Title: Distinct bacterial and protist plankton diversity dynamics uncovered through DNA-based monitoring in the Baltic Sea area

Authors:

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], Markus Lindh[3], Agneta Andersson[4,5], Bengt Karlson[3], Anders F Andersson[1,*]

Affiliations:

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 and Environmental Sciences, Umeå, Sweden*
5. *Umeå Marine Sciences Centre, Umeå University, SE-905 71, Hörnefors, Sweden*
6. *Current address: State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China and University of Chinese Academy of Sciences, Beijing 100049, China*

*Corresponding author: Anders F Andersson (anders.andersson@scilifelab.se)

Abstract:

Planktonic microorganisms in coastal waters form the base of food web and biogeochemical cycles. The Baltic Sea area, with its pronounced environmental gradients, serves as a model coastal environment. Yet, microbial diversity assessment across these environmental gradients has so far lacked either taxonomic scope or the integration of spatial and temporal scales. Here, we analyzed protist and bacterial diversity using DNA metabarcoding across 398 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, Bayesian modeling showed that bacterial lineages were less likely than protists to occur in both lower (<9 PSU) and higher (>15 PSU) brackish salinities. Nonetheless, protist alpha diversity increased with salinity. Changes in bacterial alpha diversity were primarily seasonal and linked to influx of deepwater taxa through vertical mixing in winter. We propose that protists are ecologically less sensitive to salinity because compartmentalization allows them to disconnect basic metabolic processes from the cell membrane. Additionally, further and more frequent dispersal of bacteria might impede local adaptation. Ultimately, DNA-based environmental monitoring expands our understanding of microbial diversity patterns and the underlying factors.

Introduction:

Aquatic microorganisms play indispensable roles in biogeochemical cycles[1]. Phototrophic planktonic bacteria and protists account for around half of primary production on Earth[2], with a disproportionately high contribution from coastal ecosystems[3]. More than a third of the human population lives near these ecosystems[4] and depends on them[5,6], while these environments experience extreme anthropogenic pressures[5,7]. Furthermore, coastal zones comprise multiple land-to-sea environmental gradients, including a range of salinities: freshwater-brackish-marine, leading to a large diversity of habitats[5,8].

The Baltic Sea and the adjacent Kattegat-Skagerrak (hence “Baltic Sea area” Fig. 1a) are among the best-studied coastal sea areas[9]. Recent formation (8000 years ago[10]), pronounced environmental gradients, and narrow connecting straits make the Baltic Sea a useful model for studying eco-evolutionary processes driven by dispersal and adaptation[11]. In addition, its early exposure to anthropogenic pressures, especially eutrophication, can foreshadow the processes awaiting other coastal regions[12]. The Baltic Sea area also has a long record of microbial research, with the earliest data coming from the 19th century[13]. Regular microscopy-based phytoplankton monitoring has been run by multiple countries in the region for decades[14] and since the 1980s in Sweden[15]. More recently, the Baltic Sea area was one of the first brackish waters surveyed using DNA metabarcoding[15–17] and metagenomics[18,19]. Nevertheless, most of the unknown diversity in the Baltic Sea area is believed to be microbial[20]. Up until now, research has mainly focused on specific groups of phytoplankton and metazooplankton[20], with much of the microbial diversity missing from the picture.

Globally, salinity is among the strongest factors structuring bacterial[21–24] and protist[25] communities. However, the importance of salinity has largely been inferred from the divide between freshwater and marine environments[26–28]. Previous research shows that brackish waters host unique bacterial species[19,29]. Both microscopy- and DNA-based surveys in the Baltic Sea area suggest pronounced shifts in surface water microbial community composition (beta diversity) along the horizontal salinity gradient[16,17,30–32]. Moreover, bacterial communities in the Baltic Sea change more with salinity than between winter and summer[33]. More generally, results obtained across steep salinity gradients in Antarctica suggested stronger environmental filtering among planktonic bacteria than protists[34,35], while studies in other environmental contexts came to the opposite conclusion[36–39]. Whether bacterial communities are more affected by salinity in particular, or generally more/less environmentally filtered than protist ones, is yet to be determined.

The number of aquatic animal species tends to reach a minimum within the brackish spectrum across salinity gradients[20,40]. For phytoplankton, patterns of both a maximum[30] and a minimum[31] of species richness (alpha diversity) within the brackish salinity spectrum, as well as an increase along the salinity gradient[32], have been

reported based on microscopic assessments. DNA metabarcoding transect surveys of plankton showed no clear patterns for bacteria[16,17,33] nor major eukaryotic taxa[17]. At the same time, bacterial alpha diversity has been observed to be lower in winter than in summer in the Baltic Sea[33,41], consistently with patterns observed in marine environments[42–44]. In contrast, microscopy-based analysis of phytoplankton alpha diversity has shown different seasonal patterns in low and high salinities in the Baltic Sea area[32]. Little is known about the factors driving the seasonal patterns or confounding diversity and community composition changes along the salinity gradient.

Here, we present the first in-depth comparative analysis of bacterial and protist diversity across spatiotemporal scales in the Baltic Sea area and across the corresponding environmental gradients. We use a recently published DNA metabarcoding dataset[45] and new data released here, which together span two complete calendar years and include 398 samples. We show how protists and bacteria differ in their ecological sensitivity to salinity and analyze drivers of taxonomic richness. Finally, we propose an explanation of the contrasting bacterial and protist diversity patterns based on fundamental physiological and ecological differences between the groups.

Results and discussion

Processing of metabarcoding data for ecologically relevant metabarcoding-based analyses

We based the first part of our analyses on a publicly available 16S and 18S metabarcoding dataset, covering broad spatiotemporal scales and environmental gradients regularly sampled throughout a year[45]. After filtering procedures (see Methods for details), we used the data from 246 independent samples. They came from 18 locations, extending from the Bothnian Bay (minimum salinity = 2.0 PSU) to the Skagerrak (maximum salinity = 33.92 PSU), each location sampled at least six times between January 2019 and February 2020 (Fig. 1a).

From the 16S amplicon sequence variants (ASVs), we chose only those classified as *Bacteria* and excluded plastids and mitochondria. From the 18S metabarcoding results, we excluded ASVs classified as animals (*Metazoa*), fungi, land plants (*Embryophyceae*), and selected macroalgae. We thus treated the remaining 16S ASVs as bacteria and 18S ASVs as protists.

We clustered the ASVs by simultaneous genetic and distribution similarity, using distribution-based operational taxonomic unit-calling (dbOTU[46]; we thus call those taxonomic units dbOTUs). This method groups variants corresponding to different marker gene copies within one organism and other ecologically indistinguishable intra-species variation. Therefore, dbOTUs should correspond to ecologically coherent

lineages unless the sequenced regions are identical between organisms with distinct ecologies. More than half of dbOTUs were annotated to at least family level for both bacteria and protists (Supplementary Fig. S1c-d). More than half of the reads were annotated to at least genus or family level for bacteria and protists, respectively (Supplementary Fig. S1e-f).

Finally, a controlled amount of synthetic DNA amplifiable by the metabarcoding primers, i.e., spike-in, was added to the samples. This approach aims to obtain absolute values (copy number per volume sample) for the marker gene (16S/18S ribosomal DNA (rDNA)), which better estimates absolute abundance than relative counts[47]. Thus, it can uncover blooming and succession patterns obscured from relative abundance analysis[48]. At the same time, the spike-in approach is highly sensitive to differences in sample preparation and amplification[47]. This can lead to noisier results than relative abundance, especially in environmental monitoring, as samples are often collected by multiple researchers working in variable conditions.

Regionally distinct seasonal dynamics of major bacterial and protist taxa

Major bacterial (Fig. 1b) and protist (Fig. 1c) clades displayed distinct seasonal abundance dynamics. Both their abundance 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).

We observed an increased abundance of cyanobacteria as early as May in the Baltic Proper (Fig. 1b), consistent with microscopic observations[49]. The increase in cyanobacterial abundance was smaller and peaked later moving northwards within the Baltic Sea. In the Baltic Proper, both dinoflagellates and *Gyrista* (a protist subdivision that includes diatoms), peaked in abundance in March (Fig. 1c), as observed using microscopic methods[49,50]. The dynamics of *Dinoflagellata* and *Gyrista* differed vastly between basins. However, in all cases except for Bothnian Bay, those groups increased in abundance in the first half of the year. Furthermore, in the Kattegat and the Skagerrak, there was an additional dinoflagellate peak in September, corresponding to autumn blooms. *Haptophyta* generally displayed increased abundance between May and August. *Cryptophyta* were abundant in the Baltic Sea, with higher abundance in the second half of the year than in the first.

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[41,51], 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[16–18]. 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[41,51], *Alphaproteobacteria*, and *Gammaproteobacteria* comprised a stable proportion of the whole community throughout the year in all basins but the Kattegat (Supplementary Fig. S2a). Nevertheless, consistent with metabarcoding[16] and metagenomic transects[18], *Alphaproteobacteria* and *Gammaproteobacteria* were more abundant in the Kattegat-Skagerrak than in the Baltic Sea.

Overall, we observed geographic differences in both total abundance and seasonal dynamics of major bacterial and protist taxa. There were some latitudinal shifts, e.g., for cyanobacteria and *Acidimicrobiia*, and apparent differences between Kattegat-Skagerrak and the Baltic Sea. Compared to relative abundance-based analysis, spike-in correction added ecologically relevant information, as best exemplified by the March dinoflagellate bloom in the Baltic Proper (Fig. 1c) being indistinguishable based on the relative abundance of *Dinoflagellata* (Supplementary Fig. S2b). Still, the spike-in approach entailed substantial random noise, especially in less sampled basins (Fig. 1b-c) and relative abundances better captured gradual succession patterns, especially for bacteria (Supplementary Fig. S2a). Thus, the best normalization approach depends on the goals of the analysis and suitable statistical methods, at the very least until the consistency of absolute abundance quantification improves.

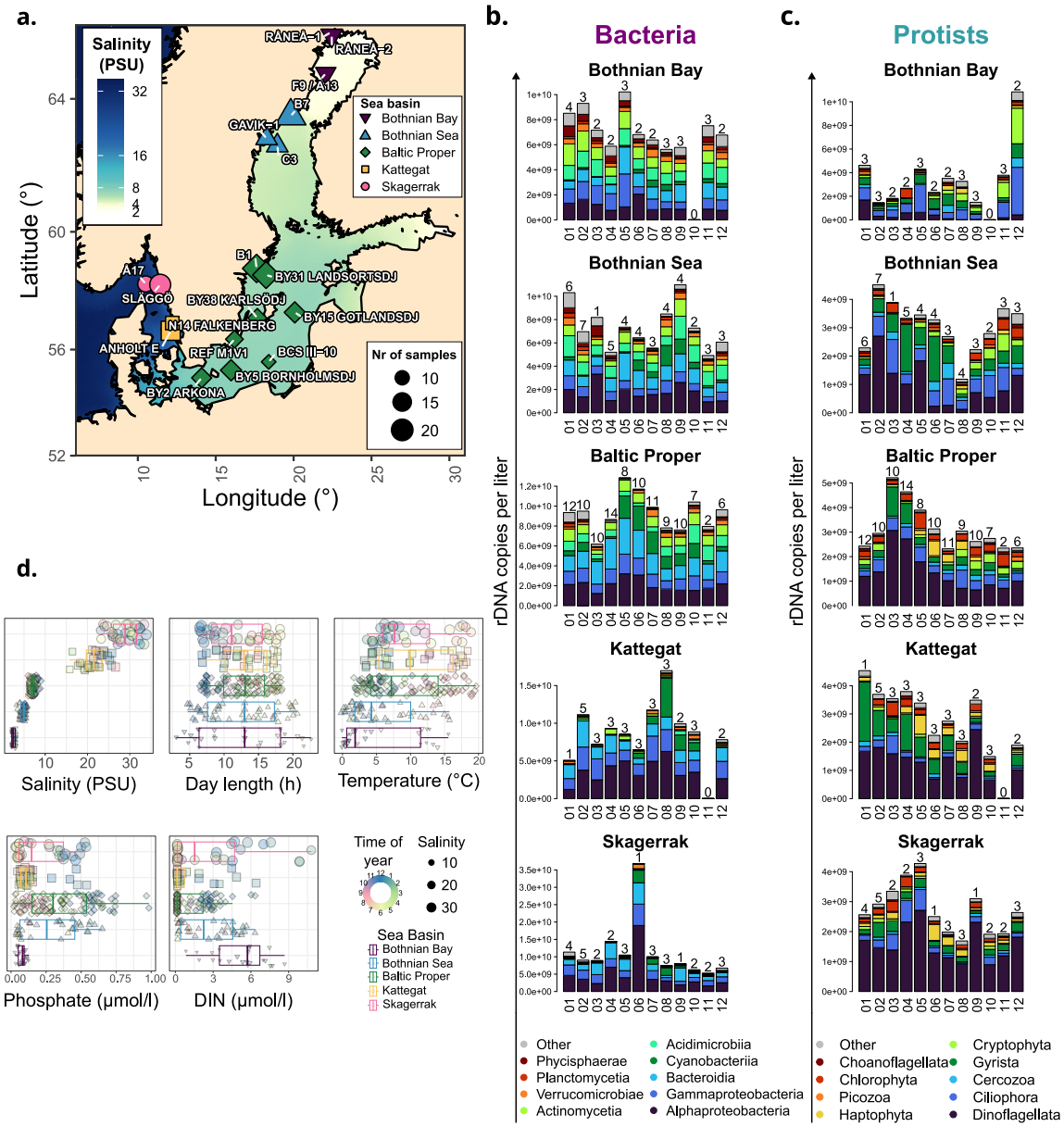


Fig. 1 | Spatiotemporal changes in abundance of major microbial taxa, 2019-2020 a. A map of 18 sampling locations labeled by the name of the station and size-coded by the number of samples collected over a period of 13 months (January 2019 to February2020). **b-c.** The numbers of rRNA marker gene copies per liter annotated to **b.** bacterial classes and **c.** protist subdivisions averaged over each 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 >0.01 of reads across samples are shown. 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. 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 (R2 = 0.58). However, the protist communities were generally less similar to each other than bacterial ones (Supplementary Fig. S3a), as has been reported for picoeukaryotes and bacteria in ocean surface waters52.

Principal Coordinate Analysis (PCoA) showed distinguishable geographic and seasonal structuring of bacterial and protist communities (Supplementary Fig. S4). 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 (R2 < 0.4, Supplementary Fig. S5). The dbRDA-based variation partitioning showed that salinity could explain almost 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). Importantly, these results contradict the proposed rule that the distribution of smaller organisms is less limited by environmental factors than that of larger organisms[53], and in particular that bacteria are less environmentally filtered than protists[38]. At the same time, our results are consistent with a previous report of a stronger selection of bacterial than protist communities across a salinity gradient[35]. However, they do not support the generalization of the stronger selection to the effects of other environmental factors.

Nevertheless, for both bacterial and protist communities, there was a pronounced salinity divide between the Baltic Sea and the Kattegat-Skagerrak and a more gradual latitudinal shift within the Baltic Sea (Fig 2c-d). The latter corresponded to decreasing salinity and phosphate-to-nitrogen ratios with increasing latitude (Fig. 1d). We connected those spatial changes in beta-diversity, calculated from ASV-based dissimilarities, with high relative abundance of specific microbial families (Fig 2c-d). For example, typically freshwater *Nanopelagicaceae* (acl *Actinobacteria*)[54] and *Limnocylindraceae* (*Chloroflexi*)[55] were associated with the lowest salinity samples, while commonly marine *Rhodobacteraceae*[56] and *Thioglobaceae*[57] were at the other end of the salinity spectrum (Fig 2c).

Moreover, 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. S6). Diverse groups significantly contributed to the seasonal patterns in beta diversity, ranging from known seasonally blooming diatoms *Thalassiosiraceae* and *Skeletonemaceae*[58] to multiple uncharacterized *Bacteroides* families. In all the basins, picocyanobacteria (*Cyanobiaceae*) were associated with high

temperatures, while BACL11 (*Bacterioides*[19]) were typical of winter conditions (Fig. 2e, Supplementary Fig. S6a).

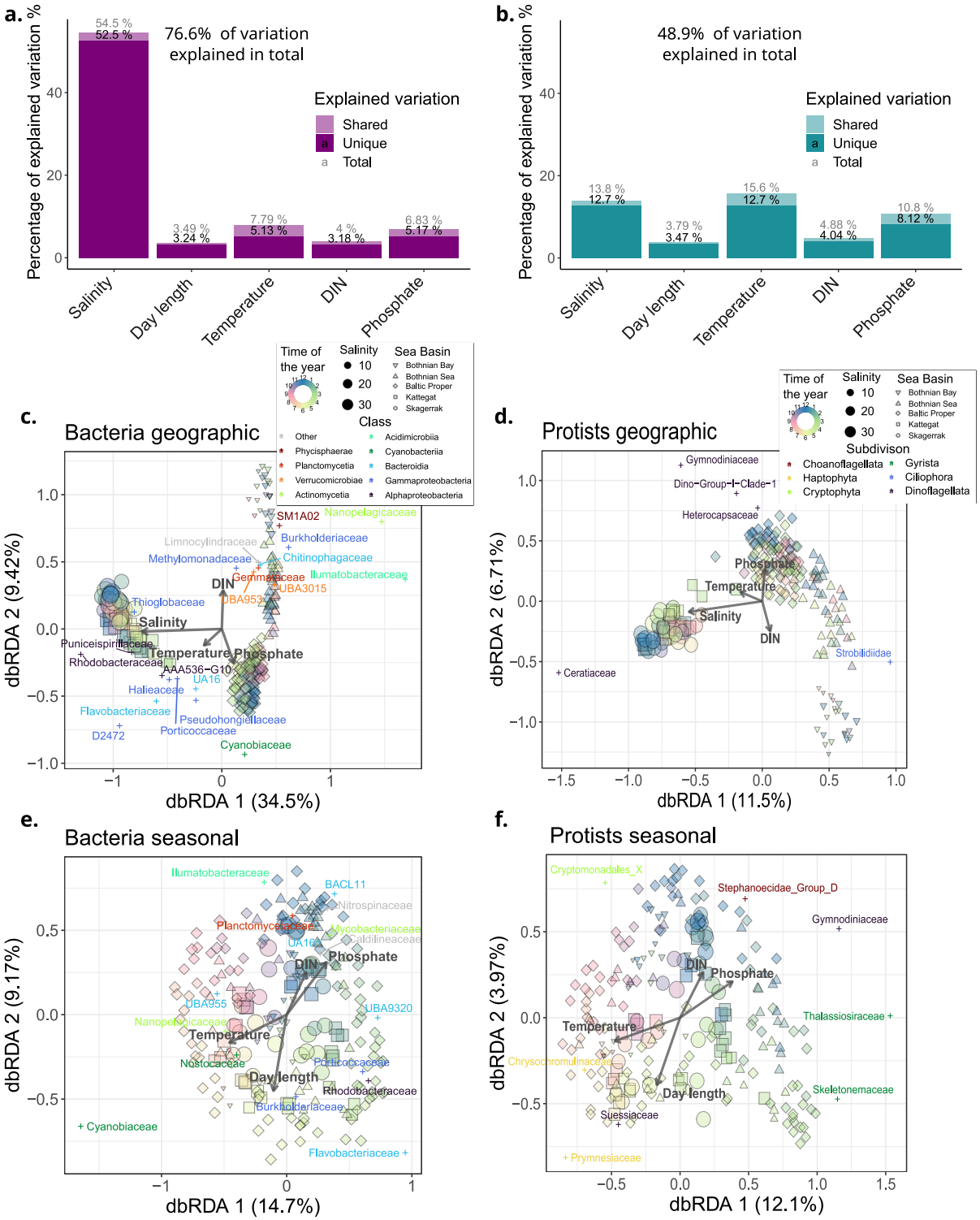


Fig. 2 | Environmental factors explaining geographic and seasonal community structuring and the most correlated families of bacteria and protists. a-b. Variance partitioning of bacterial (**a.**) and protist (**b.**) beta diversity, based on Bray-Curtis distances and

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 deviating (Bonferroni corrected $P < 0.05$) were placed on the plot, with values rescaled by a factor of 0.5. Percentages of variation explained by each dbRDA component are given in brackets by the axes labels. DIN - dissolved inorganic nitrogen. Salinity is given in practical salinity units (PSU).

Interannual stability of microbial community composition

To assess the interannual stability of the community composition patterns we observed, we performed metabarcoding on stored DNA extracted from samples collected between 2015 and 2017 (Fig, 5a-b). The new dataset, corresponding to 143 samples, covered the southern range of the 2019-2020 dataset (Fig. 1a, Fig. 3a). Similarly to the 2019-2020 dataset, it mainly corresponds to a 13-month long time-series, between February 2016 and March 2017, with a few extra samples from 2015 from one station (Släggö, Fig. 3a). There were minor differences in sample processing between the two datasets, most notable being: DNA extraction kit, different Illumina sequencing systems, and lack of spike-in sequences in the 2015-2017 dataset (see Methods for more details).

Overall, the community composition in the two datasets differed minimally, following the spatial and seasonal environmental gradients analogously (Fig. 3c-f). This apparent interannual stability of community composition is consistent with previous observations from the Baltic Sea[51,59] and other waters[39]. Finally, the similarity of communities between the two datasets shows that the technical differences had little impact on the observed general community composition (Fig. 3e-f).

Only a two-year break separated the datasets. The seasonal patterns likely change over longer time scales, as, e.g., climate change has altered the phytoplankton bloom dynamics in the Baltic Sea over decadal scales[60]. As climate change accelerates[61] and nutrient inputs are being decreased in the Baltic Sea[62], the microbial communities will surely undergo long-term changes. The introduction of invasive species can also alter microbial community dynamics[63]. A reproducible monitoring scheme encompassing diverse taxa, achievable with DNA-based methods, is crucial for understanding and informed responses to unforeseen environmental changes.

In contrast to community composition, alpha diversity differed significantly ($P < 10^{-14}$ in both cases) for both bacteria and protists (Fig. 3g-h). We doubt the increased diversity in 2015-2017 dataset, amounting to hundreds of dbOTUs on average, represented an actual biological change. Instead, it likely originated from the technical differences between the datasets. We used rarefaction for alpha-diversity estimates, as it performs best in removing sequencing depth biases[64] (expected due to different Illumina sequencing systems used for the datasets). Nevertheless, in both datasets,

273 rarefaction curves were saturated, suggesting that extra reads would not change alpha
274 diversity much (Supplementary Fig. S7). Thus, another technical difference, remaining
275 to be identified, likely caused the dbOTU difference between datasets. Ultimately, our
276 results suggest more caution is needed when comparing alpha diversity than beta
277 diversity across datasets. As a consequence, in subsequent alpha diversity analyses,
278 we focused solely on the 2019-2020 dataset because of its larger geographic scope.
279

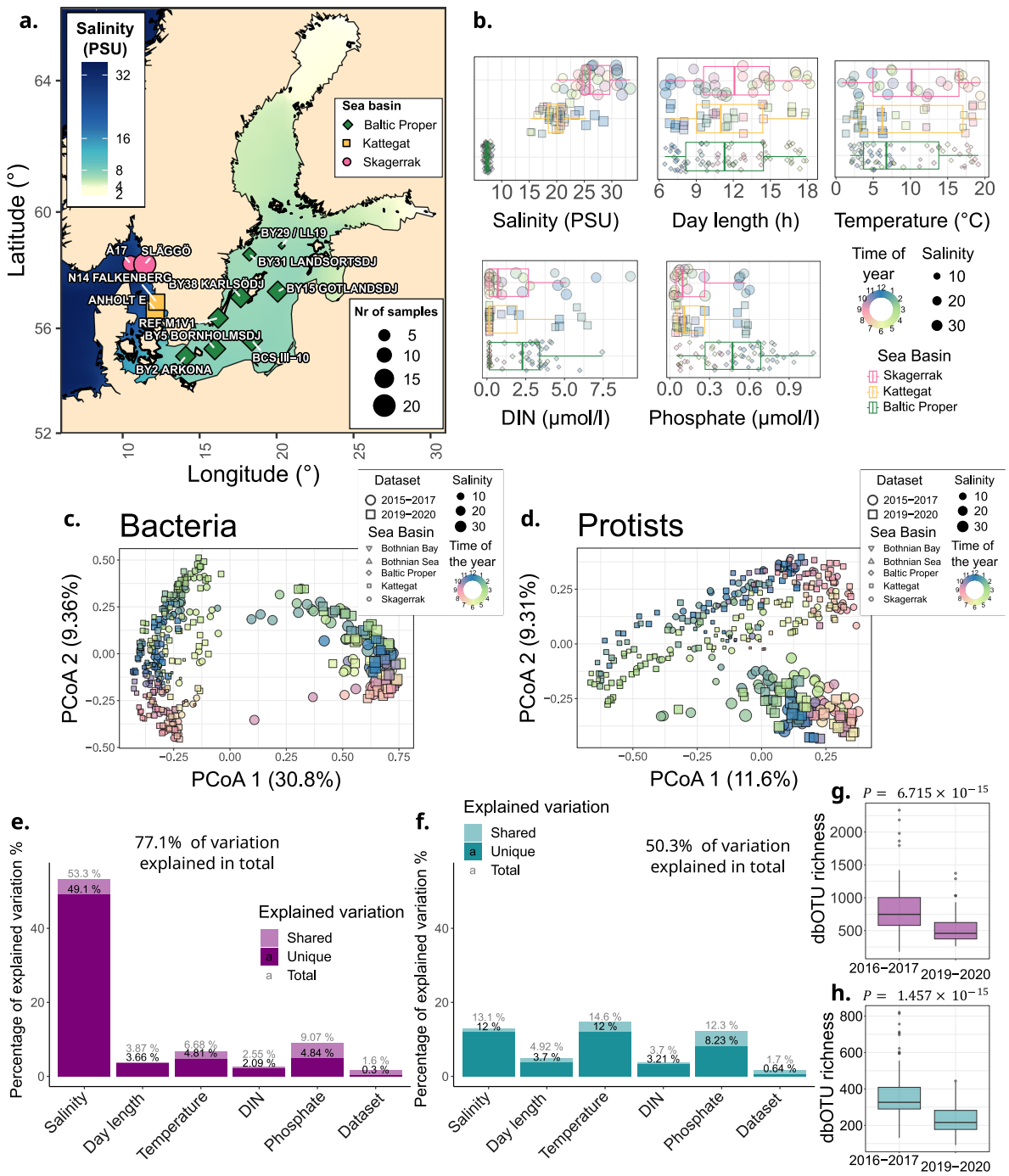


Fig. 3 | Bacterial and protist diversity across datasets collected in 2015-2017 and in 2019-2020. a. A map of 12 sampling locations labeled by the number of samples collected between August 2015 and March 2017 (the 2015-2017 dataset), with all the 2015 samples coming from station Släggö in Skagerrak **b.** The distribution of measurements of major physicochemical and geographic parameters for each basin in the 2015-2017 dataset. **c-d**. PCoA based on Bray-Curtis dissimilarities of bacterial (**c.**) and protist (**d.**) communities across the 2015-2017 and 2019-2020 datasets. **e-f.** Variance partitioning of bacterial (**e.**) and protist (**f.**) diversity, based on Bray-Curtis distances and dbRDA. Dataset stands for a dummy variable differentiating the

samples collected in 2015-2017 and 2019-2020. **g-h.** dbOTU richness of bacteria (**g.**) and protists (**h.**) 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 Wilcoxon (signed-rank) test are given above each plot. DIN - dissolved inorganic nitrogen. dbOTU - distribution-based operational taxonomic unit.

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

We assessed the changes in alpha diversity along the salinity gradient, for both bacteria and protists, following the methodology used to argue for or against maximum[17,30] and minimum[31] protist diversity along the gradient (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. 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 its variation (Fig. 4e). Moreover, DIN followed similar seasonal patterns of rapid decline and increase (Supplementary Fig. S8a) and correlated with seasonal rather than geographic variation in dbOTU richness (Supplementary Fig. S8b).

However, the 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 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[33]. However, these patterns may originate from similar conditions across the water column, leading to related bacteria thriving. Evidence that deepwater bacteria actually influx to the surface in winter is missing.

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[33], from here on referred to as the “transect-based dataset”. We matched the sequences representative of the OTUs (operational taxonomic units) distinguished in the transect-based dataset to ASVs representative of our dbOTUs (see Methods for

details). 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[68].

We further chose from the transect-based dataset only the summer (July) samples, corresponding to the time of strong thermocline. Among those, we categorized OTUs as found only in deep (>10m depth) or in surface (≤10m depth) waters, or in both. We assessed how many OTUs from each category matched the dbOTUs from winter (December-February) and summer (June-August) months (Fig. 5a). We excluded Bothnian Bay from the analysis since there were no deepwater samples from this region in the transect-based dataset. Among the deepwater OTUs, significantly more ($P < 10^{-75}$) were found in winter than in summer despite originating from summer samples. In contrast, more OTUs found in both deep and surface water were detected in summer ($P < 10^{-8}$), while among the surface OTUs, there was no significant difference between the seasons ($P = 0.67$). Ultimately, these results suggest a substantial proportion of taxa found only in deep waters in summer can be found at the surface during winter.

The proportion of detected deepwater OTUs in a sample correlated strongly with dbOTU richness, with a clear seasonal pattern (Fig. 5b). In contrast, the proportion of detected non-deepwater-specfic OTUs ('surface' and 'both' from Fig. 5a) showed a weaker correlation with richness (Supplementary Fig. S8c), and did not follow the proportion of deepwater OTUs matched (Supplementary Fig. S8d). Thus, it was not the detection biases that underpinned the correlation between the proportion of deepwater OTUs and richness. Moreover, after partitioning out the correlations with deepwater OTUs, DIN no longer correlated with dbOTU richness (Fig. 5c). This relation was not bilateral, as the proportion of deepwater OTUs still correlated with dbOTU richness after partitioning out DIN (Supplementary Fig. S8e). Therefore, the correlation between DIN and bacterial diversity comes from a common confounder: seasonal convective water mixing.

However, the richness of dbOTUs matching to non-deepwater-specific OTUs still correlated with the proportion of deepwater OTUs (Fig. 5d). Simultaneously, we detected more non-deepwater-specific OTUs in summer than in winter (Fig. 5a). Therefore, the same taxa were more persistently detected across the samples in winter. This could not be explained by a few dominant taxa pushing others below detection limits in summer, as the correlation remained after partitioning out Pielou's evenness index[69] (Supplementary Fig. S8f). Thus, some consequences of vertical mixing or parallel processes, e.g., lower irradiation[44,70,71], lead to a more persistent or geographically widespread presence of the same taxa in the surface waters in winter than in summer.

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 near-marine conditions (in Skagerrak; Fig. 5e). In contrast, surface water-specific taxa never reached cumulative relative abundance higher than 0.1 (Supplementary Fig. S8g), while taxa

368 found in both surface and deep waters always corresponded to more than half of the
369 reads (Supplementary Fig. S8h). Thus, the ‘both’ category likely includes abundant
370 surface water taxa found also in deeper waters in summer due to sinking.
371 With cumulative relative abundance commonly above 0.1 in winter (Fig. 5e), the
372 deepwater taxa likely impact the surface water ecosystem, if not through metabolic
373 activity, at least by providing an alternative food source. Even if the deepwater taxa can
374 survive only briefly at the surface, they can still substantially impact longer-term
375 community dynamics[72]. A further investigation of the consequences of these
376 community coalescence events is needed.
377 Ultimately, the higher mean bacterial richness in the least saline Bothnian Sea
378 and Bothnian Bay may well be explained by longer vertical mixing periods due to colder
379 climates. In summer, when mixing ceased, richness dropped to similar levels across
380 basins (Fig. 4c), consistent with previous findings reporting no salinity-dependent
381 pattern of bacterial alpha diversity[16,17,33].

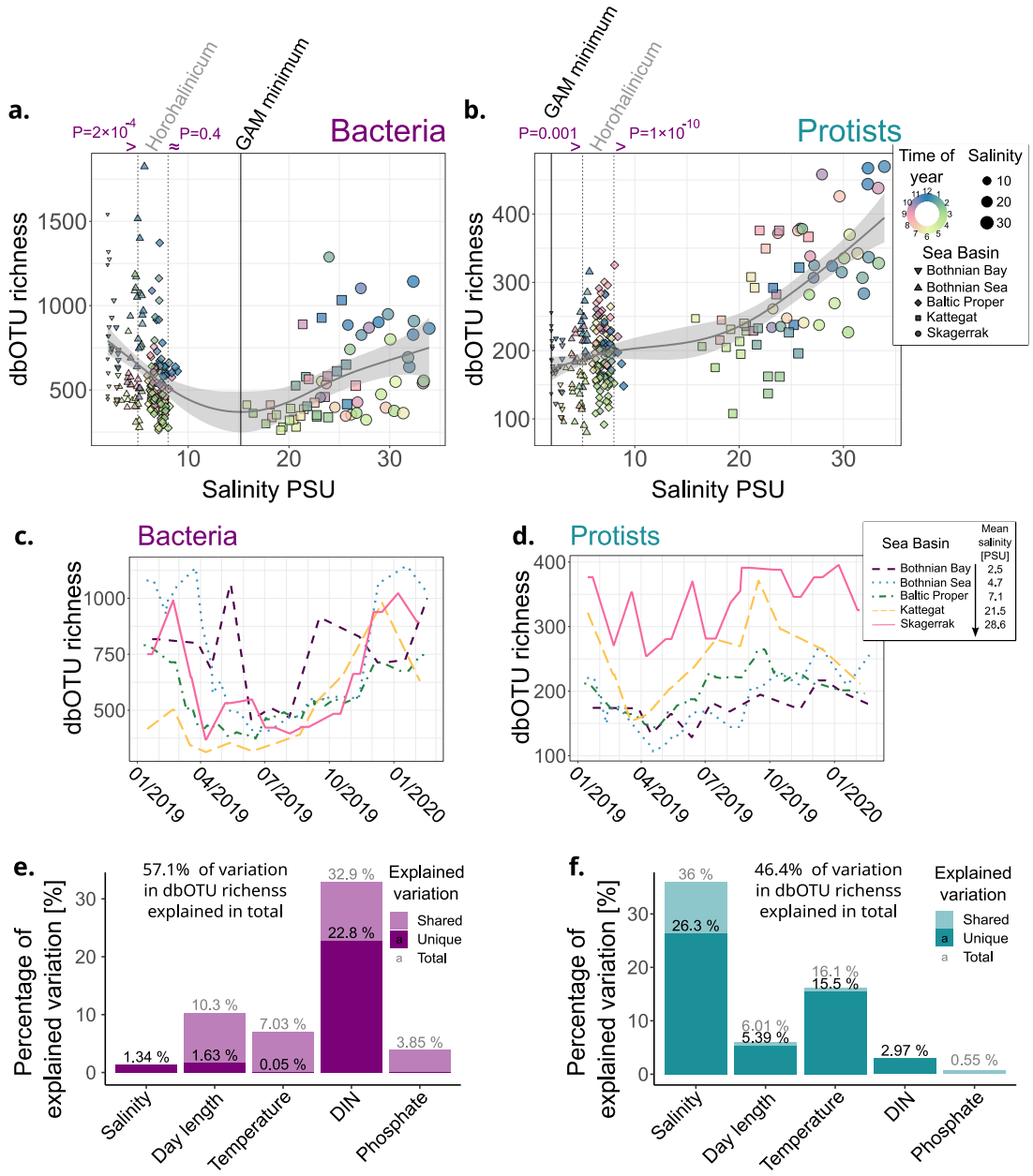


Fig. 4 | Bacterial and protist richness Mean number of bacterial (**a.**) and protist (**b.**) dbOTUs found across 100 refraction iterations (dbOTU richness). The fit corresponds to smoothened generalized additive models (GAMs). The dotted vertical lines correspond to boundaries of horohalinicum (5-8 PSU). The colored (in)equality signs over the dotted lines mark the significant differences of dbOTU richness between horohalinicum and lower/higher salinities, and the P-values from 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 panels a-b. **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 two weeks before and after each date. The same legend applies to panels c-d. **e-f**. Variance partitioning of bacterial (**e.**) and protist (**d.**) dbOTU richness.

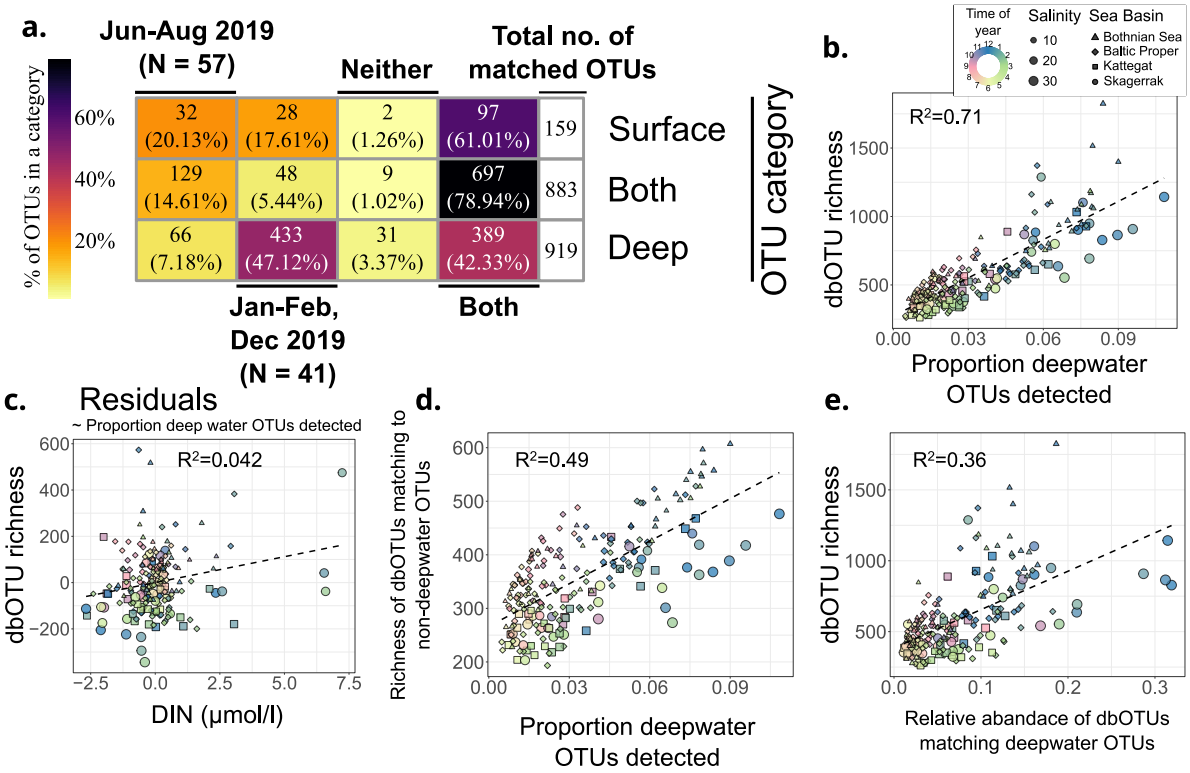


Fig. 5 | Bacterial alpha diversity and the proportion of deepwater OTUs detected a. Number of OTUs from a previous, transect-based study[33] matched to dbOTUs found in winter and summer samples or in neither or both seasons. The OTUs are categorized by being found exclusively in deep water (> 10m depth) samples, exclusively in surface waters, or in both, in the original study. Only OTUs matching to any dbOTUs from this study are counted. **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.** Proportion of all deepwater OTUs vs. richness of bacterial dbOTUs matching to non-deepwater-specific OTUs ("surface" and "both" categories). **e.** Relative abundance of all dbOTUs matching a deep water 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[33]. The same legend applies to figures b-e. DIN - dissolved inorganic nitrogen.

Higher protist alpha diversity in near-marine salinities

In contrast to the situation for bacteria, protist richness displayed a clear salinity-dependent pattern, increasing with salinity (Fig. 4b; linear fit yields R^2=0.44). This is consistent with a recent microscopy-based study[32]. However, unlike in the microscopy-based study, richness in the near-marine Kattegat-Skagerrak was higher than in the Baltic Sea across the year (Fig. 4d).

415 Similarly to our observations on protists, metazooplankton and fish alpha
416 diversity in the Kattegat-Skagerrak is higher than in the Baltic Sea due to many marine
417 species in thriving in higher but not lower brackish salinities, outweighing the presence
418 of freshwater species in lower salinities[20]. The Baltic fauna comes mainly from a limited
419 number of local adaptations to brackish conditions following the last glaciation
420 period[11,20]. In contrast, for Baltic Sea bacteria, we recently showed evidence against
421 widespread occurrence of such local adaptation, but rather colonization by a global
422 brackish microbiota[29]. Consistently, this (Fig. 4a, c) and previous studies[16,17,33] showed
423 no evidence of decreased bacterial diversity in the Baltic Sea compared to Kattegat-
424 Skagerrak. Whether the observed protist diversity pattern originates from a selected
425 group of locally adapted species, as is the case for animals, is yet to be determined.
426 Temperature was the second factor most strongly associated with protist alpha
427 diversity (Fig. 4f), and in most basins, dbOTU richness increased from March to
428 September (Fig. 4d). Temperature was also the factor best explaining protist beta
429 diversity (Fig. 2b). Therefore, our results suggest that both common and rare protists
430 might be more susceptible to the direct effects of climate change than bacteria.

431 Protist dbOTUs more often occur at both high and low brackish 432 salinities

433 The salinity gradient in the Baltic Sea area includes a rapid shift across the Danish
434 straits. Consequently, the data we used contained geographically proximate samples
435 from lower (<9 PSU) and higher (>15 PSU) brackish salinities (Fig. 1a, Fig. 3a). The
436 above analyses showed this salinity barrier to have the strongest structuring effect on
437 the microbial communities, though less so for protists than for bacteria (Fig. 2 and 3).
438 On the other hand, protist but not bacterial alpha diversity was higher on the more
439 saline side of the barrier (Fig. 4).
440 We further investigated whether bacterial or protist dbOTUs were generally less
441 likely to cross this salinity barrier between lower and higher salinities. To minimize the
442 effects of other environmental filtering effects (especially climate), we chose the same
443 number of stations and samples on both sides of the salinity barrier (see Methods for
444 details). Additionally, we based further analyses on the 2019-2020 dataset, which
445 allowed us to use spike-in normalized counts as an abundance metric comparable
446 between bacteria and protists. The proportion of protist dbOTUs crossing the salinity
447 barrier across the subsequently selected samples was significantly higher than that of
448 bacterial dbOTUs (Fig. 6a).
449 Still, the observed proportion of barrier-crossing bacteria and protists may be
450 affected by detection biases coming from differences in relative abundance, length of
451 vegetative season, and environmental filtering by factors other than salinity. Bacteria
452 were, on average, more abundant in terms of rDNA gene marker copy numbers
453 (Supplementary Fig. S9a). As protists tend to have higher rDNA copy numbers per cell

454 than bacteria[73], this difference likely translates to an even larger difference in cell
455 numbers, a more relevant abundance measure in the context of dispersal capabilities
456 and detection probability. Meanwhile, on average, protists were present in more
457 samples on their preferred side (side with higher occupancy) of the salinity barrier
458 (Supplementary Fig. S9b). Overall, similarly widespread bacteria were more abundant
459 than protists (Supplementary Fig. S9c-f). Still, across the occupancy values, a higher
460 proportion of protists crossed the salinity barrier (Fig. 6b). The same was true across
461 the abundance values, apart from the most abundant outliers, which were barrier-
462 crossing dbOTUs among bacteria but not protists (Fig. 6c). Generally, across different
463 occupancy and abundance values, a higher proportion of protists crossed the salinity
464 barrier (Fig. 6d, Supplementary Fig. S9g-h).
465 To confirm that protists are more likely to cross the salinity barrier, we build a
466 Bayesian model using a Monte Carlo Markov Chain (MCMC) with Just Another Gibbs
467 Sampler[74] (JAGS). The model, based on the Bernoulli distribution, calculates the
468 probability of crossing a barrier from logistic regression from the maximal occupancy on
469 one side of the barrier, geometric mean abundance, and whether a dbOTU is a protist
470 or a bacterium. After model selection using the Watanabe-Akaike information criterion[75],
471 we chose a version of the model assuming the probability dependence on the other
472 variables follows different functions for bacteria and protists (see Methods for details).
473 The model predicted protists to be more likely to cross the salinity barrier across the
474 occupancy and abundance values, both on average (Fig. 6e, Supplementary Fig. S9i-j)
475 and in a majority of MCMC iterations (Fig. 6f). The probability that protists are more
476 likely to cross the salinity barrier (Fig. 6f) was consistently close to one except for
477 dbOTUs with low occupancy and high abundance, in which case detection biases might
478 outweigh the biological signal.
479 Finally, using a similar approach, we built a Bayesian model of total occupancy
480 as a function of the occupancy on the more occupied side of the salinity barrier, based
481 on the Poisson distribution. From this model, we calculated the predicted occupancy on
482 the less occupied side of the salinity barrier and compared it to the experimental values
483 (Fig. 6g). The model and the data showed that protists not only more often crossed the
484 salinity barrier but also were, in general, more widespread in the less preferable salinity
485 regime.

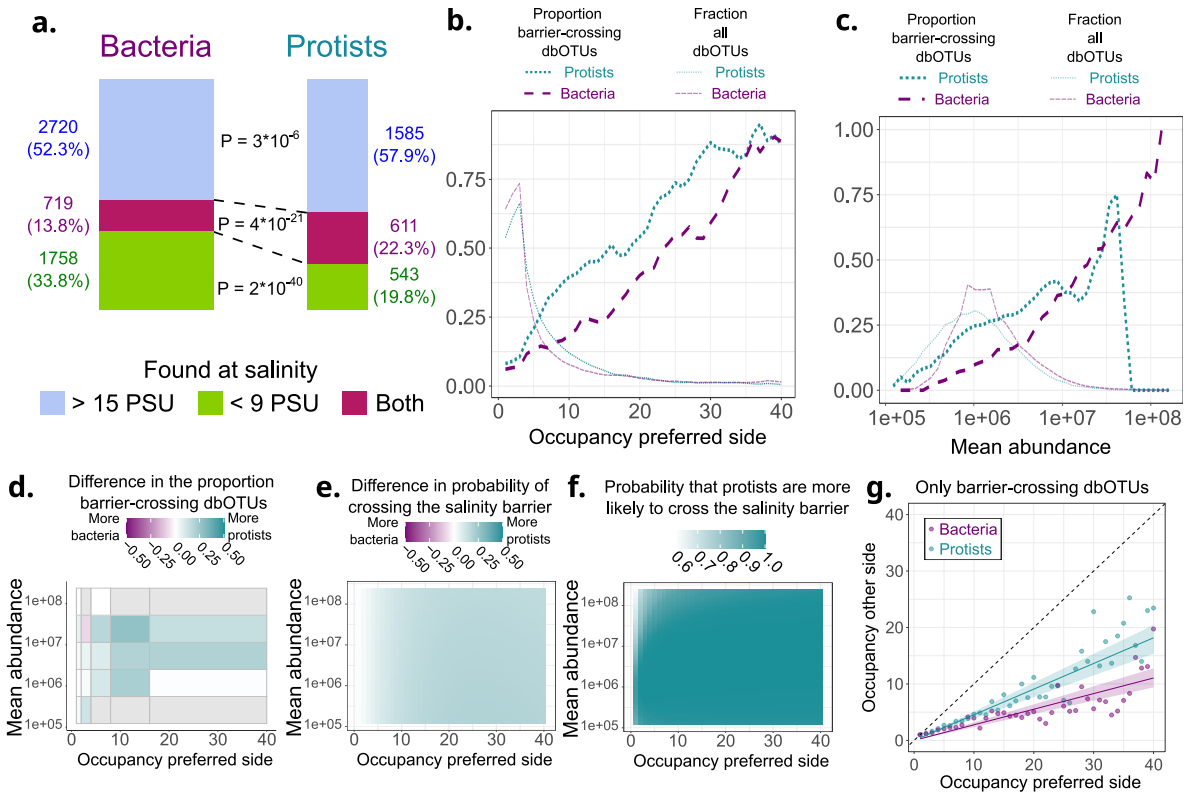


Fig. 6 | Proportion and probability of bacteria and protists (being) found in both high (>15 PSU) and low (<9 PSU) brackish salinity. a. The number and proportion of bacterial and protists dbOTUs found in high (>15 PSU), low (<9 PSU) brackish salinity, and across the salinity barrier. **b-c.** The proportion of barrier crossing dbOTUs for each value of maximal occupancy at one side of the salinity barrier (**b.**) and across forty bins of mean abundance (**c.**). The proportion of all dbOTUs that lie within each bin is also given. The values are rolling means across +/- two bins. **d.** The difference in the proportion of protist and bacterial barrier crossing dbOTUs. Grey tiles correspond to the values across which either no protist or no bacterial dbOTUs were found. **e-g.** Results from Bayesian modeling. **e.** The mean difference in the predicted probability of a protist and a bacterial dbOTU crossing the salinity barrier. **f.** The probability (proportion of MCMC iterations) that protists are more likely to cross 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. The points correspond to means of the experimental values. The lines correspond to the values predicted from the model with 95% confidence intervals. The dashed black line is the identity line. For all the analyses, only the four stations with salinity >15 PSU, and four out of the most saline stations with salinity <9 PSU are chosen. The data for each station is downsampled to the same number of observations. During downsampling, samples closest in time to the samples from the least sampled station were chosen. Only observations from 2019-2020 are included. The abundance values correspond to numbers of rDNA copies per liter, based on the spike-in normalization, and are always given on a logarithmic scale. “Occupancy preferred side” and “occupancy other side” refer to the occupancy at the side with highest and lowest occupancy, respectively.

Proposed explanation of the stronger ecological sensitivity to salinity among bacteria than protists

We propose that compartmentalization makes eukaryotic cells (thus, protists) more acclimatable than prokaryotic ones (thus, bacteria) to a different salinity regime (Fig. 7). In a compartmentalized cell, crucial metabolic processes take place across internal membranes. External ion concentrations and pH do not directly affect these processes. Notably, in eukaryotic cells, the proton motive force (PMF) used for ATP synthesis is maintained along the inner membranes of mitochondria and plastids. In contrast, prokaryotes usually produce PMF across the cellular membrane directly exposed to the external environment. While eukaryotic cytosolic pH is kept at stable, neutral levels[76], marine and brackish waters are buffered and usually slightly alkaline (pH~8)[27,77]. Thus, in the presence of sea salt, some transferred protons are consumed in reactions with (bi-)carbonates or hydroxide. Consequently, many marine bacteria preferably use sodium motive force (SMF), i.e., Na^+-gradient across the membrane, for transmembrane transport, motility, and/or ATP production[18,27,78,79]. The competitive advantage of this strategy diminishes with decreasing salinity (in a logarithmic manner, assuming stable intracellular ion concentrations[80]), and so does likely the relative abundance of bacteria using it. Additionally, many protists have organelles with osmoregulating functions, especially the contractile vacuole[81].

The bioenergetics-based explanation would imply that eukaryotes are also ecologically less sensitive to pH. Indeed, that has been observed when comparing bacterial and fungi soil communities[82]. The role of bioenergetics would also imply testable predictions. One is that adding a salt mix that includes (bi-)carbonates, compared to adding pure NaCl, should have a substantially larger impact on bacterial communities, but less so on protist ones. This is expected since (bi-)carbonates introduce a pH buffering and alkalizing effect, neutralizing the protons exported outside the outer membrane. Another prediction would be that the shifts in community composition along the salinity gradients largely follow a shift towards more bacteria using SMF and less using PMF. However, PMF has bioenergetic advantages, and even some halophilic bacteria developed adaptations to use it[83]. Similarly, the bacteria that recently transitioned to higher salinity rarely acquired the bioenergetic machinery to maintain SMF[29]. Thus, SMF-independent adaptations may also allow bacteria to thrive in higher salinities. Nevertheless, around 70% of marine bacteria were predicted to use SMF based on their genomes, compared to 4% in freshwater[27]. The same SMF-producing genes increase in abundance in metagenomes across the Baltic Sea salinity gradient[18]. How well bioenergetic strategy shifts explain the salinity-associated taxonomic turnover is yet to be determined.

In addition to compartmentalization, protist plankton seem to be more constrained in their geographic ranges than bacteria[28,84]. In general, dispersal limitation increases with the size of a planktonic organism[84]. Thus, smaller bacterial cells reach

new environments more frequently and from further away. In turn, a locally adapting bacterium will, more likely than a locally adapting protist, be outcompeted by a colonizer from a distant yet environmentally similar location. Previous research shows that the same brackish bacterial species more often inhabit distant brackish basins than closely located freshwater or marine waters[19,29]. This might be less common for protists. Our observations that protists display less community structuring by salinity, more frequently cross the salinity barrier, and exhibit increasing species richness with salinity, are all compatible with local adaptation playing a more important role for protistan than bacterial distribution in the region. More direct, genomic studies are needed to test whether local protist adaptations after a shift of salinity regime are indeed more common than bacterial ones.

An important aspect of testing those hypotheses is that ecological sensitivity does not necessarily equal physiological sensitivity in laboratory conditions. An organism that can survive a wide range of salinities in isolation might lose its competitive edge after a minor disruption.

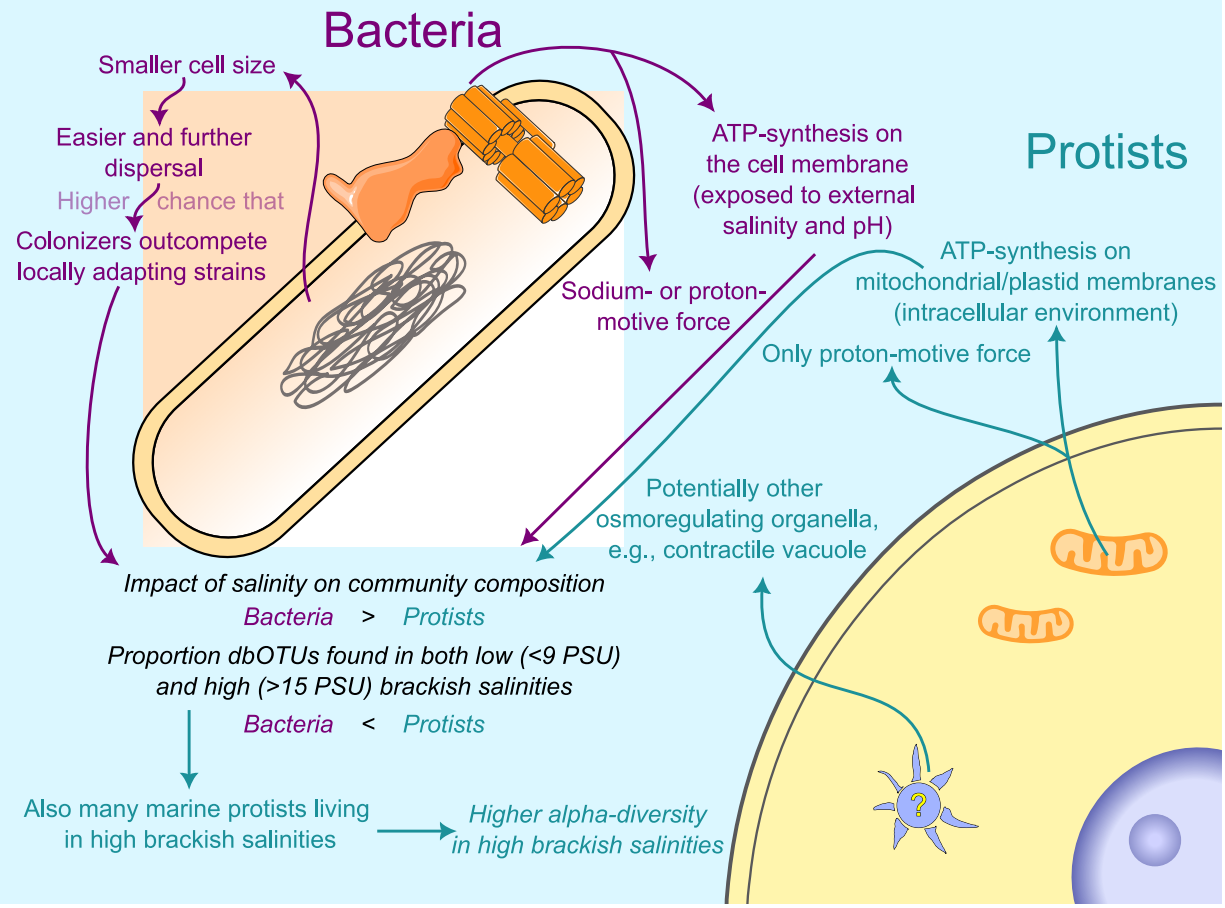


Fig. 7 | The proposed mechanistic explanation for higher ecological sensitivity to salinity for bacteria than for protists. The cells and their components are drawn schematically and do not accurately represent sizes and shapes. All the cell components, except the protist nucleus and bacterial nucleoid, correspond to the machinery responsible for processes indicated by the arrows leading from them. Statements based on our results are written in italics. This figure was created using images from Togo Picture Gallery[85] Servier Medical Art (http://smart.servier.com).

Conclusions:

We show multiple lines of evidence that salinity, generally, has a stronger impact on the distribution of bacteria than protists. Both patterns of overall community composition and the likelihood of crossing a salinity barrier support this notion. This cannot be generalized to stronger environmental filtering among bacteria, as other environmental factors had similar or stronger effects on the protist communities. To explain these patterns, we speculate that compartmentalization allows protists to acclimatize to a different salinity regime more easily (Fig. 7). Additionally, more frequent and distant dispersal of bacteria may hamper local adaptation, as locally adapting strains are outcompeted by colonizers.

Bacterial alpha diversity followed primarily seasonal patterns. We connected those patterns to changes in water column stratification by detecting deepwater taxa at

584 the surface in vertically mixed winter waters. Moreover, the total relative abundance of
585 deepwater taxa reached substantial levels in winter. While seasonal vertical mixing has
586 previously been connected to higher bacterial alpha diversity[33,44,66,67], we provide more
587 conclusive evidence of an influx of deepwater taxa to the surface in winter. The impacts
588 of these community coalescence events on surface water ecosystems require further
589 investigation.
590 In contrast to bacteria, protist alpha diversity primarily followed a geographic
591 pattern, being higher in the more saline Kattegat-Skagerrak. This pattern is analogous
592 to that of metazooplankton and fish, which originates from the limited scope of marine
593 taxa's adaptations to the Baltic Sea conditions following the last glaciation period[11,20].
594 Whether the diverse protists in the Kattegat-Skagerrak represent marine taxa that did
595 not adapt to lower salinities in the Baltic Sea is yet to be confirmed.
596 We show geographic differences not only in microbial community composition
597 but also in seasonal succession dynamics. Thus, time series results from multiple and
598 diverse locations should be integrated for generalizability and proper interpretation. We
599 observed interannual stability of community composition relative to the environmental
600 conditions, consistent with previous results. However, minor technical differences
601 between the datasets confounded alpha diversity analysis. Thus, a standardized DNA-
602 based scheme is needed to monitor rare microbial taxa.
603 Using DNA metabarcoding across strong spatiotemporal gradients, we find new
604 patterns of microbial diversity and expand our understanding of previously observed.
605 Ultimately, as Earth's environments undergo rapid change, diverse microorganisms,
606 often understudied groups, will be impacted and/or consequential for the trajectory of
607 coastal and other ecosystems. DNA-based monitoring covers the phylogenetic scope
608 needed to understand and respond to those unprecedented changes.

609 Methods:

610 Data collection and sequencing

611 The 2019-2020 data comes from Latz et al. 2024[45]. The collection and sequencing
612 procedures are described in detail therein. In brief, samples were collected monthly or
613 every second week, depending on the location, as a part of the Swedish National
614 Marine Monitoring Program. The samples for metabarcoding were collected using a
615 depth-integrating hose, covering 0-10m (except 0-5m at RÅNEÅ-1, and 0-20m at B1
616 and BY31). 500 ml of water from the hose sampling was filtered using 0.22 µm pore
617 size. DNA was extracted using the Zymobiomics™ DNA miniprep kit following a
618 modified protocol[86], and spike-in DNA was added. The DNA was sequenced using
619 Illumina™ MiSeq flow cells. For 16S rDNA gene-based metabarcoding of prokaryotes,
620 primers targeting the hypervariable V3-V4 regions[16], and for 18S metabarcoding of

621 eukaryotes, primers targeting the V4 region[87], were used. Phased primers[45,88] were
622 used for both 18S (on the forward primer) and 16S (on both primers) metabarcoding.
623 Sequences of the primers and the details of the phasing strategy and sample indexing
624 through a second PCR can be found in Latz et al. 2024[45]. Physico-chemical and
625 geographic data has been obtained from the Swedish Meteorology and Hydrology
626 Insitute's service SHARKweb: https://sharkweb.smhi.se/hamta-data/
627 The 2015-2017 data was used for the first time for this study. The samples were
628 stored at -80°C. DNA was extracted in the summer of 2017 and stored at -20°C until
629 February 2023. The sampling and sequencing procedures were analogous to the one
630 for 2019-2020 data, with a few exceptions:
631 1. Volumes varied from 200ml to 800ml. According to our previous tests run in the
632 same setting, volumes ≥200ml should capture the microbial community structure
633 equally well[45].
634 2. DNA extraction was performed using Qiagen™ DNeasy PowerWater Kit, and not
635 ZymoBIOMICS™ DNA Miniprep Kit with a small modification[86] as for the 2019-
636 2020 dataset.
637 3. No spike-in DNA was added.
638 4. Illumina™ NextSeq, rather than MiSeq system, was used.
639 5. For 18S, the initial use of 20 cycles for the first PCR amplification was
640 insufficient, leading to many library preparation failures. It was then re-run with 24
641 cycles, and the results of both runs were merged after processing the tables by
642 summing up the counts in the output count table.
643 6. For 18S, reverse primers were also phased, following a sequence
644 (CTACGA)CTTTCGTTCTTGATYRR, with versions of the primer starting from
645 any base in the brackets included in the primer mix.
646 7. After library preparation, for most samples, 18S and 16S libraries were mixed
647 and sequenced together. A cutadapt[89]-based pipeline was used to separate
648 reads with only 16S or only 18S samples for downstream analysis, as described
649 in the following section.
650 Samples from other projects, which we did not analyze in this study, were sequenced
651 simultaneously and submitted to the European Nucleotide Archive (ENA) all together.

652 Processing of sequencing data

653 The data processing for 2019-2020 samples has been described in detail in Latz et al.
654 2024[45]. In brief, the phased primer sequences were removed from the obtained reads
655 using a modified version of a cutadapt[89]-based pipeline (original pipeline taken from:
656 https://github.com/kjurdzinski/amplicon-multi-cutadapt-18S-16S-monitoring). Only reads
657 containing 16S or 18S primers were kept, which allowed us to separate reads from the
658 two methods when the libraries were mixed. DADA2[90] version 1.18.0 was used to
659 denoise the data, infer amplicon sequence variants (ASVs), and taxonomically annotate

660 them. For taxonomic annotation of 16S ASVs, we used 16S sequences from GTDB[91]
661 version R06-RS202-1 corrected for mislabeled sequences using SATIVA[92,93].
662 Annotation was additionally conducted using the SILVA database[94] version 138.1 to
663 identify plastids and mitochondria. The 18S ASVs were annotated using the PR2
664 database[95] version 5.0.1.
665 The 2015-2017 data was analyzed the same way, with the only difference being
666 that the denoising and ASV inference with DADA2 was performed using the nf-
667 core/ampliseq pipeline[96] version 2.7.0 instead of an in-house R script. We kept the
668 DADA2 parameters as in Latz et al. 2024[45].
669 We further combined the 2015-2017 and 2019-2020 data. We matched ASVs
670 using exact matching. On the merged ASV-sequence list and count table, we ran an
671 additional chimera removal step using UCHIME v1[97] and clustered the remaining ASVs
672 based on their simultaneous genetic and distribution similarity using dbOTU3[46]. Both
673 these steps were performed using an in-house pipeline available at
674 https://github.com/kjurdzinski/chimera_dbOTU_pipeline. Default parameters were used
675 for UCHIME, and for dbOTU2, only the maximum genetic distance for 16S has been
676 changed to 3% (instead of the default 10%). The procedure should minimize the
677 influence of intra-specific variation on our data and further reduce the effects of PCR
678 amplification and sequencing errors. We called thus obtained taxonomic units dbOTUs
679 (distance-based operational taxonomic units, as inferred using dbOTU3[46])
680 For both 16S and 18S, we removed dbOTUs with no annotation at the
681 phylum/supergroup taxonomic level. From 16S dbOTUs, *Archaea*, plastids, and
682 mitochondria were removed. Thus, the remaining ASVs should correspond to bacteria.
683 From 18S ASVs, animals (*Metazoa*), *Fungi*, land plants (*Embryophyceae*), and selected
684 macroalgae (*Phaeophyceae*, and Rhodophyta excluding *Rhodella* and *Cyanidiales*)
685 were removed. Thus, the remaining ASVs were primarily protists.
686 We kept spike-in dbOTUs in a separate table and used them only to obtain spike-
687 in corrected abundance, i.e., the inferred number of rDNA copies per liter. The rDNA
688 copies per liter were calculated according to the equation:

689 $$rDNA\ copies\ per\ liter$$

690 $$= \frac{dbOTU\ counts}{spike - in\ counts}$$

691 $$\times \frac{amount\ of\ spike - in\ added\ [g]}{volume\ water\ sampled\ [l]\ \times\ spike - in\ molar\ mass\ (g/mol]}$$

692 $$\times 6.022 \times 10^{23}$$

693 We have excluded over- and under-sequenced samples to remove the
694 amplification errors (including distorted relative abundances) and not to rarefy to a too
695 low read number during rarefaction. We have excluded three samples with less than
696 30,000 reads for bacteria, as well as one sample with more than 2,000,000 reads.
697 Moreover, we have excluded two samples from 2019-2020 with more than 200,000

698 protist reads (this number of reads was normal for 2015-2017 data since Illumina™
699 NextSeq was used instead of MiSeq, yielding overall higher numbers of reads).
700 Additionally, one sample (20191015_263953_1) with suspected contamination has
701 been excluded.
702 For comparability of results for protists and bacteria, we used only the samples
703 with data available from both 16S and 18S amplicon sequencing (excluding the under-
704 and over-sequenced samples) for all the further analyses. Additionally, in two samples
705 (station RÅNEÅ-2 on 2019-01-29 (sample ID: 20190129_263856_1) and B7 on 2019-
706 09-17 (sample ID: 20190917_264450_1)) 18S spike-in addition failed. Those samples
707 were excluded from analyses in which spike-in normalization was used, except for
708 bacteria-only focused analyses in Fig. 1b.

709 Abundance, alpha and beta diversity spatiotemporal patterns

710 For the spatiotemporal analysis of abundance, beta- and alpha diversity across the
711 samples from 2019-2020, we removed the samples from stations H4, BY29 / LL 19, and
712 SR3. Those stations were sampled at only a few unevenly distributed occasions, and
713 their exclusion/inclusion at certain seasons could obscure seasonal patterns.
714 All the statistical analyses and data visualization was performed using R version
715 4.2.2[98]. Inkscape[99] was used to assemble and edit multi-panel figures, and to make Fig.
716 7.
717 The multivariate analyses were performed using vegan R package[100], and all the
718 functions in this paragraph can be found there unless another package is explicitly
719 stated. Bray–Curtis distances between communities were obtained using rarefaction via
720 the avgdist function. Those distances were used for Principal Coordinate Analysis
721 (PCoA) and distance-based Redundancy Analysis (dbRDA), performed with,
722 respectively, capscale (without setting constraints) and dbrda functions. Variance
723 partitioning was performed using rdacca.hp R package[101]. Families were positioned on
724 the ordination plots by calculating weighted averages (sppscores function). Only the
725 families deviating (Bonferroni corrected P-value < 0.05) from normal distribution around
726 the center of the dbRDA coordinate system were shown on the plots. To find the
727 corresponding two-dimensional ellipsoid confidence intervals, aspace[102] package was
728 used. Whenever only samples were visualized, their ordination scores were scaled
729 using the eigenvalues, and symmetric scaling was used if both the families and the
730 samples were visualized (using, respectively, “sites” and “symmetric” scaling options of
731 the plotting functions respectively). For dbRDA and variance partitioning, we used a
732 subset of 228 (out of 246) samples with all the environmental variables of interest
733 available. For more informative visualization, to keep the sample structuring visible
734 despite very high family scores, we divided the family scores in half.

735 dbOTU richness was calculated as the mean number of dbOTUs detected in a
736 sample over 100 rarefaction iterations. Variance partitioning was performed using
737 rdacca.hp R package[101].
738 For both beta and alpha diversity estimation, in each rarefaction iteration, we
739 rarefied the bacterial 16S reads to 34382 per sample, and the protists 18S reads to
740 8084 per sample.
741 For deepwater taxa detection, we used the representative sequences of 99%
742 average nucleotide identity OTUs (operational taxonomic units) detected in samples
743 from the summer of 2008 and coming from a previous study[33]. The 2008 summer
744 transect included sampling across different depths, and the deepwater OTUs in this
745 study were present exclusively at depths ranging from 11 to 300m, classified as
746 mesopelagic in the original study[33]. We connected the OTUs from the original, transect-
747 based study to ASVs from the monitoring data used here, using the 99% nucleotide
748 identity threshold in VSEARCH[68]. We deemed an OTU and a dbOTU to be matching if
749 their representative sequences (most abundant ASV for a dbOTU) matched in the
750 VSEARCH output. For comparing the number of OTUs from deepwater, surface water,
751 and both (Fig. 5a) in winter and summer we used the chi-squared test as implemented
752 in the prop.test function in R[103].

753 Proportion salinity barrier crossing dbOTUs and Bayesian modelling

754 Any dbOTU found among included samples with salinity >15 PSU and among
755 samples with salinity <9 PSU was regarded as barrier crossing. For balancing out the
756 number of samples on both sides of the salinity barrier, apart from the four stations with
757 salinity >15 PSU (Å17, SLÄGGÖ), four out of twelve stations with salinity <9 PSU: BY2
758 ARKONA, BY5 BORNHOLMSDJ, BY15 GOTLANDSDJ, REF M1V1 (see Fig. 1a for
759 locations of the stations). First, we chose the four most saline stations from among
760 those with salinity <9 PSU. Then, we switched BCS III-10 to REF M1V1 (see Fig. 1),
761 since the latter had more samples, and, unlike other chosen stations from lower
762 salinities, was located close to the mainland coastline. For each station but the least
763 sampled (REF M1V1), the samples were downsampled, choosing samples closest in
764 time to the ones collected at the least sampled stations. Whenever available, those
765 samples should correspond to the same monitoring cruises.
766 The Bayesian models were built and implemented in JAGS [74], using the R
767 package rjagS10[104]. The parameters, probabilities, and predictions were calculated
768 based on every fifth out of 10000 MCMC iterations.
769 The model for probability of crossing the salinity barrier was optimized based on
770 Bernoulli distribution:
771 $$if_crossed_barrier_i \sim Bernoulli(p_i)$$
772 where for each dbOTU *i*, $if_crossed_barrier_i$ is a binary variable = 1 if it is a barrier
773 crossing dbOTU. p_i was calculated according to the formula

$$logit(p_i) = (1 - if_protist_i) \times (\alpha_{bacterium} + \beta_{abundance_bacterium} \times log(mean_abundance_i) + \beta_{occurence_bacterium} \times log(max_occurence_one_side_i)) + if_protist_i \times (\alpha_{protist} + \beta_{abundance_protist} \times log(mean_abundance_i) + \beta_{occurence_protist} \times log(max_occurence_one_side_i))$$

where $if_protist_i$ is a binary variable = 1 if the dbOTU *i* is a protist. $mean_abundance_i$ is geometric mean of the abundance of the dbOTU *i* in the samples where it was present. $max_occurence_one_side_i$ 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.

We compared this model with one assuming the same abundance and occupancy dependence for bacteria and for protists (one $\beta_{abundance}$ and $\beta_{occupancy}$, same for protists and bacteria, different $\alpha_{bacterium}$ and $\alpha_{protist}$). For both the model with different and the same abundance and occupancy dependencies, we also compared the versions without log-transforming either only the occupancy, or both abundance and occupancy values. Based on the Watanabe-Akaike criterion[75], the model we used performed the best.

The model for total occupancy was build 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:

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

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.

Acknowledgements:
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 Garefeld 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. Mathew Low and Malin Aronsson for help with Bayesian modeling.

Study funding information:

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. AFA was additionally funded by a research grant (2021-05563) from the Swedish Research Council VR. 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).

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).

Competing interests:

The authors declare no conflict of interest.

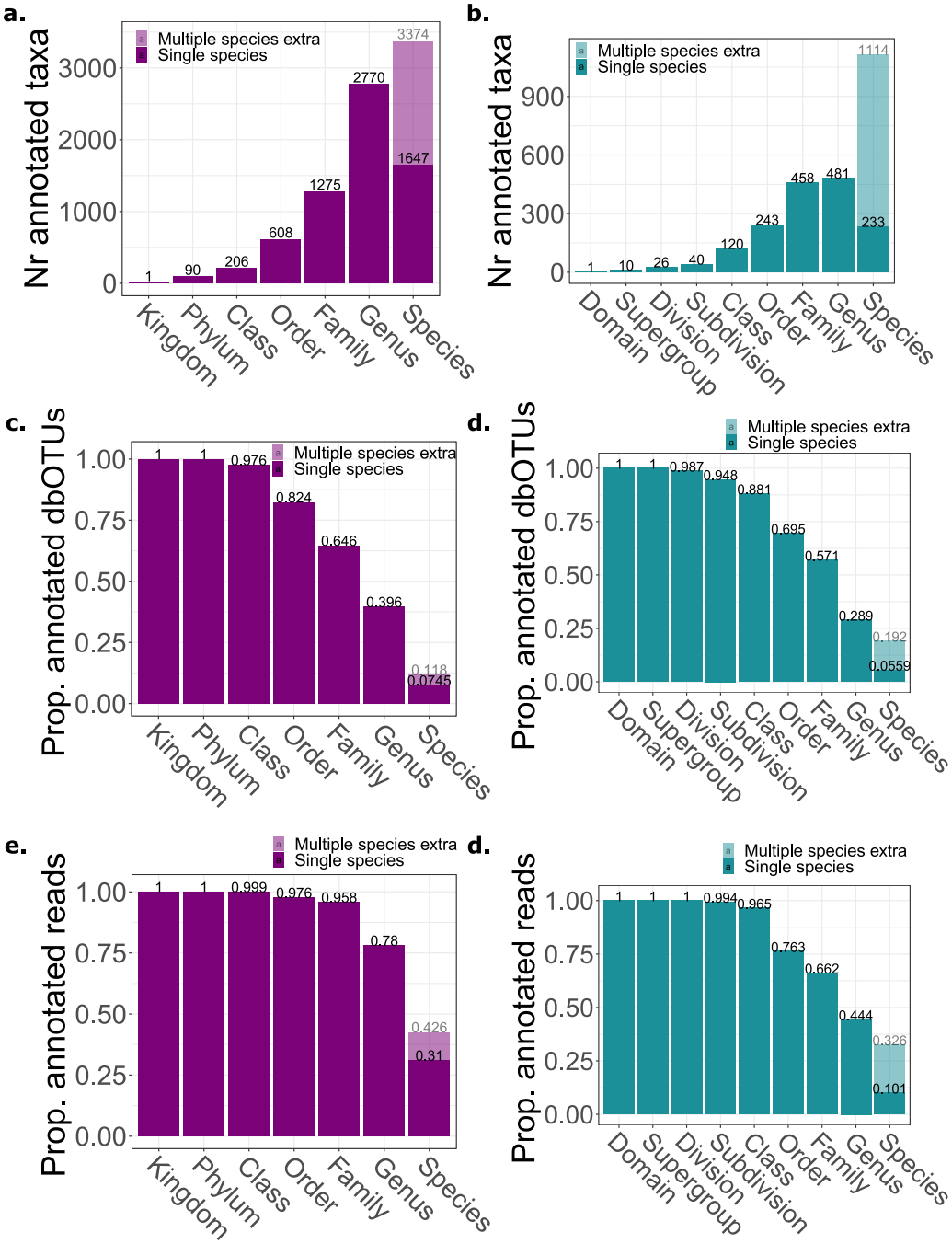
Author contributions:

KTJ analyzed the data under the supervision of AFA and with help from MACL and AT. YOOH prepared the DNA for the 2015-2017 dataset. KTJ wrote the first version of the manuscript under supervision of AFA. All authors contributed to conceptualization, interpretation of the data, revised the manuscript, and approved the final version.

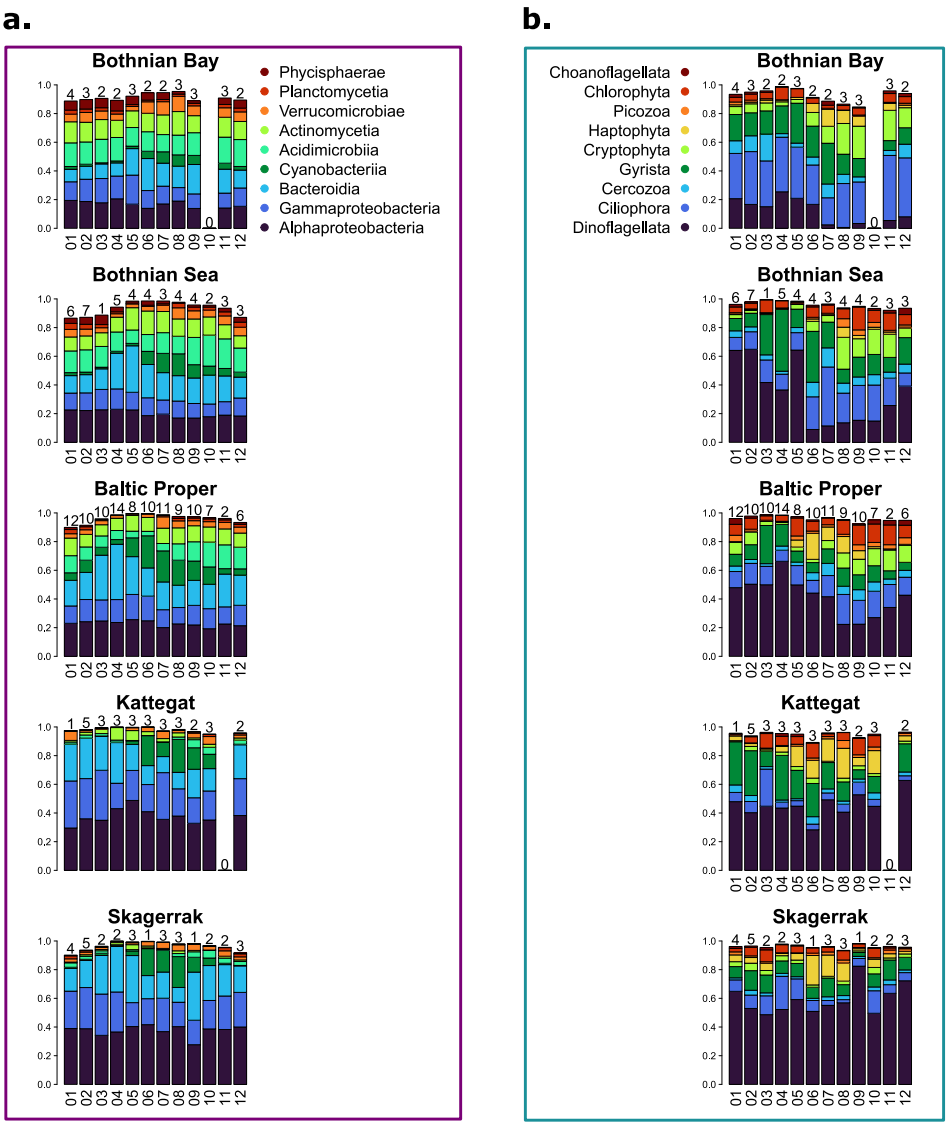
Data and code availability:

The 2019-2020 data have been published before as a data paper[45]. 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 2015-2019 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). The code used to analyze the data and generate the figures, together with intermediate processing output, can be found at https://doi.org/_/scilifelab.

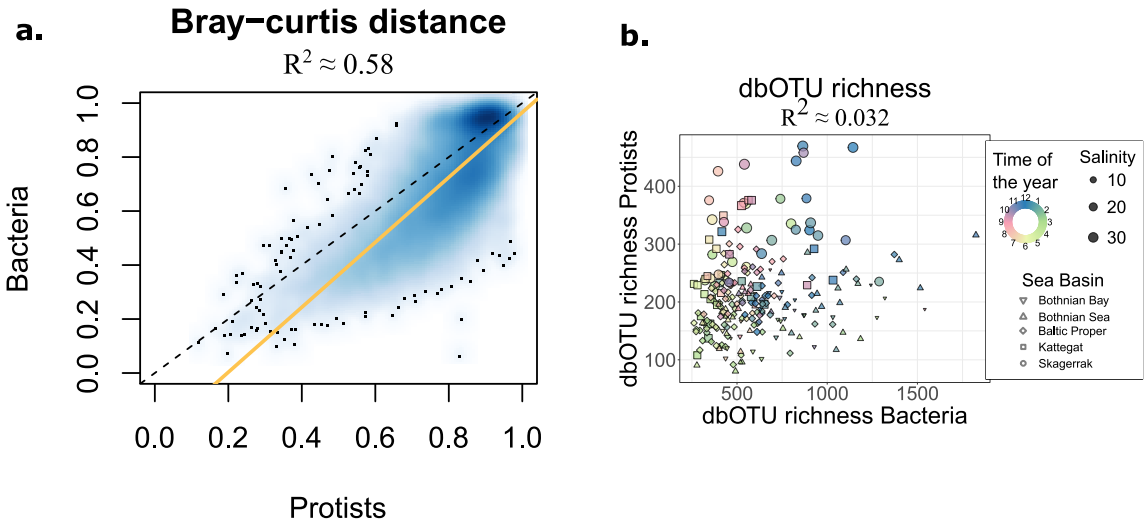
Supplementary figures:



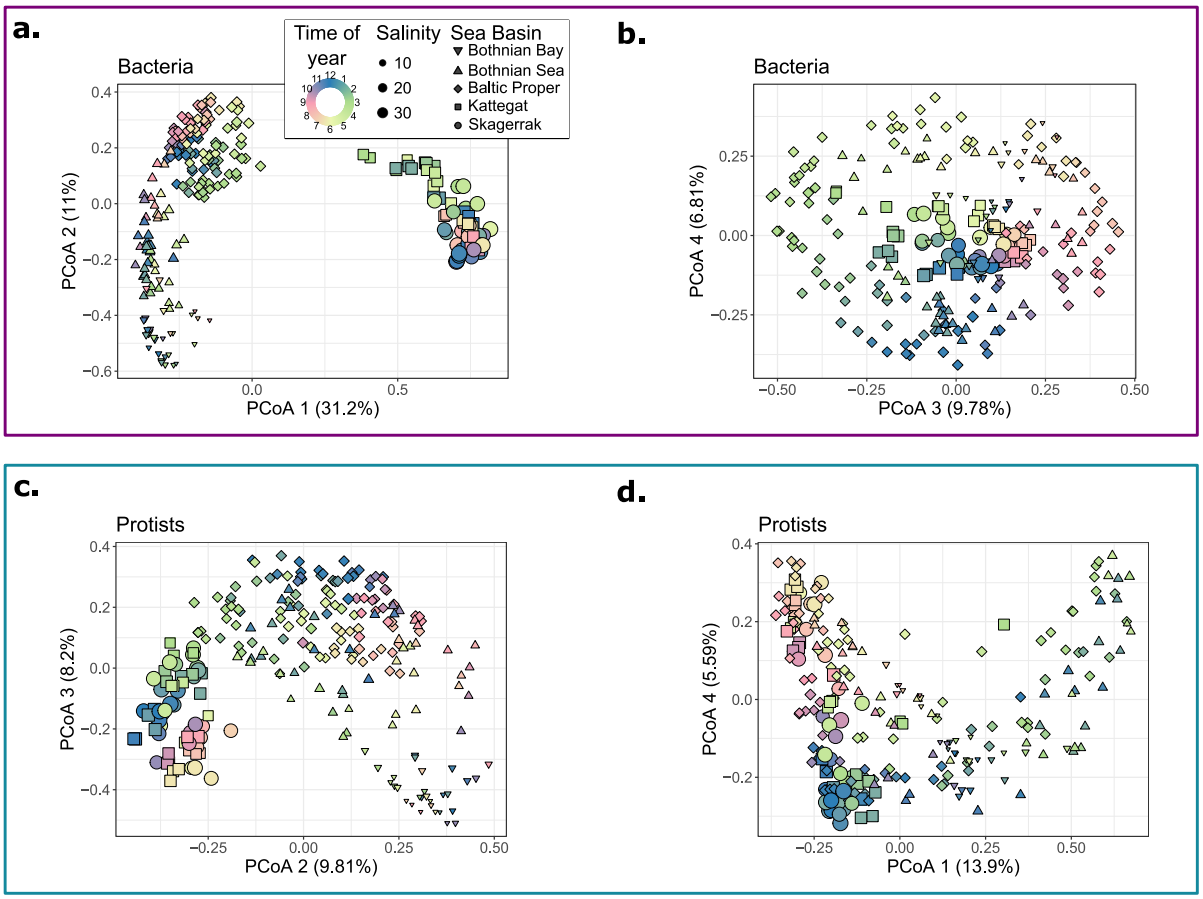
Supplementary Fig. S1 | Taxonomic annotation statistics a-b. Number of different bacterial (**a.**) and protist (**b.**) taxa found at different taxonomic levels. **c-d**. Proportions of bacterial (**c.**) and protist (**d.**) taxa annotated at different taxonomic levels. **e-f**. Proportions of bacterial (**e.**) and protist (**f.**) reads annotated at different taxonomic levels. “Single species” corresponds to an analysis that provides a species-level annotation only if the query sequence has identical matches to database sequences of a single species. “Multiple species extra” is the extra number/proportion of annotated species when exact matches to multiple species are allowed.



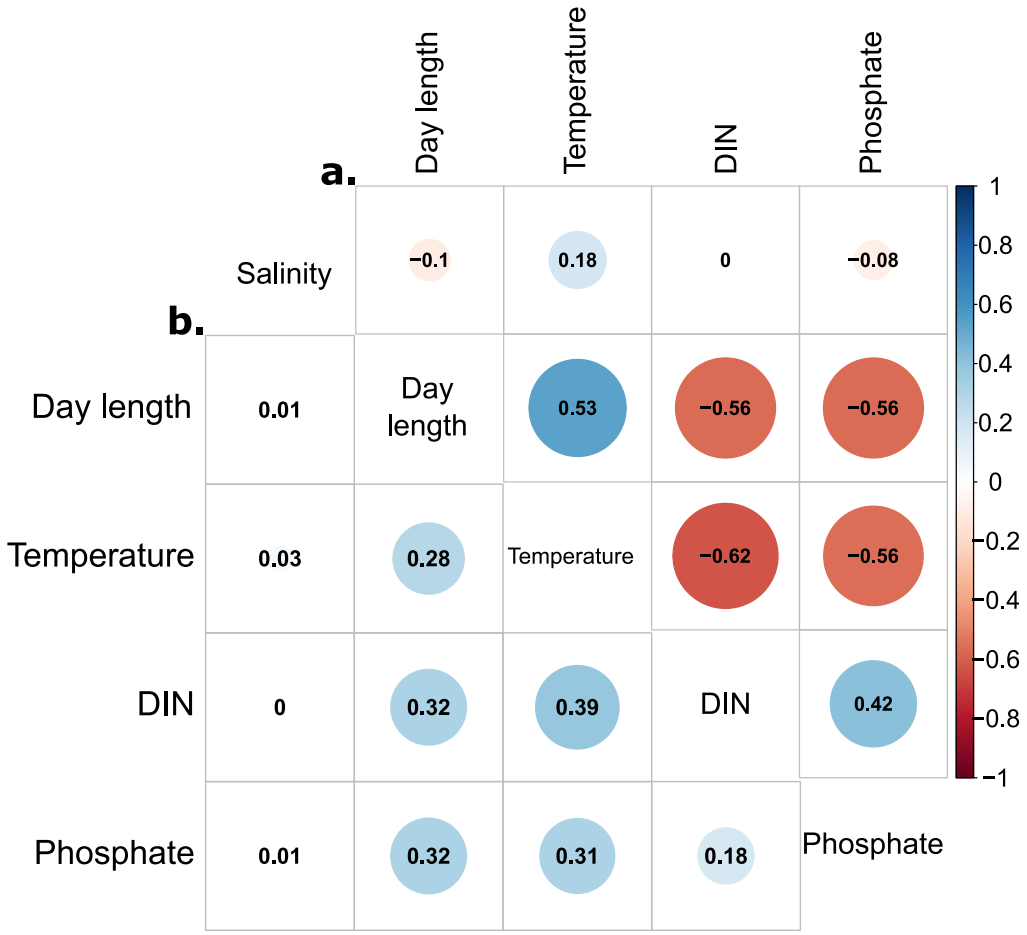
Supplementary Fig. S2 | Spatiotemporal changes in abundance of major microbial taxa, 2019-2020 a-b. The proportion of rDNA marker gene reads annotated to **a.** bacterial classes and **b.** protist subdivisions averaged over each month and Baltic Sea area basin. Only the classes/subdivisions which, on average, corresponded to >0.01 of reads across samples are shown. Above each bar, the number of samples collected in the respective month and basin is given.



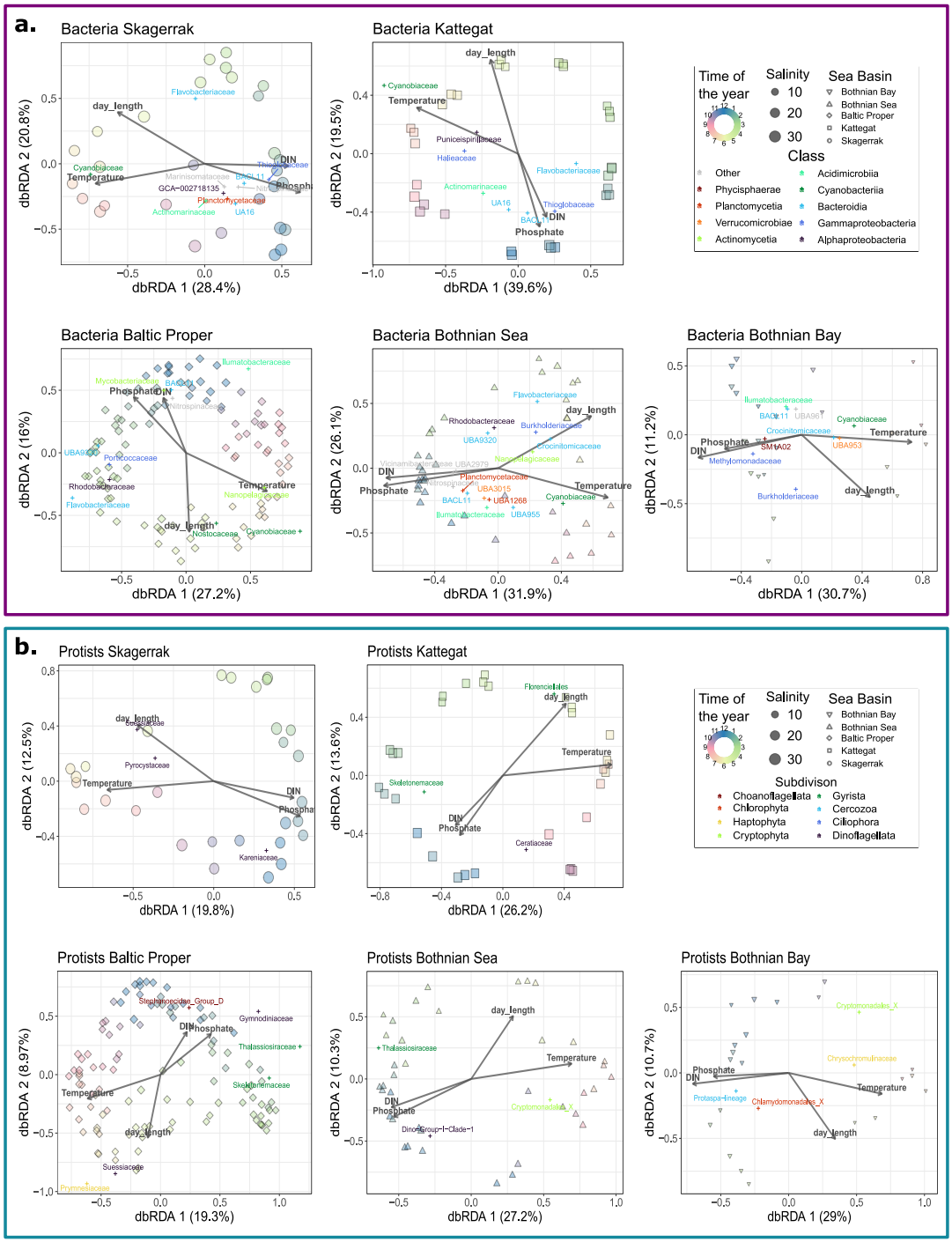
Supplementary Fig. S3 | Dissimilarity of diversity metrics between bacteria and protist, 2019-2020 a. Beta diversity calculated as Bray-Curtis distances. The golden line represents the best linear fit, while the dashed black line is the identity line. **b.** Number of detected dbOTUs (alpha diversity).



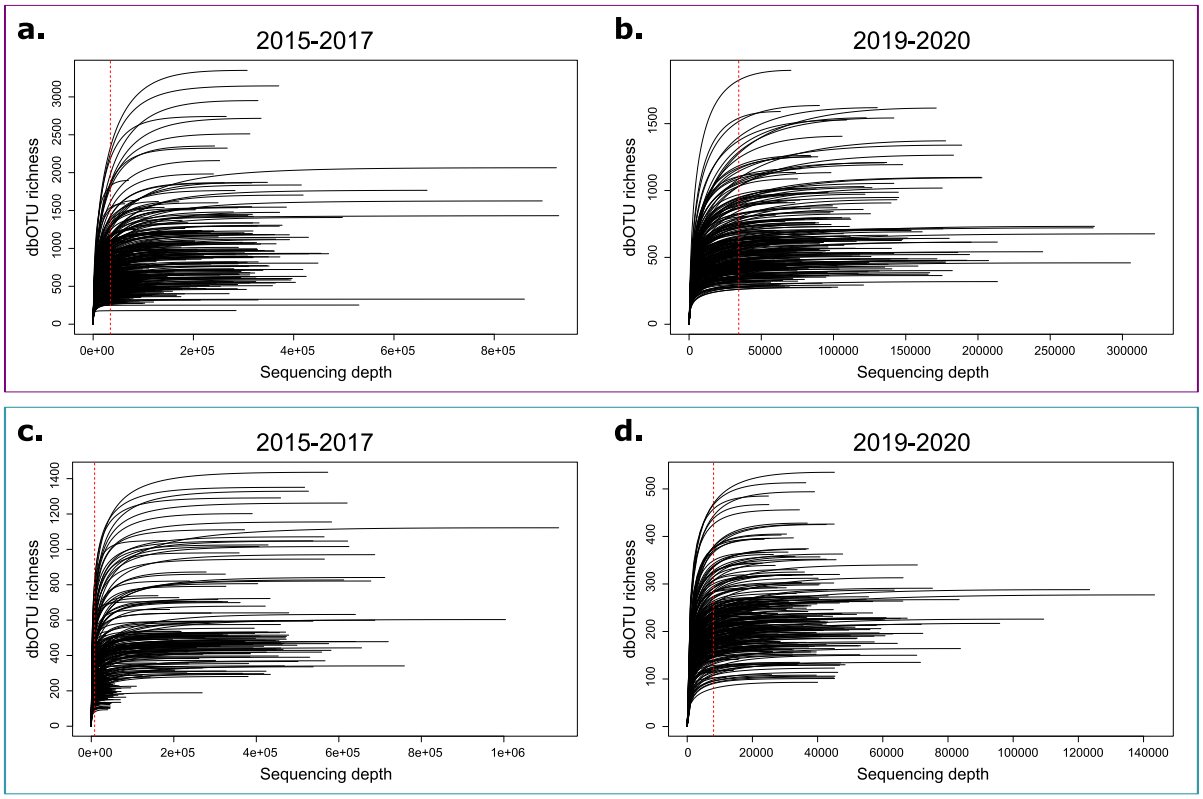
Supplementary Fig. S4 | Principal coordinate analysis (PCoA) 2019-2020. The first four ordination axes are based on Bray-Curtis distances between bacterial (**a-b**) and protist (**c-d**) communities. Note: The axes are grouped and ordered by which ones, after visual examination, correspond to geographic structuring (**a**, **c)** and which ones to seasonal effects (**b**, **d**). This ordering corresponded to showing ordination axes with decreasing variation explained for bacteria (**a-b**). However, for protists, the first plot (**c**) shows the second and the third components, while the second plot (**d**) shows the first and the fourth components.



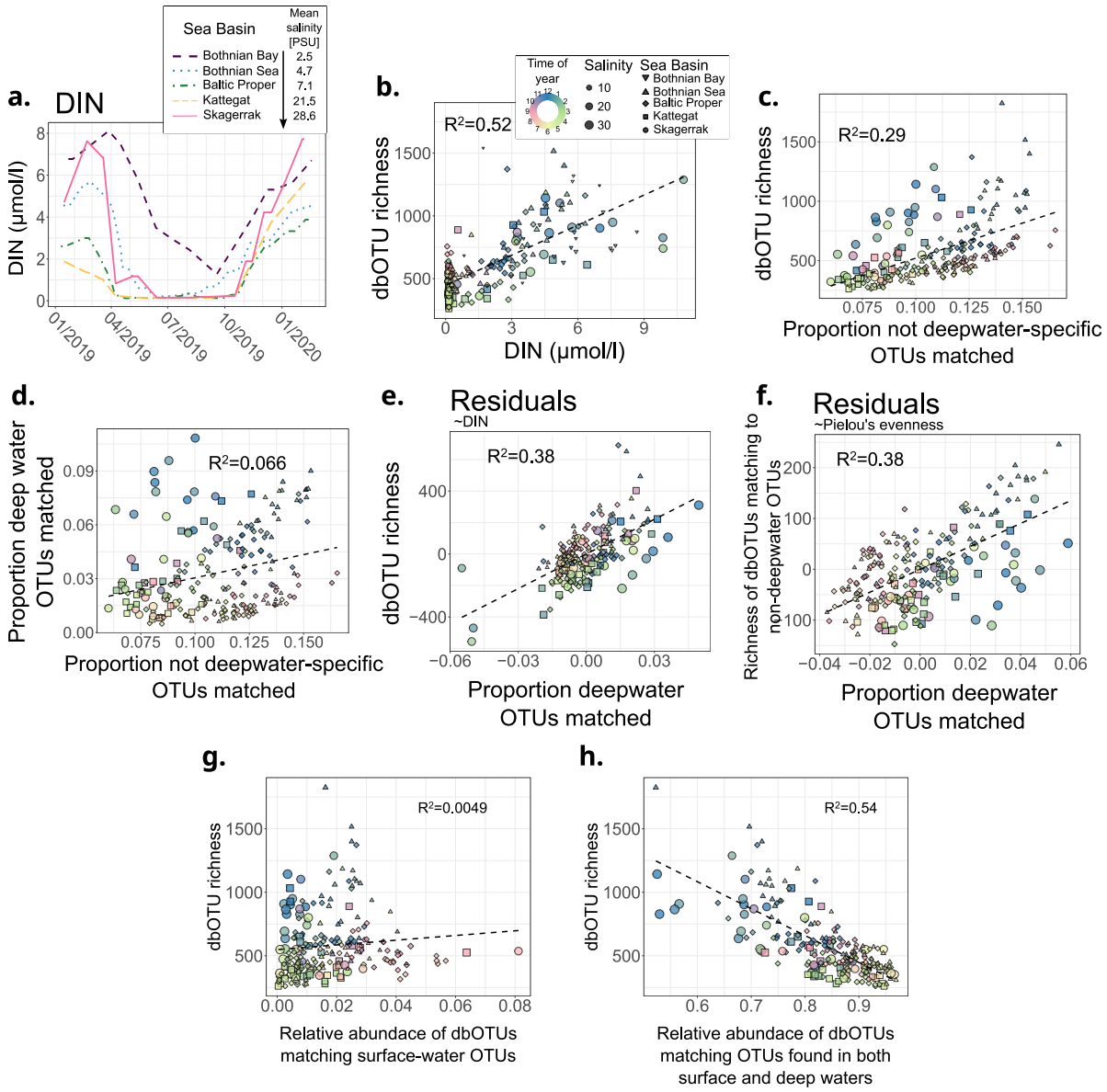
Supplementary Fig. S5 | Correlations between chosen environmental variables for the 2019-2020 dataset a. Pearson correlation coefficients (ρ) on an upper-triangle correlation plot. **b.** Coefficients of determination (R^2) on a lower-triangle correlation plot.



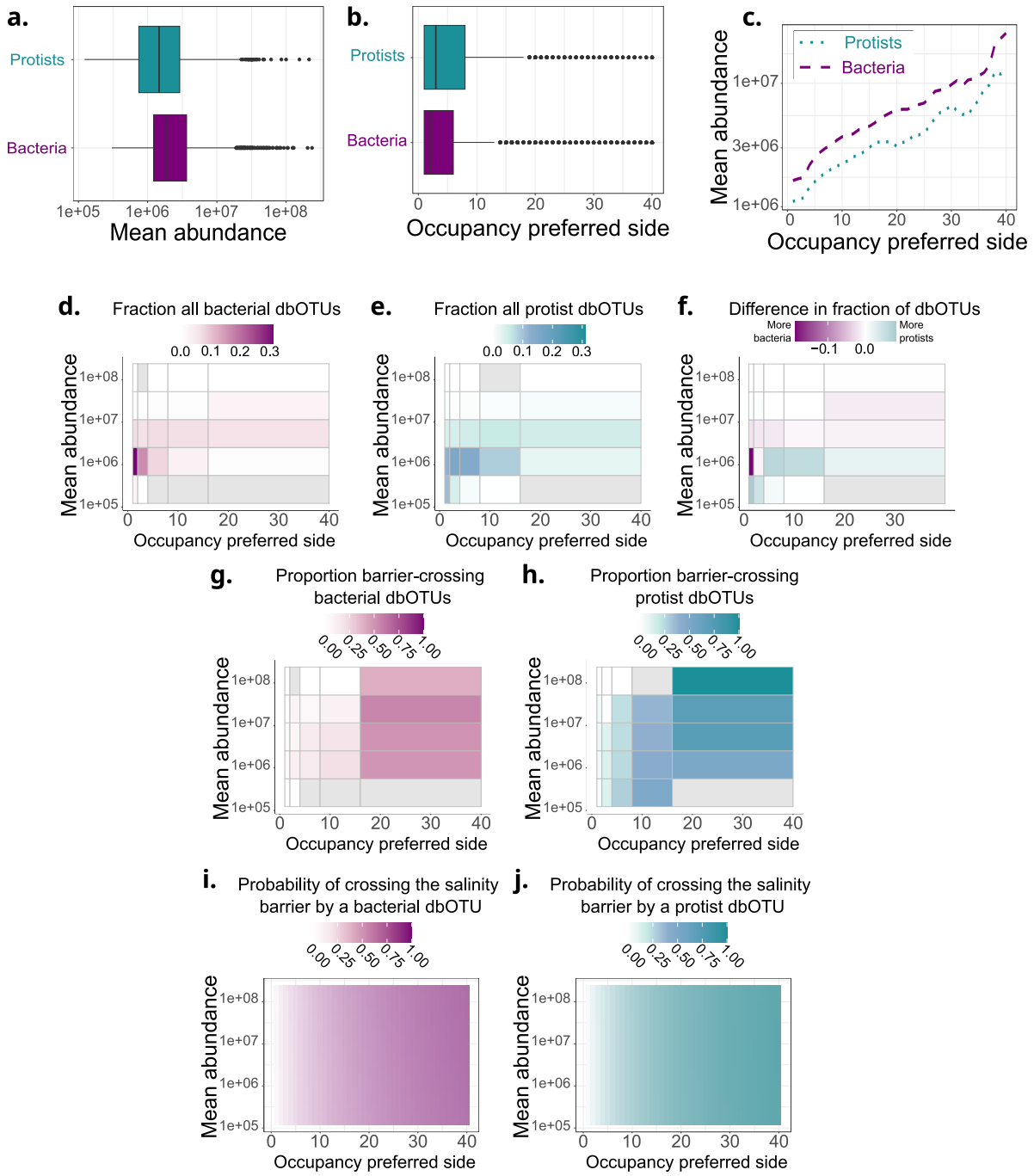
Supplementary Fig. S6 | Seasonal changes in community composition within distinguished basins of the Baltic Sea area. distance-based Redundancy Analysis (dbRDA) based on Bray-Curtis distances between the bacterial (**a.**) and protist (**b.**) communities within each of the distinguished basins of the Baltic Sea area. Weighted averages of families deviating (Bonferroni corrected P < 0.05) were placed on the plot, with values rescaled by a factor of 0.5. Percentages of variation explained by each dbRDA component are given in brackets by the axes labels.



Supplementary Fig. S7 | Refraction curves a-b. Based on 16S filtered (only bacteria) read from the 2015-2017 (**a.**) and 2019-2020 (**b.**) datests. **c-d**. Based on 18S filtered (only protists) read from the 2015-2017 (**c.**) and 2019-2020 (**d.**) datests. dbOTU - distribution-based operational taxonomic unit.



Supplementary Fig S8 | Bacterial alpha diversity a. A rolling mean of DIN values, averaged for measurements two weeks before and after each date. **b**. DIN vs. dbOTU richness across samples. **c**. Proportion of all non-deepwater-specific OTUs ("surface" and "both" categories) having a match to a dbOTU from a sample vs. bacterial dbOTU richness. **c**. Proportion of all non-deepwater-specific OTUs vs. proportion of all deepwater OTUs. **e.** The residuals of proportion of all deepwater OTUs vs. dbOTU richness from models correlating the factors to DIN, corresponding to partitioning out the unvisualized parameter. **f. e.** The residuals of proportion of all deepwater OTUs vs. dbOTU richness from models correlating the factors to Pielou's evenness. **g.** Relative abundance of all dbOTUs matching a surface water OTU vs bacterial dbOTU richness. **e.** Relative abundance of all dbOTUs matching an OTU found in both surface and deep waters vs bacterial dbOTU richness. Bothnian Bay was excluded from plots c-g, since there were no deepwater samples taken in this region in the transect-based study[33]. The same legend applies to figures b-g. DIN - dissolved inorganic nitrogen.



Supplementary Fig. S9 | The fractions of all dbOTUs, the proportions of barrier-crossing dbOTUs, and the probability of crossing the barrier across occupancy and abundance values. a-b. The distribution of mean abundance (**a.**) and occupancy at the preferred side of the salinity barrier (the side where the dbOTU has higher occupancy) (**b.**). **c.** Rolling mean (+/- two bins) of abundance across the occupancy values. **d-f.** The fraction of all bacterial (**d.**) and protist (**e.**) dbOTUs, and the difference between the fraction of protist and bacterial dbOTUs (**f.**). **g-h**. The proportion of barrier crossing bacterial (**g.**) and protist (**h.**) dbOTUs. **d-h.** Grey tiles correspond to the values across which either no protist or no bacterial dbOTUs were found. **I-j.**

The predicted probabilities of a bacterial (**i.**) and protist (**j.**) dbOTU crossing the salinity barrier. For all the analyses, only the four stations with salinity >15 PSU, and four out of the most saline stations with salinity <9 PSU are chosen. The data for each station is downsampled to the same number of reads. During downsampling, observations closest in time to the samples from the least sampled station were chosen. Only observations from 2019-2020 are included. The abundance values correspond to the numbers of rDNA copies per liter, based on the spike-in normalization, and are always given on a logarithmic scale. “Occupancy preferred side” is the maximum occupancy at one of the two sides of the salinity barrier.

References:

1. Falkowski, P. G., Fenchel, T. & Delong, E. F. The microbial engines that drive Earth’s biogeochemical cycles. *Science* **320**, 1034–1039 (2008).
2. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).
3. Mathis, M. *et al.* Enhanced CO2 uptake of the coastal ocean is dominated by biological carbon fixation. *Nat. Clim. Chang.* **14**, 373–379 (2024).
4. Cohen, J. E., Small, C., Mellinger, A., Gallup, J. & Sachs, J. Estimates of Coastal Populations. *Science* **278**, 1209–1213 (1997).
5. Williams, B. A. *et al.* Global rarity of intact coastal regions. *Conserv. Biol.* **36**, e13874 (2022).
6. De Valck, J. *et al.* Valuing ecosystem services in complex coastal settings: An extended ecosystem accounting framework for improved decision-making. *Mar. Policy* **155**, 105761 (2023).
7. O’Hara, C. C., Frazier, M. & Halpern, B. S. At-risk marine biodiversity faces extensive, expanding, and intensifying human impacts. *Science* **372**, 84–87 (2021).
8. Ray, G. C. Coastal-Zone Biodiversity Patterns: Principles of landscape ecology may help explain the processes underlying coastal diversity. *Bioscience* **41**, 490–498 (1991).
9. Kuliński, K. *et al.* Biogeochemical functioning of the Baltic Sea. *Earth Syst. Dyn.* **13**, 633–685 (2022).

10. Björck, S. A review of the history of the Baltic Sea, 13.0-8.0 ka BP. *Quat. Int.* **27**, 19–40 (1995).
11. Johannesson, K., Le Moan, A., Perini, S. & André, C. A Darwinian Laboratory of Multiple Contact Zones. *Trends Ecol. Evol.* **35**, 1021–1036 (2020).
12. Reusch, T. B. H. *et al.* The Baltic Sea as a time machine for the future coastal ocean. *Sci Adv* **4**, eaar8195 (2018).
13. 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**, (2024).
14. 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).
15. 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).
16. 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).
17. 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).
18. Dupont, C. L. *et al.* Functional tradeoffs underpin salinity-driven divergence in microbial community composition. *PLoS One* **9**, e89549 (2014).
19. Hugerth, L. W. *et al.* Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biol.* **16**, 279 (2015).
20. 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).
21. Wu, Q. L., Zwart, G., Schauer, M., Kamst-van Agterveld, M. P. & Hahn, M. W.

963 Bacterioplankton community composition along a salinity gradient of sixteen high-mountain
964 lakes located on the Tibetan Plateau, China. *Appl. Environ. Microbiol.* **72**, 5478–5485
965 (2006).
966 22. Lozupone, C. A. & Knight, R. Global patterns in bacterial diversity. *Proc. Natl. Acad. Sci. U.*
967 *S. A.* **104**, 11436–11440 (2007).
968 23. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of
969 sequences per sample. *Proc. Natl. Acad. Sci. U. S. A.* **108 Suppl 1**, 4516–4522 (2011).
970 24. Thompson, L. R. *et al.* A communal catalogue reveals Earth's multiscale microbial diversity.
971 *Nature* **551**, 457–463 (2017).
972 25. Singer, D. *et al.* Protist taxonomic and functional diversity in soil, freshwater and marine
973 ecosystems. *Environ. Int.* **146**, 106262 (2021).
974 26. Logares, R. *et al.* Infrequent marine-freshwater transitions in the microbial world. *Trends*
975 *Microbiol.* **17**, 414–422 (2009).
976 27. Walsh, D. A., Lafontaine, J. & Grossart, H.-P. On the Eco-Evolutionary Relationships of
977 Fresh and Salt Water Bacteria and the Role of Gene Transfer in Their Adaptation. in *Lateral*
978 *Gene Transfer in Evolution* (ed. Gophna, U.) 55–77 (Springer New York, New York, NY,
979 2013).
980 28. Burki, F., Sandin, M. M. & Jamy, M. Diversity and ecology of protists revealed by
981 metabarcoding. *Curr. Biol.* **31**, R1267–R1280 (2021).
982 29. Jurdzinski, K. T. *et al.* Large-scale phylogenomics of aquatic bacteria reveal molecular
983 mechanisms for adaptation to salinity. *Sci Adv* **9**, eadg2059 (2023).
984 30. Telesh, I. V., Schubert, H. & Skarlato, S. O. Revisiting Remane's concept: evidence for high
985 plankton diversity and a protistan species maximum in the horohalinicum of the Baltic Sea.
986 *Mar. Ecol. Prog. Ser.* **421**, 1–11 (2011).
987 31. Olli, K., Ptacnik, R., Klais, R. & Tamminen, T. Phytoplankton Species Richness along
988 Coastal and Estuarine Salinity Continua. *Am. Nat.* **194**, E41–E51 (2019).

989 32. Olofsson, M., Hagan, J. G., Karlson, B. & Gamfeldt, L. Large seasonal and spatial variation
990 in nano- and microphytoplankton diversity along a Baltic Sea-North Sea salinity gradient.
991 *Sci. Rep.* **10**, 17666 (2020).
992 33. Herlemann, D. P. R., Lundin, D., Andersson, A. F., Labrenz, M. & Jürgens, K. Phylogenetic
993 Signals of Salinity and Season in Bacterial Community Composition Across the Salinity
994 Gradient of the Baltic Sea. *Front. Microbiol.* **7**, 1883 (2016).
995 34. Torstensson, A. *et al.* Physicochemical control of bacterial and protist community
996 composition and diversity in Antarctic sea ice: Physicochemical control of sea ice microbial
997 communities. *Environ. Microbiol.* **17**, 3869–3881 (2015).
998 35. Logares, R. *et al.* Contrasting prevalence of selection and drift in the community structuring
999 of bacteria and microbial eukaryotes. *Environ. Microbiol.* **20**, 2231–2240 (2018).
1000 36. Soininen, J., Korhonen, J. J., Karhu, J. & Vetterli, A. Disentangling the spatial patterns in
1001 community composition of prokaryotic and eukaryotic lake plankton. *Limnol. Oceanogr.* **56**,
1002 508–520 (2011).
1003 37. De Bie, T. *et al.* Body size and dispersal mode as key traits determining metacommunity
1004 structure of aquatic organisms. *Ecol. Lett.* **15**, 740–747 (2012).
1005 38. Wu, W. *et al.* Contrasting the relative importance of species sorting and dispersal limitation
1006 in shaping marine bacterial versus protist communities. *ISME J.* **12**, 485–494 (2018).
1007 39. Yeh, Y.-C. & Fuhrman, J. A. Contrasting diversity patterns of prokaryotes and protists over
1008 time and depth at the San-Pedro Ocean Time series. *ISME Commun* **2**, 36 (2022).
1009 40. Remane, A. Die Brackwasserfauna : mit besonderer Berücksichtigung der Ostsee. *Zool.*
1010 *Anz.* **Suppl. 7**, (1934).
1011 41. Lindh, M. V. *et al.* Disentangling seasonal bacterioplankton population dynamics by high-
1012 frequency sampling. *Environ. Microbiol.* **17**, 2459–2476 (2015).
1013 42. Gilbert, J. A. *et al.* Defining seasonal marine microbial community dynamics. *ISME J.* **6**,
1014 298–308 (2012).

1015 43. Ladau, J. *et al.* Global marine bacterial diversity peaks at high latitudes in winter. *ISME J.* **7**,
1016 1669–1677 (2013).
1017 44. Raes, E. J. *et al.* Seasonal patterns of microbial diversity across the world oceans. *Limnol.*
1018 *Oceanogr. Lett.* (2024) doi:10.1002/lol2.10422.
1019 45. Latz, M. A. C. *et al.* A comprehensive dataset on spatiotemporal variation of microbial
1020 plankton communities in the Baltic Sea. *Sci Data* **11**, 18 (2024).
1021 46. Olesen, S. W., Duvallet, C. & Alm, E. J. dbOTU3: A new implementation of distribution-
1022 based OTU calling. *PLoS One* **12**, e0176335 (2017).
1023 47. Tourlousse, D. M. *et al.* Synthetic spike-in standards for high-throughput 16S rRNA gene
1024 amplicon sequencing. *Nucleic Acids Res.* **45**, e23 (2017).
1025 48. Lloréns-Rico, V., Vieira-Silva, S., Gonçalves, P. J., Falony, G. & Raes, J. Benchmarking
1026 microbiome transformations favors experimental quantitative approaches to address
1027 compositionality and sampling depth biases. *Nat. Commun.* **12**, 3562 (2021).
1028 49. Wesslander, K., Viktorsson, L., Thor, P., Nilsson, M. & Skjevik, A.-T. *The Swedish National*
1029 *Marine Monitoring Programme 2019 : Hydrography Nutrients Phytoplankton*. 59 (2020).
1030 50. Zettler, M. L., Kremp, A. & Dutz, J. Biological assessment of the Baltic Sea 2019.
1031 *Meereswiss. Ber. , Warnemünde* **115**, (2020).
1032 51. Fridolfsson, E. *et al.* Multiyear analysis uncovers coordinated seasonality in stocks and
1033 composition of the planktonic food web in the Baltic Sea proper. *Sci. Rep.* **13**, 11865
1034 (2023).
1035 52. Logares, R. *et al.* Disentangling the mechanisms shaping the surface ocean microbiota.
1036 *Microbiome* **8**, 55 (2020).
1037 53. Farjalla, V. F. *et al.* Ecological determinism increases with organism size. *Ecology* **93**,
1038 1752–1759 (2012).
1039 54. Neuenschwander, S. M., Ghai, R., Pernthaler, J. & Salcher, M. M. Microdiversification in
1040 genome-streamlined ubiquitous freshwater Actinobacteria. *ISME J.* **12**, 185–198 (2018).

55. Mehrshad, M. *et al.* Hidden in plain sight—highly abundant and diverse planktonic freshwater Chloroflexi. *Microbiome* **6**, 176 (2018).
56. 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).
57. Morris, R. M. & Spietz, R. L. The Physiology and Biogeochemistry of SUP05. *Ann. Rev. Mar. Sci.* **14**, 261–275 (2022).
58. 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**, (2018).
59. 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).
60. Hjerne, O., Hajdu, S., Larsson, U., Downing, A. S. & Winder, M. Climate driven changes in timing, composition and magnitude of the Baltic Sea phytoplankton spring bloom. *Front. Mar. Sci.* **6**, (2019).
61. Masson-Delmotte, V. *Climate Change 2021: The Physical Science Basis : Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. (Cambridge University Press, 2021).
62. HELCOM. Baltic Sea Action Plan – 2021 update. Preprint at (2021).
63. Rohwer, R. R., Hale, R. J., Vander Zanden, M. J., Miller, T. R. & McMahon, K. D. Species invasions shift microbial phenology in a two-decade freshwater time series. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2211796120 (2023).
64. Schloss, P. D. Rarefaction is currently the best approach to control for uneven sequencing effort in amplicon sequence analyses. *mSphere* **9**, e0035423 (2024).
65. Reissmann, J. H. *et al.* Vertical mixing in the Baltic Sea and consequences for eutrophication – A review. *Prog. Oceanogr.* **82**, 47–80 (2009).

1067 66. García, F. C., Alonso-Sáez, L., Morán, X. A. G. & López-Urrutia, Á. Seasonality in
1068 molecular and cytometric diversity of marine bacterioplankton: the re-shuffling of bacterial
1069 taxa by vertical mixing. *Environ. Microbiol.* **17**, 4133–4142 (2015).
1070 67. Alonso-Sáez, L., Díaz-Pérez, L. & Morán, X. A. G. The hidden seasonality of the rare
1071 biosphere in coastal marine bacterioplankton: Seasonality of the rare biosphere. *Environ.*
1072 *Microbiol.* **17**, 3766–3780 (2015).
1073 68. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open
1074 source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
1075 69. Pielou, E. C. The measurement of diversity in different types of biological collections. *J.*
1076 *Theor. Biol.* **13**, 131–144 (1966).
1077 70. Alonso-Sáez, L., Gasol, J. M., Lefort, T., Hofer, J. & Sommaruga, R. Effect of natural
1078 sunlight on bacterial activity and differential sensitivity of natural bacterioplankton groups in
1079 northwestern Mediterranean coastal waters. *Appl. Environ. Microbiol.* **72**, 5806–5813
1080 (2006).
1081 71. Ruiz-González, C., Simó, R., Sommaruga, R. & Gasol, J. M. Away from darkness: a review
1082 on the effects of solar radiation on heterotrophic bacterioplankton activity. *Front. Microbiol.*
1083 **4**, 131 (2013).
1084 72. Liu, X. & Salles, J. F. Drivers and consequences of microbial community coalescence.
1085 *ISME J.* **18**, (2024).
1086 73. Lavrinienko, A., Jernfors, T., Koskimäki, J. J., Pirttilä, A. M. & Watts, P. C. Does
1087 Intraspecific Variation in rDNA Copy Number Affect Analysis of Microbial Communities?
1088 *Trends Microbiol.* **29**, 19–27 (2021).
1089 74. Plummer, M. JAGS: A program for analysis of Bayesian graphical models using Gibbs
1090 sampling. *3rd International Workshop on Distributed Statistical Computing (DSC 2003);*
1091 *Vienna, Austria* (2003).
1092 75. Watanabe, S. A widely applicable Bayesian information criterion. *J. Mach. Learn. Res.* **14**,

1093 867–897 (2012).
1094 76. Casey, J. R., Grinstein, S. & Orlowski, J. Sensors and regulators of intracellular pH. *Nat.*
1095 *Rev. Mol. Cell Biol.* **11**, 50–61 (2010).
1096 77. Schneider, B. *et al.* Biogeochemical cycles. in *Biological Oceanography of the Baltic Sea*
1097 (eds. Snoeijs-Leijonmalm, P., Schubert, H. & Radziejewska, T.) 87–122 (Springer
1098 Netherlands, Dordrecht, 2017).
1099 78. Kogure, K. Bioenergetics of marine bacteria. *Curr. Opin. Biotechnol.* **9**, 278–282 (1998).
1100 79. Zhang, H. *et al.* Repeated evolutionary transitions of flavobacteria from marine to non-
1101 marine habitats. *Environ. Microbiol.* **21**, 648–666 (2019).
1102 80. Benarroch, J. M. & Asally, M. The Microbiologist's Guide to Membrane Potential Dynamics.
1103 *Trends Microbiol.* **28**, 304–314 (2020).
1104 81. More, K. J., Kaur, H., Simpson, A. G. B., Spiegel, F. W. & Dacks, J. B. Contractile vacuoles:
1105 a rapidly expanding (and occasionally diminishing?) understanding. *Eur. J. Protistol.* **94**,
1106 126078 (2024).
1107 82. Rousk, J. *et al.* Soil bacterial and fungal communities across a pH gradient in an arable soil.
1108 *ISME J.* **4**, 1340–1351 (2010).
1109 83. Mulkidjanian, A. Y., Dibrov, P. & Galperin, M. Y. The past and present of sodium
1110 energetics: may the sodium-motive force be with you. *Biochim. Biophys. Acta* **1777**, 985–
1111 992 (2008).
1112 84. de Vargas, C. *et al.* Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean.
1113 *Science* **348**, 1261605 (2015).
1114 85. Kawano, S., Ono, H., Takagi, T. & Bono, H. Tutorial videos of bioinformatics resources:
1115 online distribution trial in Japan named TogoTV. *Brief. Bioinform.* **13**, 258–268 (2012).
1116 86. Agneta, A. *et al.* DNA extraction protocol for DNA-metabarcoding of marine phytoplankton
1117 using Zymobiomics DNA minprep kit (Zymo Research; D4300) v1. (2021)
1118 doi:10.17504/protocols.io.bucjnsun.

1119 87. Balzano, S., Abs, E. & Leterme, S. C. Protist diversity along a salinity gradient in a coastal
1120 lagoon. *Aquat. Microb. Ecol.* **74**, 263–277 (2015).
1121 88. Wu, L. *et al.* Phasing amplicon sequencing on Illumina Miseq for robust environmental
1122 microbial community analysis. *BMC Microbiol.* **15**, 125 (2015).
1123 89. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
1124 *EMBnet.journal* **17**, 10–12 (2011).
1125 90. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data.
1126 *Nat. Methods* **13**, 581–583 (2016).
1127 91. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny
1128 substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
1129 92. Kozlov, A. M., Zhang, J., Yilmaz, P., Glöckner, F. O. & Stamatakis, A. Phylogeny-aware
1130 identification and correction of taxonomically mislabeled sequences. *Nucleic Acids Res.* **44**,
1131 5022–5033 (2016).
1132 93. Lundin, D. & Andersson, A. SBDI Sativa curated 16S GTDB database. *SciLifeLab*
1133 *https://doi.org/10.17044/scilifelab.14869077*, v3 (2021).
1134 94. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data
1135 processing and web-based tools. *Nucleic Acids Res.* **41**, D590–6 (2013).
1136 95. Guillou, L. *et al.* The Protist Ribosomal Reference database (PR2): a catalog of unicellular
1137 eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41**,
1138 D597–604 (2013).
1139 96. Straub, D. *et al.* Interpretations of Environmental Microbial Community Studies Are Biased
1140 by the Selected 16S rRNA (Gene) Amplicon Sequencing Pipeline. *Front. Microbiol.* **11**,
1141 550420 (2020).
1142 97. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves
1143 sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).
1144 98. R Core Team. R: A Language and Environment for Statistical Computing. Preprint at

https://www.R-project.org/ (2022).

99. Inkscape Project. *Inkscape*. (2020).

100. Oksanen, J. *et al.* vegan: Community Ecology Package. Preprint at https://CRAN.R-project.org/package=vegan (2022).

101. 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).

102. 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).

103. Wilson, E. B. Probable inference, the law of succession, and statistical inference. *J. Am. Stat. Assoc.* **22**, 209–212 (1927).

104. Plummer, M. rjags: Bayesian Graphical Models using MCMC. Preprint at https://CRAN.R-project.org/package=rjags (2023).