suggesting that vertical mixing dominates over the effects of irradiation on the common seasonal pattern of bacterial alpha diversity. Adding an additional temporal scale, we observed interannual spatiotemporal recurrence of community composition following the environmental gradients, consistent with previous results[28,53,63]. However, technical differences between the datasets confounded alpha diversity analysis. Thus, a standardized DNA-based scheme is needed to monitor rare microbial taxa, and our results can inform its development.

Using DNA metabarcoding across strong spatiotemporal gradients, we uncover patterns of microbial diversity and expand our understanding of previously observed ones. Ultimately, as Earth's environments undergo rapid change, diverse microorganisms, often understudied groups, will be impacted and/or will impact the trajectory of brackish and other ecosystems. DNA-based monitoring covers the phylogenetic scope needed to understand those changes and respond to them. Understanding the ecoevolutionary processes behind the diversity patterns will be critical for the correct interpretation of their changes.

Methods

Data collection and sequencing

The 2019-2020 data come from Latz et al. 2024[49]. The collection and sequencing procedures are described in detail therein. In brief, samples were collected monthly or every second week, depending on the location, as part of the Swedish National Marine Monitoring Program. The samples for metabarcoding were collected using a depth-integrating hose, covering 0-10 m (except 0-5 m at RÅNEÅ-1, and 0-20 m at B1 and BY31). 500 ml of water from the hose sampling was filtered using 0.22 µm pore size. DNA was extracted using the Zymobiomics™ DNA miniprep kit following a modified protocol, and spike-in DNA was added[79]. The DNA was sequenced using Illumina™ MiSeq flow cells. For 16S rDNA gene-based metabarcoding of prokaryotes, primers targeting the hypervariable V3-V4 regions[14], and for 18S metabarcoding of eukaryotes, primers targeting the V4 region[80], were used. Phased primers[49,81] were used for both 18S (on the forward primer) and 16S (on both primers) metabarcoding. Sequences of the primers and the details of the phasing strategy and sample indexing through a second PCR can be found in Latz et al. 2024[49]. Physico-chemical and geographic data

have been obtained from the Swedish Meteorology and Hydrology Institute's service SHARK: https://shark.smhi.se/en/ (some of the data used here are currently (at the time of publication) not available in SHARK, but can be found in the figshare repository for this study).

The 2015-2017 data were used for the first time for this study. The samples were stored at -80°C. DNA was extracted in the summer of 2017 and stored at -20°C until February 2023. The sampling and sequencing procedures were analogous to those for 2019-2020 data, with a few exceptions:

1. Volumes varied from 200 ml to 800 ml. According to our previous tests run in the same setting, volumes ≥ 200 ml should capture the microbial community structure equally well[49].
2. DNA extraction was performed using Qiagen™ DNeasy PowerWater Kit, and not ZymoBIOMICS™ DNA Miniprep Kit with a small modification[79] as for the 2019-2020 dataset.
3. No spike-in DNA was added.
4. Illumina™ NextSeq, rather than MiSeq system, was used.
5. For 18S, the initial use of 20 cycles (as in Latz et al.[49]) for the first PCR amplification was insufficient, leading to many library preparation failures. It was then re-run with 24 cycles, and the results of both runs were merged after processing the tables by summing up the counts in the output count table.
6. For 18S, reverse primers were also phased, following a sequence (CTACGA)CTTTCGTTCTTGATYRR, with versions of the primer starting from any base in the brackets included in the primer mix.
7. After library preparation, for most samples, 18S and 16S libraries were mixed and sequenced together. A cutadapt[82]-based pipeline was used to separate reads with only 16S or only 18S samples for downstream analysis, as described in the following section.

Samples from other projects, which we did not analyze in this study, were sequenced simultaneously and submitted to the European Nucleotide Archive (ENA) all together.

Processing of sequencing data

The data processing for 2019-2020 samples has been described in detail in Latz et al. 2024[49]. In brief, the phased primer sequences were removed from the obtained reads using a modified version of a cutadapt[82]-based pipeline (original pipeline taken from: https://github.com/insect-biome-atlas/amplicon-multi-cutadapt). Only reads containing 16S or 18S primers were kept, which allowed us to separate reads from the two methods when the libraries were mixed. DADA2[83] version 1.18.0 was used to denoise the data, infer amplicon sequence variants (ASVs), and taxonomically annotate them. For taxonomic annotation of 16S ASVs, we used 16S sequences from GTDB[84] version R06-RS202-1 corrected for mislabeled sequences using SATIVA[85,86]. Annotation was additionally conducted using the SILVA database[87] version 138.1 to identify plastids and mitochondria. The 18S ASVs were annotated using the PR2 database[88] version 5.0.0.

The 2015-2017 data was analyzed the same way, with the only difference being that the denoising and ASV inference with DADA2 was performed using the nf-core/ampliseq pipeline[89] version 2.7.0 instead of an in-house R script. We kept the DADA2 parameters as in Latz et al. 2024[49].

We further combined the 2015-2017 and 2019-2020 data. We matched ASVs using exact matching. On the merged ASV-sequence list and count table, we ran an additional chimera removal step using UCHIME v1[90] and clustered the remaining ASVs based on their simultaneous genetic and distribution similarity using dbOTU3[50]. Both these steps were performed using an in-house pipeline available at https://github.com/kjurdzinski/chimera_dbOTU_pipeline. Default parameters were used for UCHIME, and for dbOTU3, only the maximum genetic distance for 16S has been changed to 3% (instead of the default 10%). dbOTU aims to identify multiple ASVs stemming from the same population, i.e., multiple rRNA gene copies with some nucleotide differences encoded in the same genome, and keep only one of them. This procedure is particularly important for alpha diversity analyses, in which multiple distinct rRNA gene copies could inflate the observed richness. We called the resulting sequence clusters dbOTUs (distance-based operational taxonomic units, as inferred using dbOTU3[50]). This clustered the 122,371 and 49,600 ASVs for bacteria and protists into 48,961 and 9,909 dbOTUs,

For both 16S and 18S, we removed dbOTUs with no annotation at the phylum/supergroup taxonomic level. From 16S dbOTUs, *Archaea*, plastids, and mitochondria were removed. Thus, the remaining dbOTUs should correspond to bacteria. From 18S dbOTUs, animals (*Metazoa*), *Fungi*, land plants (*Embryophyceae*), and selected macroalgae (*Phaeophyceae* and Rhodophyta excluding *Rhodella* and *Cyanidiales*) were removed. Thus, the remaining dbOTUs were primarily protists.

We kept spike-in dbOTUs in a separate table and used them only to obtain spike-in corrected abundance, i.e., the inferred number of rDNA copies per mililiter. The rDNA copies per liter were calculated according to the equation:

$$rDNA\ copies\ per\ milliliter = \frac{dbOTU\ counts}{spike-in\ counts} \times \frac{amount\ of\ spike-in\ added\ [g]}{volume\ water\ sampled\ [ml] \times spike-in\ molar\ mass\ [g/mol]} \times 6.022 \times 10^{23}$$

We have excluded over- and under-sequenced samples to remove the amplification errors (including distorted relative abundances) and not to rarefy to a too low read number during rarefaction. We have excluded three samples with fewer than 30,000 reads for bacteria, as well as one sample with more than 2,000,000 reads. Moreover, we have excluded two samples from 2019-2020 with more than 200,000 protist reads (this number of reads was normal for 2015-2017 data since Illumina™ NextSeq was used instead of MiSeq, yielding overall higher numbers of reads). We have chosen the thresholds based on visual examination of the distribution of the sequencing depths, removing outliers. Additionally, one sample (20191015_263953_1) with suspected contamination has been excluded.

For comparability of results for protists and bacteria, we used only the samples with data available from both 16S and 18S amplicon sequencing (excluding the under- and over-sequenced samples) for all further analyses. Additionally, in two samples (station RÅNEÅ-2 on 2019-01-29 (sample ID: 20190129_263856_1) and B7 on 2019-09-17 (sample ID: 20190917_264450_1)) 18S spike-in addition failed. Those samples were excluded from analyses in which spike-in normalization was used, except for bacteria-only focused analyses in Fig. 1b.

Abundance, alpha and beta diversity spatiotemporal patterns

For the spatiotemporal analysis of abundance, beta- and alpha diversity across the samples from 2019-2020, we removed the samples from stations H4, BY29 / LL 19, and SR3. Those stations were sampled at only a few unevenly distributed occasions, and their exclusion/inclusion at certain seasons could obscure seasonal patterns.

All the statistical analyses and data visualization were performed using R version 4.2.2[91]. Inkscape[92] was used to assemble and edit multi-panel figures.

The multivariate analyses were performed using vegan R package[93], and all the functions in this paragraph can be found there unless another package is explicitly stated. Bray–Curtis distances between communities were obtained using rarefaction via the avgdist function. Those distances were used for Principal Coordinate Analysis (PCoA) and distance-based Redundancy Analysis (dbRDA), performed with, respectively, capscale (without setting constraints) and dbrda functions. We chose dbRDA to base the constrained analysis on Bray-Curtis distances, which are a more robust metric of ecological dissimilarity than Euclidean distances used in redundancy analysis (RDA) or chi-square distances used in Canonical Correspondence Analysis (CCA)[94]. Variance partitioning was performed using rdacca.hp R package[95]. Partial dbRDA was performed by using the special term Condition() in the dbRDA function's formula. Families were positioned on the ordination plots by calculating weighted averages (sppscores function). Only the families deviating (Bonferroni-corrected P-value < 0.05)

from normal distribution around the center of the dbRDA coordinate system were shown on the plots. To find the corresponding two-dimensional ellipsoid confidence intervals, aspace[96] package was used. Whenever only samples were visualized, their ordination scores were scaled using the eigenvalues, and symmetric scaling was used if both the families and the samples were visualized (using, respectively, "sites" and "symmetric" scaling options of the plotting functions). For dbRDA and variance partitioning, we used a subset of 228 (out of 246) samples with all the environmental variables of interest available. For more informative visualization, to keep the sample structuring visible despite very high family scores, we divided the family scores in half. We chose day length as one of the factors, as it has been connected with seasonal changes in bacterial alpha diversity[36,38] and, together with temperature, should catch the full scope of seasonality (as temperature increase is lagging behind day length).

dbOTU richness was calculated as the mean number of dbOTUs detected in a sample over 100 rarefaction iterations. Variance partitioning was performed using rdacca.hp R package[95].

For both beta and alpha diversity estimation, in each rarefaction iteration, we rarefied the bacterial 16S reads to 34493 per sample, and the protists 18S reads to 8009 per sample.

For deepwater taxa detection, we used the representative sequences of 99% average nucleotide identity OTUs (operational taxonomic units) detected in samples from the summer of 2008 and coming from a previous study[17]. The 2008 summer transect included sampling across different depths, and the deepwater OTUs in this study were present exclusively at depths ranging from 11 to 300 m, classified as mesopelagic in the original study[17]. We connected the OTUs from the original, transect-based study to ASVs from the monitoring data used here, using the 99% nucleotide identity threshold in VSEARCH[97]. Despite substantial methodological differences, we found a match for 36.6% of the OTUs and 16.1% of dbOTUs at a 99% nucleotide identity threshold. We deemed an OTU and a dbOTU to be matching if their representative sequences (most abundant ASV for a dbOTU) matched in the VSEARCH output. For comparing the number of OTUs from deepwater, surface water, and both (Fig. 5a) in winter and summer, we used the chi-squared test as implemented in the prop.test function in R[98]. We excluded Bothnian Bay from the analyses since there were no deepwater samples from this region in the transect-based dataset.

Niche breadths

We have estimated the niche breadths with the nichePOSNB function of the ecospat package[99]. This function calculates niche breadth as the standard deviation of a parameter's values weighted by the relative abundance of a taxon.

We used the estimated niche breadths to build a Bayesian model in JAGS[100], using the R package rjags[101]. The parameters, probabilities, and predictions were calculated based on every fifth out of 10000 MCMC iterations. We modeled salinity niche breadth based on the normal/Gaussian distribution:

$$salinity_niche_breadth_i \sim Normal(\mu_i, \sigma)$$

where $salinity_niche_breadth_i$ is the estimated salinity niche breadth for a dbOTU, σ is the standard deviation of the distribution from which these values are modeled to be sampled, while μ_i is the mean of the distribution. μ_i was calculated based on the equation:

$$\begin{aligned}\mu_i = {} & (1 - if_protist_i) \times (\alpha_{bacterium} + \beta_{temperature_bacterium} \times temperature_niche_breadth_i + \\ & \beta_{DIN_bacterium} \times DIN_niche_breadth_i + \beta_{phosphate_bacterium} \\ & \times phosphate_niche_breadth_i \beta_{day_length_bacterium} \times day_length_niche_breadth_i) + if_protist_i \\ & \times (\alpha_{protist} + \beta_{temperature_protist} \times temperature_niche_breadth_i + \beta_{DIN_protist} \\ & \times DIN_niche_breadth_i + \beta_{day_length_protist} \times day_length_niche_breadth_i)\end{aligned}$$

The prior for each parameter was set to 0, corresponding to a lack of knowledge of any effect on the dependent variable.

Proportion salinity barrier crossing dbOTUs

We analyzed which dbOTUs were found in at least one sample from near-marine stations (Å17, SLÄGGÖ) and from comparable stations in the Baltic Sea (B1 and BY31 LANDSORTSDJ). These samples were far away from the Danish Straits, imposing the salinity barrier, reducing the risk of false detection of dbOTUs whose DNA or dead/dormant cells were carried with the water flow or aerosols across the straits. The two stations on each side of the barrier comprised one near-coast and one offshore station. For each station but the least sampled (Å17 and B1), we chose a subset of samples (n = 22), keeping samples closest in time to the ones collected at the least sampled stations. Whenever available, those samples should correspond to the same monitoring cruises.

To reduce detection biases, 16S and 18S rRNA gene reads were downsampled to the same level (10,863) across the samples.

The Bayesian models were built and implemented in JAGS[100], using the R package rjags[101]. The parameters, probabilities, and predictions were calculated based on every fifth out of 10000 MCMC iterations.

The model for the probability of crossing the salinity barrier was optimized based on the Bernoulli distribution:

$$if_crossed_barrier_i \sim Bernoulli(p_i)$$

where for each dbOTU i, $if_crossed_barrier_i$ is a binary variable = 1 if it is a barrier-crossing dbOTU. p_i was calculated according to the formula

$$\begin{aligned}logit(p_i) = {} & (1 - if_protist_i) \\ & \times (\alpha_{bacterium} + \beta_{occurence_bacterium} \times \max_occurence_one_side_i) \\ & + if_protist_i \times (\alpha_{protist} + \beta_{occurence_protist} \times \max_occurence_one_side_i)\end{aligned}$$

where $if_protist_i$ is a binary variable = 1 if the dbOTU i is a protist. $\max_occurence_one_side_i$ is the maximum out of the numbers of samples in which the dbOTU i was found at one side of the salinity barrier (in salinity <9 PSU or >15 PSU). The other variables are the parameters of the model.

The model for total occupancy was built based on the Poisson distribution

$$total_occupancy_i \sim Poisson(\lambda_i)$$

where $total_occupancy_i$ is the total number of samples in which dbOTU i was found. λ_i was calculated according to the formula:

$$\begin{aligned}\lambda_i = {} & \max_occurence_one_side_i + (1 - if_protist_i) \\ & \times (\alpha_{bacterium} + \beta_{bacterium} \times \log(\max_occurence_one_side_i) \\ & + if_protist_i \times (\alpha_{protist} + \beta_{protist} \times \max_occurence_one_side_i)\end{aligned}$$

with $if_protist_i$ and $\max_occurence_one_side_i$ having the same meaning as described above. The occupancy at the less occupied side of the salinity barrier was calculated by subtracting $\max_occurence_one_side$ from the predicted λ.

The prior for each parameter was set to 0, corresponding to a lack of knowledge of any effect on the dependent variable.

Matching to freshwater and marine ASVs

For bacteria, we downloaded 16S rRNA gene V4 fragment sequences from MiCoDa v2 falling into "saltwater" and "lake" categories ("Environment - broad scale"). We then used entrez_search() and entrez_summary() functions from the rentrez R package to obtain metadata for the samples. We selected samples from less than 10 m depth, and more than 34 PSU salinity for saltwater samples, and "Freshwater" as "envrionment_feature" for the lake samples. We thus identified 182 marine and 918 freshwater samples, and downsampled the freshwater samples to match the number of marine samples. Altogether, we got 11532 marine ASVs and 12973 freshwater ASVs, with 167 ASVs found both in freshwater and marine samples.

For protists, we downloaded 18S ASVs from the metaPR^2 database version 3[69] on October 20th 2025. We have chosen the "oceanic" (hereafter referred to as marine) and "freshwater" ecosystems ("freshwater lake" and "freshwater river"), all size fractions, "surface" depth level, and only ASVs with a minimum of 100 reads in the database (default option). Since all the downloaded freshwater samples came from the temperate climate, we limited the marine samples to the temperate climate as well. Furthermore, we limited the marine samples to the ones with recorded salinity above 34 PSU. We were then left with 238 marine and 208 freshwater samples. To equalize the number of samples per biome, we randomly chose 208 out of the 238 marine samples. Out of the downloaded ASVs, we chose only the protist ones, as described in the "Processing of sequencing data" section. We then rarefied the count tables to the minimum sampling depth (1161 reads) and kept only the ASVs with reads remaining after rarefying. We also removed the one ASV found in both freshwater and marine samples from the downstream analyses. Overall, we obtained 5568 marine and 1913 freshwater ASVs. We kept both all the marine ASVs and a list with as many marine ASVs as available for the freshwater biome, chosen randomly, with the probability of an ASV being chosen weighted by its average abundance in the database sample. We then trimmed the ASVs to 90 bp length to be comparable with matching to MiCoDa for bacteria.

We have connected the ASVs from our study to the database ASVs, obtaining perfect matches (100% nucleotide identity threshold) with VSEARCH[97]. We considered a dbOTU to be marine or freshwater if the representative ASV had a perfect match in the reference database.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The 2019-2020 data has been published before as a data paper[49], and is available in the European Nucleotide Archive under the study accession code PRJEB55296. The 2015-2017 raw sequencing data is available in the European Nucleotide Archive under the study accession code https://identifiers.org/ena.embl:PRJEB84926. The ASVs identified in the 2015-2017 sequencing data, together with a count table and contextual data, can be accessed through the ASV portal at the Swedish Biodiversity Data Infrastructure (SBDI). Subsequently, the ASVs, count tables, and contextual information are available at the Global Biodiversity Information Facility (GBIF), under accessions https://www.gbif.org/dataset/fbf49d0a-df4c-4b89-ba71-143bf6430a91 for 16S rRNA gene and https://www.gbif.org/dataset/9a228813-6288-4aba-a8f8-480e425b5f62 for 18S rRNA gene.

Code availability

The code used to analyze the data and generate the figures, together with intermediate processing output, can be found at https://doi.org/10.17044/scilifelab.28673273.

Received: 29 April 2026; Accepted: 7 September 2026;
Published online: 07 October 2026

References

1. Wu, Q. L., Zwart, G., Schauer, M., Kamst-van Agterveld, M. P. & Hahn, M. W. Bacterioplankton community composition along a salinity gradient of sixteen high-mountain lakes located on the Tibetan Plateau, China. *Appl. Environ. Microbiol.* **72**, 5478–5485 (2006).
2. Lozupone, C. A. & Knight, R. Global patterns in bacterial diversity. *Proc. Natl. Acad. Sci. Usa* **104**, 11436–11440 (2007).
3. Caporaso, J. G. et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. Usa.* **108**, 4516–4522 (2011).
4. Thompson, L. R. et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).
5. Logares, R. et al. Infrequent marine-freshwater transitions in the microbial world. *Trends Microbiol*. **17**, 414–422 (2009).
6. Walsh, D. A., Lafontaine, J. & Grossart, H.-P. On the Eco-Evolutionary Relationships of Fresh and Salt Water Bacteria and the Role of Gene Transfer in Their Adaptation. in *Lateral Gene Transfer in Evolution* (ed. Gophna, U.) 55–77 (Springer New York, New York, NY, 2013).
7. Burki, F., Sandin, M. M. & Jamy, M. Diversity and ecology of protists revealed by metabarcoding. *Curr. Biol.* **31**, R1267–R1280 (2021).
8. Jurdzinski, K. T. et al. Large-scale phylogenomics of aquatic bacteria reveal molecular mechanisms for adaptation to salinity. *Sci. Adv.* **9**, eadg2059 (2023).
9. Snoeijs-Leijonmalm, P. & Andrén, E. Why is the Baltic Sea so special to live in? in *Biological Oceanography of the Baltic Sea* (eds. Snoeijs-Leijonmalm, P., Schubert, H. & Radziejewska, T.) 23–84 (Springer Netherlands, Dordrecht, 2017).
10. *Water in Crisis*. (Oxford University Press, New York, NY, 1993).
11. Crump, B. C. & Bowen, J. L. The microbial ecology of estuarine ecosystems. *Ann. Rev. Mar. Sci.* **16**, 335–360 (2024).
12. Hugerth, L. W. et al. Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biol.* **16**, 279 (2015).
13. Hewson, I. & Fuhrman, J. A. Richness and diversity of bacterioplankton species along an estuarine gradient in Moreton Bay, Australia. *Appl. Environ. Microbiol.* **70**, 3425–3433 (2004).
14. Herlemann, D. P. et al. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* **5**, 1571–1579 (2011).
15. Fortunato, C. S. et al. Determining indicator taxa across spatial and seasonal gradients in the Columbia River coastal margin. *ISME J.* **7**, 1899–1911 (2013).
16. Torstensson, A. et al. Physicochemical control of bacterial and protist community composition and diversity in Antarctic sea ice: Physicochemical control of sea ice microbial communities. *Environ. Microbiol.* **17**, 3869–3881 (2015).
17. Herlemann, D. P. R., Lundin, D., Andersson, A. F., Labrenz, M. & Jürgens, K. Phylogenetic Signals of Salinity and Season in Bacterial Community Composition Across the Salinity Gradient of the Baltic Sea. *Front. Microbiol.* **7**, 1883 (2016).
18. Hu, Y. O. O., Karlson, B., Charvet, S. & Andersson, A. F. Diversity of Pico- to mesoplankton along the 2000 km salinity gradient of the Baltic Sea. *Front. Microbiol.* **7**, 679 (2016).
19. Tee, H. S., Waite, D., Lear, G. & Handley, K. M. Microbial river-to-sea continuum: gradients in benthic and planktonic diversity, osmoregulation and nutrient cycling. *Microbiome* **9**, 190 (2021).
20. Henson, M. W. & Thrash, J. C. Microbial ecology of northern Gulf of Mexico estuarine waters. *mSystems* **9**, e0131823 (2024).
21. Zhang, W., Zhang, Y., Shao, Z., Sun, Y. & Li, H. Differences in biogeographic patterns and mechanisms of assembly in estuarine bacterial and protist communities. *Microorganisms* **13**, (2025).
22. Hervé, V., Morelle, J., Lambourdière, J., Lopez, P. J. & Claquin, P. Together throughout the year: seasonal patterns of bacterial and eukaryotic microbial communities in a macrotidal estuary. *Environ. Microbiome* **20**, 8 (2025).
23. Bazin, P. et al. Phytoplankton diversity and community composition along the estuarine gradient of a temperate macrotidal ecosystem: combined morphological and molecular approaches. *PLoS One* **9**, e94110 (2014).
24. Piwosz, K., Całkiewicz, J., Gołębiewski, M. & Creer, S. Diversity and community composition of pico- and nanoplanktonic protists in the Vistula River estuary (Gulf of Gdańsk, Baltic Sea). *Estuar. Coast. Shelf Sci.* **207**, 242–249 (2018).
25. Li, R., Jiao, N., Warren, A. & Xu, D. Changes in community structure of active protistan assemblages from the lower Pearl River to coastal Waters of the South China Sea. *Eur. J. Protistol.* **63**, 72–82 (2018).

26. Logares, R. et al. Contrasting prevalence of selection and drift in the community structuring of bacteria and microbial eukaryotes. *Environ. Microbiol.* **20**, 2231–2240 (2018).
27. Wu, W. et al. Contrasting the relative importance of species sorting and dispersal limitation in shaping marine bacterial versus protist communities. *ISME J.* **12**, 485–494 (2018).
28. Yeh, Y.-C. & Fuhrman, J. A. Contrasting diversity patterns of prokaryotes and protists over time and depth at the San-Pedro Ocean Time series. *ISME Commun.* **2**, 36 (2022).
29. Remane, A. Die Brackwasserfauna : mit besonderer Berücksichtigung der Ostsee. *Zool. Anz*. **Suppl**. **7**, (1934).
30. Ojaveer, H. et al. Status of biodiversity in the Baltic Sea. *PLoS One* **5**, e12467 (2010).
31. Schubert, H., Feuerpfeil, P., Marquardt, R., Telesh, I. & Skarlato, S. Macroalgal diversity along the Baltic Sea salinity gradient challenges Remane's species-minimum concept. *Mar. Pollut. Bull.* **62**, 1948–1956 (2011).
32. Telesh, I. V., Schubert, H. & Skarlato, S. O. Revisiting Remane's concept: evidence for high plankton diversity and a protistan species maximum in the horohalinicum of the Baltic Sea. *Mar. Ecol. Prog. Ser.* **421**, 1–11 (2011).
33. Olli, K., Ptacnik, R., Klais, R. & Tamminen, T. Phytoplankton Species Richness along Coastal and Estuarine Salinity Continua. *Am. Nat.* **194**, E41–E51 (2019).
34. Olofsson, M., Hagan, J. G., Karlson, B. & Gamfeldt, L. Large seasonal and spatial variation in nano- and microphytoplankton diversity along a Baltic Sea-North Sea salinity gradient. *Sci. Rep.* **10**, 17666 (2020).
35. Lindh, M. V. et al. Disentangling seasonal bacterioplankton population dynamics by high-frequency sampling. *Environ. Microbiol.* **17**, 2459–2476 (2015).
36. Gilbert, J. A. et al. Defining seasonal marine microbial community dynamics. *ISME J.* **6**, 298–308 (2012).
37. Ladau, J. et al. Global marine bacterial diversity peaks at high latitudes in winter. *ISME J.* **7**, 1669–1677 (2013).
38. Raes, E. J. *et al*. Seasonal patterns of microbial diversity across the world oceans. *Limnol. Oceanogr. Lett* https://doi.org/10.1002/lol2.10422 (2024).
39. Kuliński, K. et al. Biogeochemical functioning of the Baltic Sea. *Earth Syst. Dyn.* **13**, 633–685 (2022).
40. Björck, S. A review of the history of the Baltic Sea, 13.0-8.0 ka BP. *Quat. Int.* **27**, 19–40 (1995).
41. Johannesson, K., Le Moan, A., Perini, S. & André, C. A Darwinian Laboratory of Multiple Contact Zones. *Trends Ecol. Evol.* **35**, 1021–1036 (2020).
42. Reusch, T. B. H. et al. The Baltic Sea as a time machine for the future coastal ocean. *Sci. Adv.* **4**, eaar8195 (2018).
43. Mazur-Marzec, H. et al. Biodiversity of microorganisms in the Baltic Sea: the power of novel methods in the identification of marine microbes. *FEMS Microbiol. Rev*. **48**, fuae024 (2024).
44. Heiskanen, A. S. et al Monitoring strategies for phytoplankton in the Baltic Sea coastal waters. European Commission. *Institute for Environment and Sustainability Inland and Marine Waters*. *Unit I-21020*. *Ispra (VA), Italy* (2005).
45. SHARK. SHARK - National marine environmental monitoring of Phytoplankton in Sweden since 1983. The Swedish Meteorological and Hydrological Institute https://doi.org/10.15468/S3YG8Z (2022).
46. Andersson, A. F., Riemann, L. & Bertilsson, S. Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J.* **4**, 171–181 (2010).
47. Dupont, C. L. et al. Functional tradeoffs underpin salinity-driven divergence in microbial community composition. *PLoS One* **9**, e89549 (2014).
48. Snoeijs-Leijonmalm, P. Patterns of biodiversity. In *Biological Oceanography of the Baltic Sea* (eds. Snoeijs-Leijonmalm, P., Schubert, H. & Radziejewska, T.) 123–191 (Springer Netherlands, Dordrecht, 2017).
49. Latz, M. A. C. et al. A comprehensive dataset on spatiotemporal variation of microbial plankton communities in the Baltic Sea. *Sci. Data* **11**, 18 (2024).
50. Olesen, S. W., Duvallet, C. & Alm, E. J. dbOTU3: A new implementation of distribution-based OTU calling. *PLoS One* **12**, e0176335 (2017).
51. Kembel, S. W., Wu, M., Eisen, J. A. & Green, J. L. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Comput. Biol.* **8**, e1002743 (2012).
52. Torstensson, A. et al. Comparing DNA metabarcoding with light microscopy to identify eukaryotic phytoplankton in the Baltic Sea, Kattegat and Skagerrak. *Sci. Rep*. **16**, (2026).
53. Fridolfsson, E. et al. Multiyear analysis uncovers coordinated seasonality in stocks and composition of the planktonic food web in the Baltic Sea proper. *Sci. Rep.* **13**, 11865 (2023).
54. Zettler, M. L., Kremp, A. & Dutz, J. Biological assessment of the Baltic Sea 2019. *Meereswiss. Ber*, *Warnemünde* **115**, (2020).
55. Wesslander, K., Viktorsson, L., Thor, P., Nilsson, M. & Skjevik, A.-T. *The Swedish National Marine Monitoring Programme 2019* : *Hydrography Nutrients Phytoplankton*. 59 (2020).
56. Logares, R. et al. Disentangling the mechanisms shaping the surface ocean microbiota. *Microbiome* **8**, 55 (2020).
57. Farjalla, V. F. et al. Ecological determinism increases with organism size. *Ecology* **93**, 1752–1759 (2012).
58. Neuenschwander, S. M., Ghai, R., Pernthaler, J. & Salcher, M. M. Microdiversification in genome-streamlined ubiquitous freshwater Actinobacteria. *ISME J.* **12**, 185–198 (2018).
59. Mehrshad, M. et al. Hidden in plain sight—highly abundant and diverse planktonic freshwater Chloroflexi. *Microbiome* **6**, 176 (2018).
60. Pujalte, M. J., Lucena, T., Ruvira, M. A., Arahal, D. R. & Macián, M. C. The Family Rhodobacteraceae. in *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria* (eds. Rosenberg, E., DeLong, E. F., Lory, S., Stackebrandt, E. & Thompson, F.) 439–512 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2014).
61. Morris, R. M. & Spietz, R. L. The Physiology and Biogeochemistry of SUP05. *Ann. Rev. Mar. Sci.* **14**, 261–275 (2022).
62. Spilling, K. et al. Shifting diatom—dinoflagellate dominance during spring bloom in the Baltic sea and its potential effects on biogeochemical cycling. *Front. Mar. Sci*. **5**, 327 (2018).
63. Olli, K. et al. Harmonizing large data sets reveals novel patterns in the Baltic Sea phytoplankton community structure. *Mar. Ecol. Prog. Ser.* **473**, 53–66 (2013).
64. Kleine Bardenhorst, S., Vital, M., Karch, A. & Rübsamen, N. Richness estimation in microbiome data obtained from denoising pipelines. *Comput. Struct. Biotechnol. J.* **20**, 508–520 (2022).
65. Reissmann, J. H. et al. Vertical mixing in the Baltic Sea and consequences for eutrophication – A review. *Prog. Oceanogr.* **82**, 47–80 (2009).
66. García, F. C., Alonso-Sáez, L., Morán, X. A. G. & López-Urrutia, Á Seasonality in molecular and cytometric diversity of marine bacterioplankton: the re-shuffling of bacterial taxa by vertical mixing. *Environ. Microbiol.* **17**, 4133–4142 (2015).
67. Alonso-Sáez, L., Díaz-Pérez, L. & Morán, X. A. G. The hidden seasonality of the rare biosphere in coastal marine bacterioplankton: Seasonality of the rare biosphere. *Environ. Microbiol.* **17**, 3766–3780 (2015).
68. Jurburg, S. D. et al. Microbial community database (MiCoDa). German Centre for Integrative Biodiversity Research https://doi.org/10.25829/8D64 (2024).
69. Vaulot, D. et al. metaPR2 : A database of eukaryotic 18S rRNA metabarcodes with an emphasis on protists. *Mol. Ecol. Resour.* **22**, 3188–3201 (2022).

70. Singer, D. et al. Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems. *Environ. Int.* **146**, 106262 (2021).
71. Taylor, F. J. R., Hoppenrath, M. & Saldarriaga, J. F. Dinoflagellate diversity and distribution. in *Protist Diversity and Geographical Distribution* 173–184 (Springer Netherlands, Dordrecht, 2007).
72. Mtawali, M. et al. Phylogenomic resolution of marine to freshwater dinoflagellate transitions. *ISME J*. **19**, (2025).
73. Shalchian-Tabrizi, K. et al. Diversification of unicellular eukaryotes: cryptomonad colonizations of marine and fresh waters inferred from revised 18S rRNA phylogeny. *Environ. Microbiol.* **10**, 2635–2644 (2008).
74. de Vargas, C. et al. Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 1261605 (2015).
75. Soininen, J., Korhonen, J. J., Karhu, J. & Vetterli, A. Disentangling the spatial patterns in community composition of prokaryotic and eukaryotic lake plankton. *Limnol. Oceanogr.* **56**, 508–520 (2011).
76. De Bie, T. et al. Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. *Ecol. Lett.* **15**, 740–747 (2012).
77. Font-Muñoz, J. S. et al. Advection by ocean currents modifies phytoplankton size structure. *J. R. Soc. Interface* **14**, 130 (2017).
78. Keeling, P. J. & Campo, J. D. Marine protists are not just big bacteria. *Curr. Biol.* **27**, R541–R549 (2017).
79. Agneta, A. et al. DNA extraction protocol for DNA-metabarcoding of marine phytoplankton using Zymobiomics DNA minprep kit (Zymo Research; D4300) v1 https://doi.org/10.17504/protocols.io.bucjnsun (2021).
80. Balzano, S., Abs, E. & Leterme, S. C. Protist diversity along a salinity gradient in a coastal lagoon. *Aquat. Microb. Ecol.* **74**, 263–277 (2015).
81. Wu, L. et al. Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. *BMC Microbiol* **15**, 125 (2015).
82. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
83. Callahan, B. J. et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
84. Parks, D. H. et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
85. Kozlov, A. M., Zhang, J., Yilmaz, P., Glöckner, F. O. & Stamatakis, A. Phylogeny-aware identification and correction of taxonomically mislabeled sequences. *Nucleic Acids Res* **44**, 5022–5033 (2016).
86. Lundin, D. & Andersson, A. SBDI Sativa curated 16S GTDB database. *SciLifeLab* https://doi.org/10.17044/scilifelab.14869077, v3 (2021).
87. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590–D596 (2013).
88. Guillou, L. et al. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res* **41**, D597–D604 (2013).
89. Straub, D. et al. Interpretations of Environmental Microbial Community Studies Are Biased by the Selected 16S rRNA (Gene) Amplicon Sequencing Pipeline. *Front. Microbiol.* **11**, 550420 (2020).
90. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).
91. R Core Team. R: A Language and Environment for Statistical Computing. Preprint at https://www.R-project.org/ (2022).
92. Inkscape Project. *Inkscape*. (2020).
93. Oksanen, J. et al. vegan: Community Ecology Package. Preprint at https://CRAN.R-project.org/package=vegan (2022).
94. Faith, D. P., Minchin, P. R. & Belbin, L. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* **69**, 57–68 (1987).
95. Lai, J., Zou, Y., Zhang, J. & Peres-Neto, P. R. Generalizing hierarchical and variation partitioning in multiple regression and canonical analyses using the rdacca.hp R package. *Methods Ecol. Evol.* **13**, 782–788 (2022).
96. Buliung, R. N. & Remmel, T. K. Open source, spatial analysis, and activity-travel behaviour research: capabilities of the aspace package. *J. Geogr. Syst.* **10**, 191–216 (2008).
97. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
98. Wilson, E. B. Probable inference, the law of succession, and statistical inference. *J. Am. Stat. Assoc.* **22**, 209–212 (1927).
99. Malard, L. A. et al. Comparative analysis of diversity and environmental niches of soil bacterial, archaeal, fungal and protist communities reveal niche divergences along environmental gradients in the Alps. *Soil Biol. Biochem.* **169**, 108674 (2022).
100. Plummer, M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria* (2003).
101. Plummer, M. rjags: Bayesian Graphical Models using MCMC. Preprint at https://CRAN.R-project.org/package=rjags (2023).
102. Assis, J. et al. Bio-ORACLE v3.0. Pushing marine data layers to the CMIP6 Earth System Models of climate change research. *Glob. Ecol. Biogeogr*. https://doi.org/10.1111/geb.13813 (2024).
103. Wessel, P. & Smith, W. H. F. A global, self-consistent, hierarchical, high-resolution shoreline database. *J. Geophys. Res.* **101**, 8741–8743 (1996).
104. Schnute, J. T., Boers, N. & Haigh, R. PBSmapping: Mapping fisheries data and spatial analysis tools. *CRAN: Contributed Packages* The R Foundation https://doi.org/10.32614/cran.package.pbsmapping (2012).

Acknowledgments

We would like to thank Jenny Lycken for the effort put into sampling and sample preparation for the 2019-2020 dataset. National Genomics Infrastructure (NGI) Sweden, and especially Elísabet Einarsdóttir, for seamless and adjustable sequencing service. Karin Garefelt for help with preparing the 2015-2017 samples for sequencing. Additionally, KTJ would like to thank Maliheh Mehrshad for hosting him in Uppsala during the critical period of work on this manuscript, as well as suggesting courses which taught him the skills necessary for the crucial analyses of this study. Ulf Gardin, for help with the multivariate analyses. Matthew Low and Malin Aronsson for help with Bayesian modeling.

Author contributions

K.T.J. analyzed the data under the supervision of A.F.A and with help from MACL and A.T. Y.O.O.H prepared the DNA for the 2015-2017 dataset. K.T.J. wrote the first version of the manuscript under the supervision of A.F.A. K.T.J., M.A.C.L., A.T., S.B., M.H., Y.O.O.H., M.L., A.A., B.K., and A.F.A. contributed to conceptualization, interpretation of the data, revised the manuscript, and approved the final version.

Funding

This work was funded by the Swedish Agency for Marine and Water Management and the Swedish Environmental Protection Agency under grant number NV-03728-17. A.F.A. was additionally funded by a research grant from the Swedish Research Council VR (2021-05563) and from FORMAS (2022-01515). Y.O.O.H. was funded by Science and Technology Program of Xinjiang Uyghur Autonomous Region (E5628503). DNA-extraction and sequencing of the 2015-2017 samples was supported by the EU Horizon 2020 project JERICO-NEXT under grant agreement No 654410 and by the Swedish Research Council FORMAS (grant number 214-2013-1449). M.A.C.L. was funded by Villum Fonden (grant number: 34442). For this study, we used computational resources provided by the Swedish National Infrastructure for Computing (SNIC) through the Uppsala

Multidisciplinary Center for Advanced Computational Science (UPPMAX). Open access funding provided by Royal Institute of Technology.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-026-10977-4.

Correspondence and requests for materials should be addressed to Krzysztof T. Jurdzinski or Anders F. Andersson.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Dr Tobias Goris. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

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

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

© The Author(s) 2026