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], 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, Oceanographic Services, 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*

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

Abstract:

Planktonic microorganisms in coastal waters form the foundation of food webs and biogeochemical cycles while exposed to pronounced environmental gradients, especially brackish salinities. Yet, commonplace ecological assessment overlooks most of their diversity. Here, we analyzed the protist and bacterial diversity from new and publicly available DNA metabarcoding data collected alongside the Swedish marine monitoring program. We show that salinity, unlike other environmental factors, had a stronger effect on bacterial than protist community composition. The seasonality of protist but not bacterial families showed high geographic variation. Bacterial alpha diversity increased with dissolved inorganic nitrogen, while protist alpha diversity was highest in near-marine salinities. Microbial community composition patterns displayed interannual stability despite technical differences affecting the detection of rare taxa. Co-occurrence analysis identified clusters of potentially interdependent microorganisms.

Bayesian modeling showed that the same bacterial lineages were less likely than protists to occur in both lower (<9 PSU) and higher (>15 PSU) brackish salinities. We propose that protists are less ecologically sensitive to salinity due to the disconnection of basic metabolic processes from the cell membrane through compartmentalization. Ultimately, incorporating DNA metabarcoding into an environmental monitoring program allowed us to connect ecological and biogeographic processes with understudied taxa and biodiversity dynamics.

Introduction:

Microorganisms play indispensable roles in biogeochemical cycles[1]. Planktonic bacteria and protists account for around half of primary production on Earth[2], with a disproportionally 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]. Finally, coastal zones comprise multiple land-to-sea environmental gradients, including a full range of brackish salinities, leading to an immense diversity of habitats[5,8].

Baltic Sea and adjacent Kattegat and Skagerrak (hence “Baltic Sea area,” Fig. 1a) are among the best-studied coastal areas[9]. Recent formation (8000 years ago[10]) and pronounced environmental gradients and barriers make the Baltic Sea a model for studying eco-evolutionary processes driven by dispersal and adaptation[11]. Simultaneously, its early exposure to anthropogenic pressures, especially eutrophication, can foreshadow the processes awaiting other coastal regions[12]. Finally, the Baltic Sea area has a long record of microbial research. With the earliest data coming from the XIX century[13], regular microscopy-based phytoplankton monitoring has been run by multiple countries in the region for decades[14] and since the 1980s[15] in Sweden[15]. At the same time, the Baltic Sea area was one of the first brackish waters in which bacteria and protists were surveyed using DNA metabarcoding[16–18] and metagenomics[19,20]. Still, most of the unknown diversity in the Baltic Sea area is believed to be microbial[21]. The research so far has mainly focused on groups of phytoplankton and metazooplankton[21], while the roles of other microorganisms on ecosystem functioning are less understood.

Globally, salinity is one of the strongest factors structuring bacterial[22–24] and protist[25] communities. However, the importance of salinity has been, to a large extent, inferred from the divide between freshwater and marine environments[26–28]. Our previous research shows brackish waters host unique bacterial species[20,29]. 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[17,18,30–32]. Moreover, bacterial communities change more across the salinity

gradient than between seasons[33]. However, whether protists are as ecologically sensitive to salinity as bacteria is less known.

The number of aquatic animal species tends to reach a minimum within the brackish spectrum across salinity gradients[21,34]. For phytoplankton, patterns with a maximum[30] and minimum[31] species richness (alpha diversity) within brackish waters have been reported. DNA metabarcoding transect surveys of plankton showed no clear patterns for bacteria[17] and major eukaryotic taxa[18]. At the same time, bacterial alpha diversity has been observed to be lower in winter than in summer in the Baltic Sea[33,35], consistently with patterns observed in marine environments[36–38]. 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-detail spatiotemporal analysis of bacterial and protist diversity across the Baltic Sea area. We base the analysis on a recently published DNA metabarcoding dataset[39], as well as new data released here, together corresponding to 398 samples We show how protists and bacteria differ in their ecological sensitivity to salinity, compare seasonal patterns of closely related species in different regions, and analyze drivers of taxonomic richness. We propose an explanation of the contrasting bacterial and protist diversity patterns based on fundamental physiological differences between the groups. Finally, we use co-occurrence data to identify core microbiomes characteristic of overlapping conditions within the Baltic Sea area.

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 stations across broad environmental gradients regularly sampled throughout a year[39]. After filtering procedures, we used the data from 246 independent samples. They came from 18 locations, extending from Bothnian Bay (minimal salinity = 2.00 PSU) to Skagerrak (maximal salinity = 33.92 PSU), 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 macroalgae (Rhodophyta, *Phaeophyceae*, *Cyanidiales*). We thus treated the remaining 16S ASVs as bacteria and 18S ASVs as unicellular eukaryotes, i.e., protists.

We clustered the ASVs by simultaneous genetic and distribution similarity, using distribution-based operational taxonomic unit-calling (dbOTU[40]; 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.

In multiple analyses, we used family-level annotation to either group or characterize dbOTUs. This choice was mainly motivated by family being the lowest taxonomic level at which more than half of dbOTUs (and protist reads) were annotated (Supplementary Fig. S1c-f). Additionally, in experimental conditions, bacterial community assembly processes tend to converge at the family level but are more stochastic at lower taxonomic levels, suggesting ecologically relevant functional coherence[41].

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 rRNA), which better estimates absolute abundance than relative counts[42]. Thus, it can uncover blooming and succession patterns obscured from relative abundance analysis[43]. At the same time, the spike-in approach is highly sensitive to differences in sample preparation and amplification[42]. This can lead to noisier results, especially in environmental monitoring, as samples are 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 Skagerrak, Kattegat, and the three basins of the Baltic Sea included in this study (Baltic Proper, Bothnian Sea, Bothnian Bay).

We observed an increased abundance of cyanobacteria as early as May in the Baltic Proper (Fig. 1b), consistent with microscopic observations[44]. The increase in cyanobacterial abundance was smaller and peaked later moving northwards within the Baltic Sea. In the Baltic Proper, both dinoflagellates and *Gyrista* (the protist subdivision including diatoms), peaked in abundance in March (Fig. 1c), as observed using microscopic methods[44,45]. 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 Kattegat and 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 non-phytoplankton groups. *Bacteroidia* (Fig. 1b) peaked in abundance in May across the Baltic Sea, consistently with previous reports[35,46], but displayed less clear seasonal dynamics in Skagerrak and Kattegat. *Actinomycetia* and *Acidimicrobiia* were abundant in the Baltic Sea but not in Kattegat/Skagerrak, consistent with *Actinobacteria* abundance patterns from transect-based studies[17–19]. 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 coastline station[35,46], *Alphaproteobacteria* and *Gammaproteobacteria* comprised a stable proportion of the whole community throughout the year in all basins but Kattegat (Supplementary Fig. S2a). Still, consistently with metabarcoding[17] and metagenomic transects[19], *Alphaproteobacteria* and *Gammaproteobacteria* were more abundant in Skagerrak and Kattegat 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 or *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 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). Relative abundances better captured gradual succession patterns, especially for bacteria (Supplementary Fig. S2a). These gradual changes likely reflect sensitivity to factors other than the productivity of the system. 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.

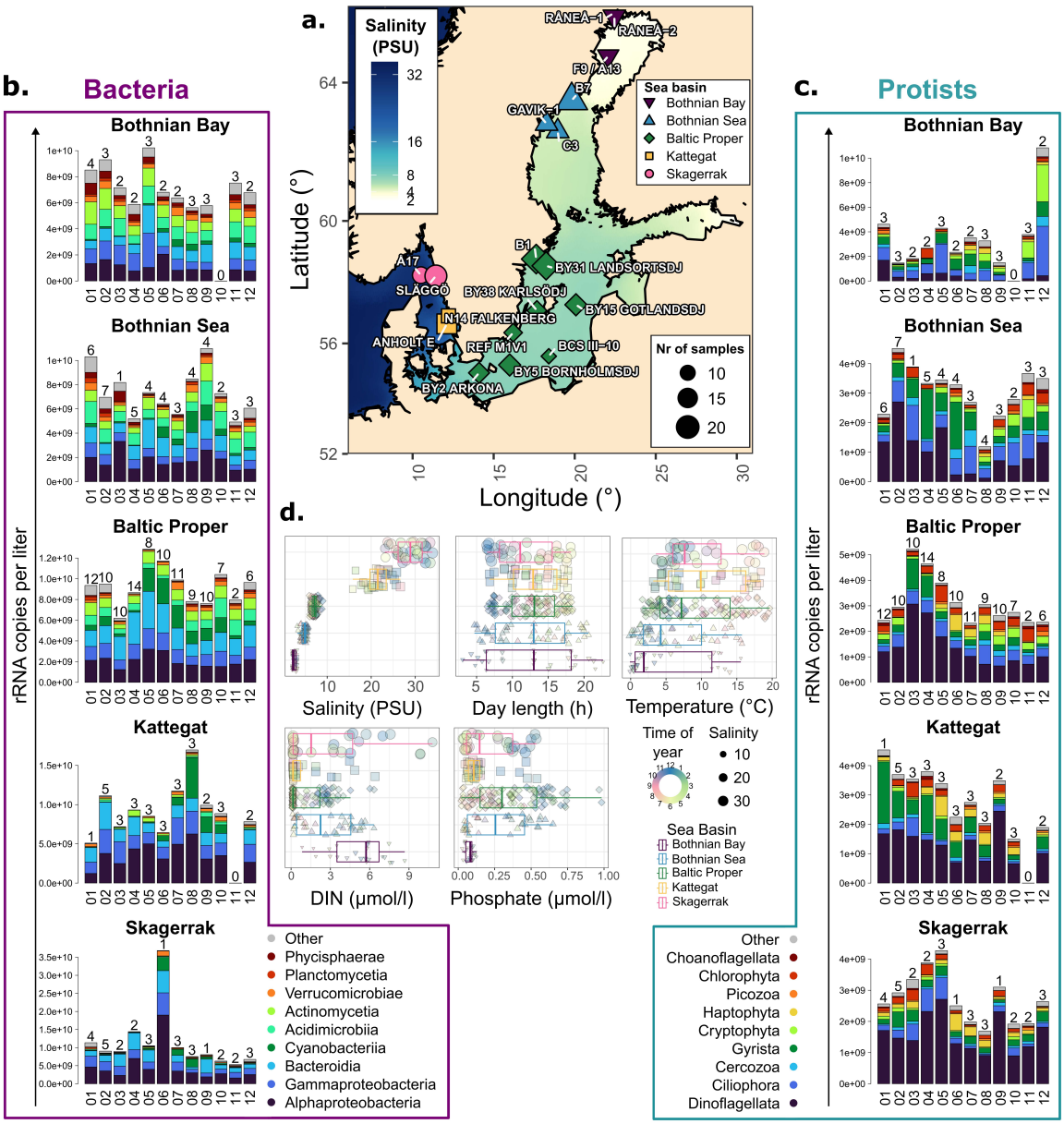


Fig. 1 | Spatiotemporal changes in abundance of major microbial taxa, 2019-2020 a. A map of 18 sampling locations labeled by the name of the station and size-coded by the number of samples collected over a period of 13 months in 2019 and early 2020. **b-c.** The numbers of rRNA marker gene copies per 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) are 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 roughly correlated ($R^2 = 0.58$). 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 waters[47].

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 ($R^2 < 0.4$, 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, while temperature and nutrients were of greater importance than for bacteria (Fig. 2b). These results contradict the proposed rule that the distribution of smaller organisms is less limited by environmental factors than that of bigger organisms[48], and in particular that bacteria are less environmentally filtered than protists[49].

Nevertheless, for both bacterial and protist communities, there was a detectable salinity divide between the Baltic Sea and 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*)[50] and *Limnocylindraceae* (*Chloroflexi*)[51] were associated with the lowest salinity samples, while commonly marine *Rhodobacteraceae*[52] and *Thioglobaceae*[53] were at the other end of the salinity spectrum (Fig 2c). Overall, many abundant and unevenly spatially distributed bacterial and protist families that we identified have unknown metabolic capabilities and preferred habitats. Our results suggest their potentially crucial ecological importance and point to candidate niche-defining factors.

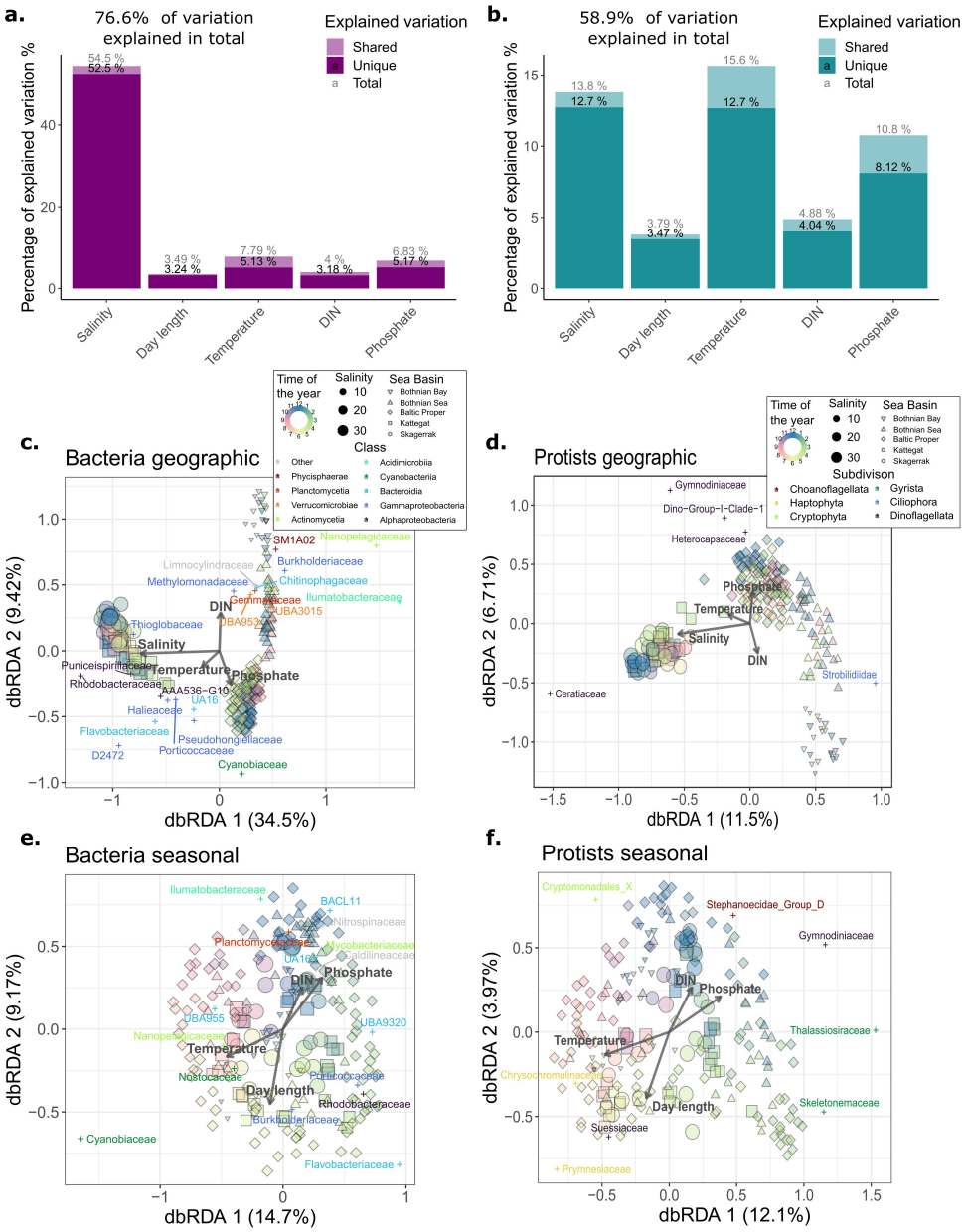


Fig. 2 | Environmental factors explaining geographic and seasonal community structuring and the most correlated families of bacteria and protists a-b. Variance partitioning of bacterial (**a.**) and protist (**b.**) beta diversity, based on Bray-Curtis distances and distance-based Redundancy Analysis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

Closely related bacteria but not protists follow analogous seasonal dynamics across the salinity gradient

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*[54] to multiple uncharacterized *Bacteroides* families.

In all the basins, picocyanobacteria (*Cyanobiaceae*) were associated with high temperatures (Fig. 3a, Supplementary Fig. S6a). The rise in picocyanobacterial abundance was followed by blooms of filamentous heterocyst-forming cyanobacteria (*Nostocacaea*) in the Baltic Proper and the Bothnian Sea (Fig. 3a-b). Expectedly, *Nostocacaea* thrived in DIN-depleted conditions (Fig. 3b-c). The nitrogen fixed by filamentous cyanobacteria leaks to picocyanobacteria, which has been suggested to promote the growth of the latter group in DIN-deficient conditions[55]. However, we observed a high abundance of *Cyanobiaceae* in DIN-depleted Skagerrak and Kattegat, where no filamentous cyanobacteria blooms occurred (Fig. 3a-c). Thus, other mechanisms, potentially other diazotrophs or efficient recycling of organic nitrogen, can sustain picocyanobacterial populations under nitrogen-depleted conditions. Finally, only in Baltic Proper did we observe an increase in the abundance of *Cyanobiaceae* in winter. Therefore, previously identified winter increases of certain *Synechococcus* sp. in Baltic Proper[56] may be specific to this region.

Similar to cyanobacteria, the abundance of haptophyte families *Prymensiaceae* and *Chrysochromulinaceae* were associated with high temperatures (Fig. 2f). However, we could not connect any family to similar conditions across all the basins—not only among haptophytes but among protists in general (Supplementary Fig. S6b). This suggests that despite stronger geographic structuring of bacterial communities along the salinity gradient (Fig. 2a-d), protist seasonal succession patterns differ more between regions. The aforementioned haptophytes may best exemplify this heterogeneity. *Prymensiaceae* were by far most abundant in summer in Baltic Proper (Fig. 3d), while *Chrysochromulinaceae* bloomed at a similar time in Bothnian Bay (Fig. 3e). These families also increased in abundance around June in Skagerrak, largely explaining the increase in overall haptophyte counts (Fig. 1c). The same cannot be said about Kattegat, where haptophytes were most abundant in May (Fig. 1c). Both in Skagerrak and Kattegat, *Gephyrocapsa huxleyi* (formerly *Emiliania huxleyi*) is known to form summer/late spring blooms[57]. As *Gephyrocapsa oceanica* and *G. huxleyi* are indistinguishable using 18S metabarcoding[58], we investigated the abundance of *Noelaerhabdaceae*, a family containing these species. Expectedly[59], we observed this group to be abundant in nutrient-deficient conditions, with the inorganic N:P ratio much lower than the Redfield ratio of 16:1, and only in Kattegat/Skagerrak (Supplementary

Fig. S7a, d). Still, it did not explain Kattegat's high abundance of haptophytes in May. We ultimately found a May bloom of *Phaeocystaceae* in Kattegat (Supplementary Fig. S7e), which was short-lived, as is typical of *Phaeocystis*[57].

Overall, in each basin of the Baltic Sea area (except the Bothnian Sea), a different set of haptophyte species bloomed under warm, nutrient-depleted conditions. Ultimately, which taxa within *Haptophyta* flourish affects ecosystem functioning, as they differ in their toxicity[57], as well as roles in carbon and sulfur cycles[60]. So is the case for many other groups, including dinoflagellates and diatoms, which can be similarly investigated using DNA monitoring data in focused studies.

BACL11, *Bacterioides* with streamlined genomes and low GC content[20], was the only family other than *Cyanobiaceae* associated with seasonal changes in community composition across all the studied regions (Supplementary Fig. S6a). However, opposite to *Cyanobiaceae,* BACL11 thrived in cold waters (Fig. 2e, Supplementary Fig. S6a). Our observations diverge from the previously suggested dinoflagellate-bloom association of BACL11[20]. Complex seasonal patterns (Fig. 3g) of BACL11 suggest dependence on a combination of factors or a factor not investigated here. On the other hand, nitrite-oxidizing *Nitrospinaceae*[61], a winter-associated family (Fig. 2e, Supplementary Fig. S6a), was clearly influenced by DIN dynamics (Fig. 3c, h). Still, their abundance usually decreased before DIN depletion, as well as during the summer in the Bothnian Bay, where nitrogen availability remained relatively high throughout the year. A better understanding of the growth-limiting factors of *Nitrospinaceae* is needed to explain these patterns.

Finally, *Flavobacteriaceae* showed seasonal patterns linearly uncorrelated with the investigated environmental factors (Fig. 2e, Supplementary Fig. S6a). However, this *Bacteroides* family rose in abundance at the periods corresponding to a rapid decrease in phosphate availability (Fig. 3f, i), as did total *Bacteroidia* abundance (Fig. 1c). Recently, it has been suggested that *Bacteroidia*, and in particular *Flavobacteria*, gain a competitive advantage under limited phosphate availability by efficiently mobilizing organic phosphorus 61[62,63]. Our results further support this notion and suggest that phosphorus availability contributes to brackish *Flavobacteriaceae* seasonality. Still, this adaptation strategy appears limited, as less *Flavobacteriaceae* was detected during phosphate-minimum conditions in summer. Apart from phosphorus mobilization, *Flavobacteriaceae* are adapted to efficiently utilize algal biomass[64], which might explain why they are abundant under decreasing but not increasing phosphate availability.

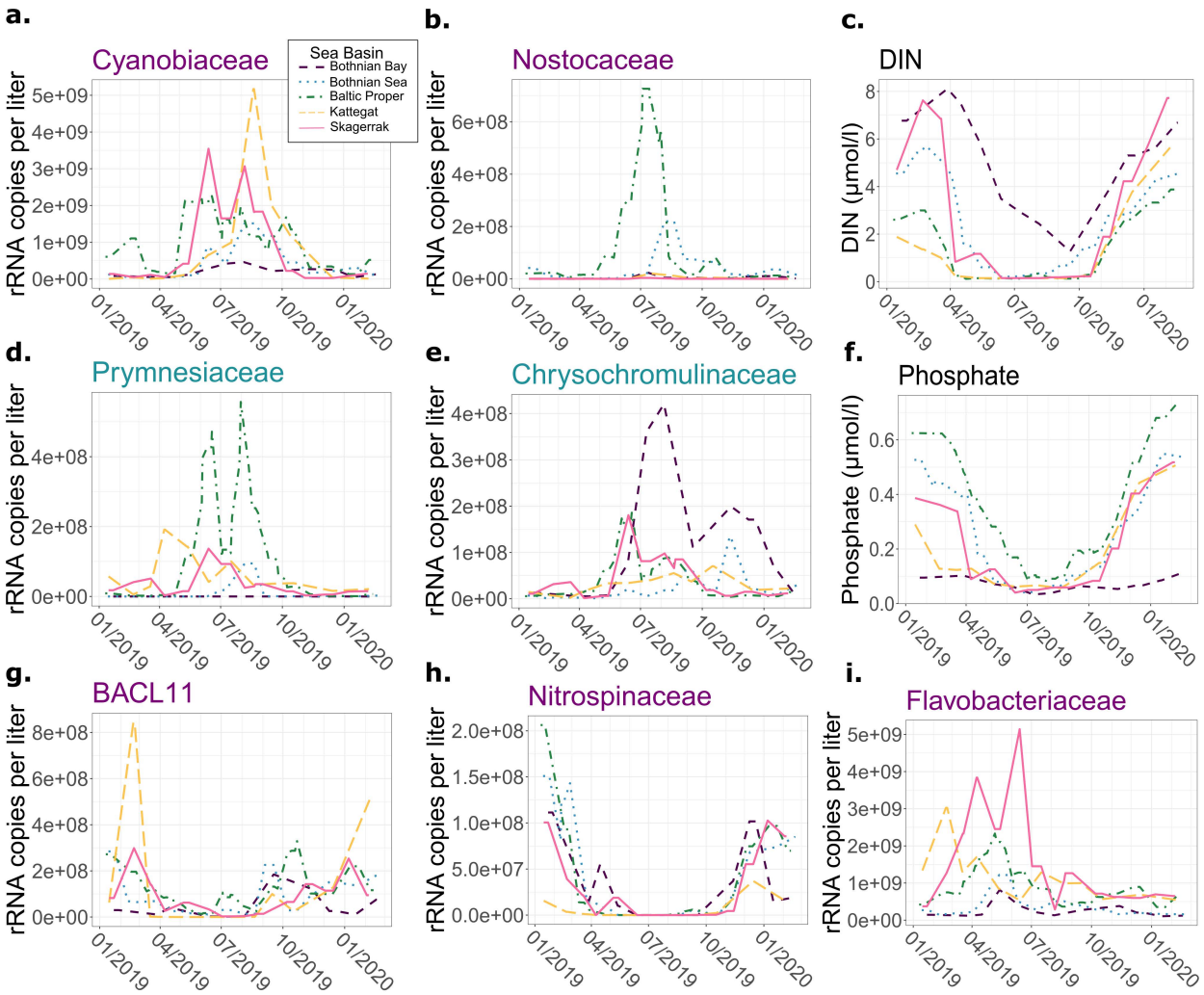


Fig. 3 | Seasonal dynamics of selected taxa and nutrients. a-b., d-e., g-i. Numbers of rRNA marker gene copies annotated to families associated with beta diversity seasonal dynamics (see Fig. 2e-f). Families of cyanobacteria (**a-b**.), haptophytes (**d-e.**), *Bacteroidia* (**g-h**.), and *Nitrospinia* (**i.**) are shown. The names of bacterial families are colored purple and of protist families green. **c, d.** Concentration of inorganic nitrogen (**c.**) and phosphate (**f.**) across the samples in 2019-2020. The lines correspond to the rolling mean for each basin, with values averaged for measurements two weeks before and after each date. The same legend (**a.**) applies to all the panels, with basins ordered from lowest to highest salinity (top to bottom). DIN - dissolved inorganic nitrogen.

Inorganic nitrogen availability associated with bacterial alpha diversity

Bacterial alpha diversity declined and increased rapidly (Fig. 4a) at the beginnings and ends of nitrogen-depletion periods (Fig. 3c). Consistently, DIN was the environmental factor by far best-explaining variation in bacterial richness using a linear model (Fig. 4c).

322 Many bacteria use inorganic nitrogen as an energy source or an electron
323 acceptor in nitrification and denitrification processes. In the previous analyses, we found
324 the nitrate oxidizers *Nitrospinaceae* associated with high-DIN winter conditions in
325 multiple basins. Among dbOTUs preferentially found in high DIN (>0.5 µmol) conditions,
326 a few families other than *Nitrospinaceae* were overrepresented (Fig. 4e). Among them
327 was UBA2979 (corresponding to *Ca*. *Bathosphaeria*[65], *Chloroflexota*) and
328 *Rhodospirillaceae*, with representatives containing genes for nitrate reduction[51,66].
329 Indeed, the latter family, while comprising diverse metabolic strategies, includes major
330 marine denitrifiers[66,67]. Moreover, representatives of UBA8477 (*Marinisomatota*, 100%
331 barcode (ASV4915) sequence identity with clade SHBH1141[68] representatives
332 confirmed using BLAST[69]) express genes for nitrous oxide reduction[68]. While
333 denitrification is an anaerobic process, anoxic microenvironments can form in organic
334 particles suspended in surface waters[70] and sustain denitrification[71]. Indeed, such
335 oxygen-depleted microniches have been found during cyanobacterial blooms in the
336 Baltic Sea but with low denitrification rates connected to insufficient nitrate
337 concentrations in the surrounding water[71]. Still, a spatiotemporal metagenomic study is
338 needed to confirm the connection between the higher taxonomic diversity and the
339 nitrogen-related metabolic capabilities.
340 Inorganic nitrogen availability opens up niches for nitrifying and denitrifying, thus
341 potentially directly contributing to increased bacterial dbOTU richness in winter. Still,
342 other factors are likely also involved. It has previously been suggested that lower
343 bacterial alpha diversity in summer than in winter results from the effects of
344 cyanobacterial blooms on metabarcoding data[33]. Cyanobacteria would thus dominate
345 the 16S rRNA pool, decreasing the probability of detecting other taxa, even if those
346 taxa's abundances remain stable. However, bacterial dbOTU correlated much more
347 weakly with the relative abundance of cyanobacteria than with DIN in our dataset
348 (Supplementary Fig. S8). Neither could the higher alpha diversity easily be explained by
349 autumn water mixing, as previously suggested[33], as bacterial dbOTU richness remained
350 high throughout the winter (Fig. 4a). Also, day length, strongly correlated with bacterial
351 alpha diversity in a marine time series[36,38], had smaller explanatory power in our
352 spatiotemporal analysis, mainly overlapping with covarying factors (Fig. 4c). Finally,
353 nutrient depletion comes from and is maintained by subsequent blooms of different
354 protist and bacterial phytoplankton groups. Thus, DIN-associated low dbOTU richness
355 might also be indicative of the biological effects of these disrupted conditions. However,
356 we would then expect a similar DIN-dependency among protists, which is not the case
357 (Fig. 4b and d).

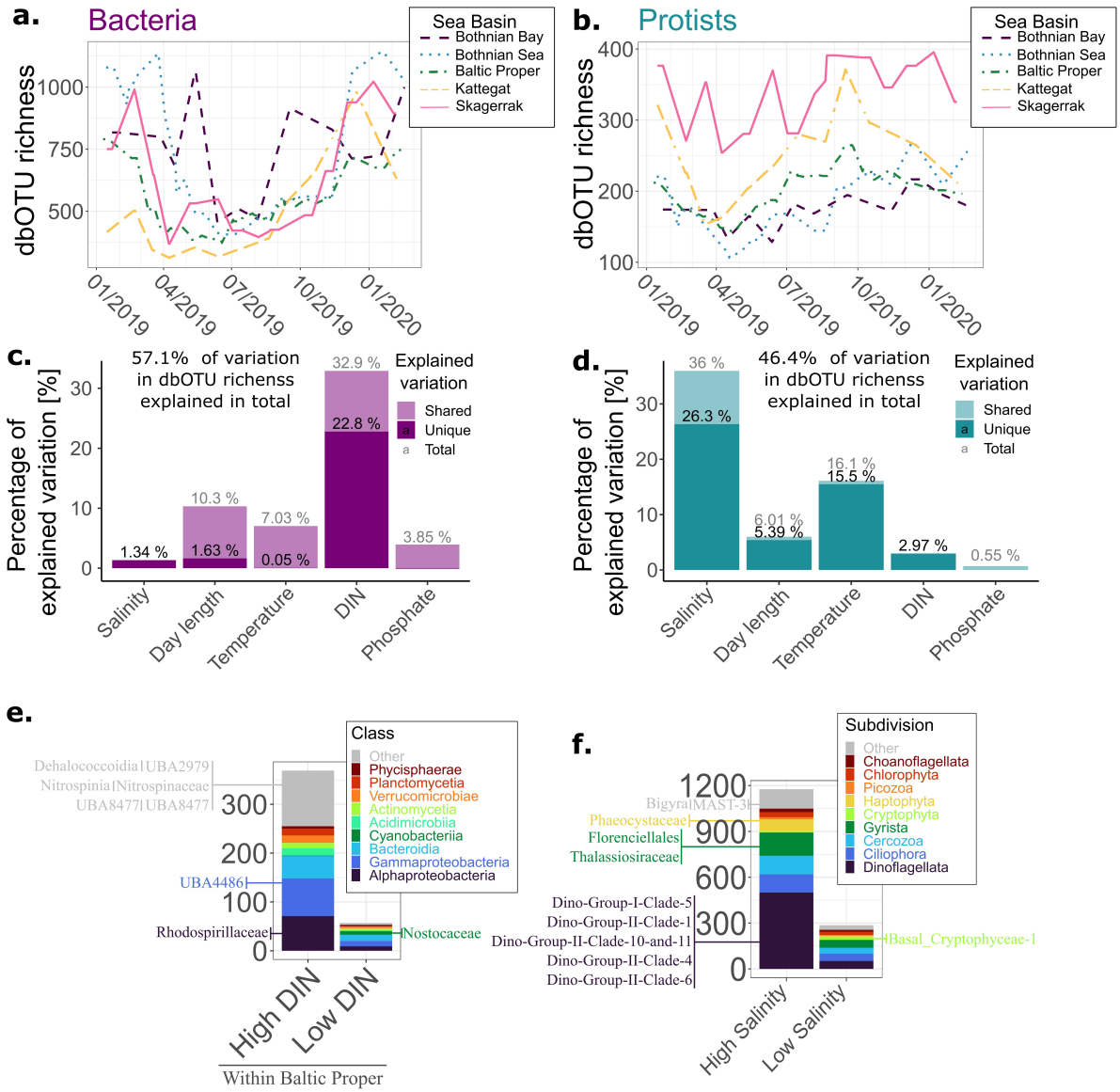


Fig. 4 | Seasonal dynamics of alpha diversity across regions a-b. Mean number of bacterial (**a.**) and protist (**b.**) dbOTUs found across 100 refraction iterations (dbOTU richness). The lines for the distinguished Baltic Sea area basins correspond to the rolling mean, with values averaged for measurements two weeks before and after each date. **c-d**. Variance partitioning of bacterial (**c.**) and protist (**d.**) dbOTU richness. **e.** Number of dbOTUs found significantly (FDR < 0.01 based on Wilcoxon test) more often in high DIN (> 0.5 µmol/l) and low DIN (≤ 0.5 µmol/l), respectively, in the Baltic Proper. The differentially present dbOTUs are grouped by major bacterial classes (classes with relative abundance in the whole dataset > 0.01). Families overrepresented among the differentially present dbOTUs (FDR < 0.01 based on hypergeometric test) have been named. If the family came from an unlabeled class, the name of the class is given on the left-hand side of the family's name. **f.** Same as e., but comparing protist dbOTUs more often found in high (>10 PSU) and low (≤ 10 PSU) salinity, grouped by major protist subdivisions.

Higher protist alpha diversity in near-marine salinities

Protist dbOTU richness was strongly associated with salinity and was especially high in Kattegat and Skagerrak (Fig. 4b, d). Similarly, metazooplankton and fish alpha diversity in Skagerrak and Kattegat is higher than in the Baltic Sea, due to many marine species in higher brackish salinities, outweighing the presence of freshwater species in lower salinities[21]. The Baltic fauna comes mainly from a limited number of local adaptations to brackish conditions following the last glaciation period[11,21]. In contrast, for Baltic Sea bacteria, we recently showed evidence against widespread occurrence of such local adaptation, but rather colonization by a global brackish microbiota[29]. Consistently, both this (Fig. 4a, c) and previous studies[17,18,33] showed no evidence for decreased bacterial diversity in the Baltic Sea compared to Kattegat/Skagerrak. Whether the observed protist diversity pattern originates from a selected group of locally adapted species, as is the case for animals, is yet to be determined. Notably, unlike bacteria[72], marine protist species show a considerable dispersal limitation[28,73]. If the same is true in brackish waters, it would decrease the chances of colonization by brackish protists from distant locations, promoting local adaptation.

Among dbOTUs more often found in high salinity (>15 PSU), there were multiple families of *Syndiniales* ("Dino-Group" clades, Fig. 4f). These parasitic dinoflagellates are known to be more abundant and diverse in marine than freshwater environments[74]. Similarly, all described[75] MAST-3[76], *Florenciellales*[77,78], and *Phaeocystis*[79] species are marine (and all *Phaeocystaceae* we found were annotated to the genus *Phaeocystis*). As *Cyclostephanos*, *Stephanodiscus*, and *Discostella* are annotated to *Stephanodiscaceae* in the PR2 database[75], it leaves *Thalassiosiraceae* as a family with a majority of representatives from marine environments[80,81]. Systematic analyses with higher taxonomic resolution are needed to determine to what extent marine protists populate the higher brackish salinity ranges, as in Kattegat and Skagerrak.

Temperature was the second factor most strongly associated with protist alpha diversity (Fig. 4d), and in most basins, dbOTU richness increased from March to September (Fig. 4b, Supplementary Fig. S7b). This pattern might, however, reflect a slow recovery after spring blooms. Still, temperature was also the factor best explaining protist beta diversity (Fig. 2b), and our results therefore suggest that both common and rare protists might be more susceptible to the direct effects of climate change than bacteria.

Interannual stability of microbial community composition

To complement the 2019-2020 data analyzed above, 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. 5a). Similarly to the 2019-2020 dataset, it mainly

410 corresponded to a 13-month long time-series, between February 2016 and March 2017,
411 with a few extra samples from 2015 from one station (Släggö, Fig. 5a). There were
412 minor differences in sample processing between the two datasets, most notable being:
413 DNA extraction kit, different Illumina sequencing systems, and lack of spike-in
414 sequences in the 2015-2017 dataset (see Methods for more details).
415 Overall, the community composition in the two datasets differed minimally,
416 following the spatial and seasonal environmental gradients in an analogous manner
417 (Fig. 5c-f). This apparent interannual stability of community composition is consistent
418 with previous observations from the Baltic Sea[46,82] and other waters[83]. Finally, the
419 similarity of communities between the two datasets shows that their technical
420 differences have little impact on the observed general community composition (Fig. 5e-
421 f).
422 Only a two-year break separated the datasets. The seasonal patterns likely
423 change over longer time scales, as, e.g., climate change has altered the phytoplankton
424 bloom dynamics in the Baltic Sea over decadal scales[84]. As climate change
425 accelerates[85] and nutrient inputs are being decreased in the Baltic Sea[86], the microbial
426 communities may shift in unforeseen directions. These changes may result not only
427 from shifting abiotic factors but also from biotic interactions, as, e.g. invasive species
428 introduction can alter microbial community dynamics[87]. A reproducible monitoring
429 scheme encompassing diverse taxa, achievable with DNA-based methods, is crucial for
430 understanding and informed responses to unforeseen changes.
431 In contrast to community composition, alpha diversity differed significantly ($P <$
432 10^{-14} in both cases) for both bacteria and protists (Fig. 5g-h). We doubt this increase,
433 amounting to hundreds of dbOTUs on average, represents an actual biological change.
434 Instead, it likely originates from the technical differences between the datasets. We
435 based our alpha-diversity estimates on rarefaction as it performs best in removing
436 sequencing depth biases[88] (expected due to different Illumina sequencing systems used
437 for the datasets). Moreover, in both datasets, rarefaction curves were saturated,
438 suggesting that extra reads would not change alpha diversity much (Supplementary Fig.
439 S9). Thus, another technical difference likely caused the dbOTU difference between
440 datasets. Its identification would require additional investigation. Ultimately, our results
441 suggest more caution is needed when comparing alpha diversity than beta diversity
442 across datasets.

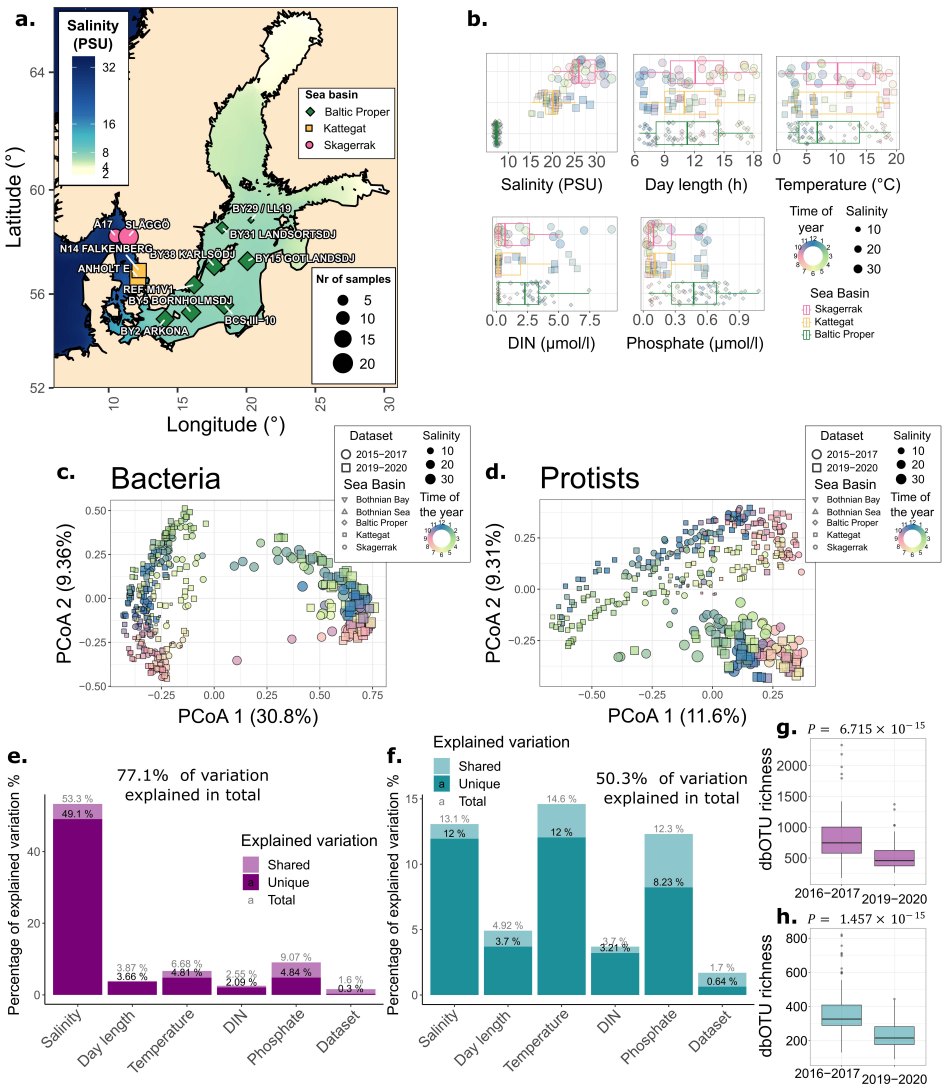


Fig. 5 | Bacterial and protist diversity across datasets collected in 2015-2017 and in 2019-2020. a. A map of 12 sampling locations labeled by the number of samples collected between August 2015 and March 2017 (the 2015-2017 dataset), with all the 2015 samples coming from Släggö in Skagerrak **b.** The distribution of measurements of major physicochemical and geographic parameters for each basin in the 2015-2017 dataset. **c-d**. Principal Coordinatet Analysis (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 distance-based Redundancy Analysis (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. 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 pairwise Wilcoxon signed-rank test are given above each plot. DIN - dissolved inorganic nitrogen. dbOTU - distribution-based operational taxonomic unit.

460 ## Clusters of exclusively co-occurring microorganisms point to
461 potential interdependencies

462 We performed a custom co-occurrence analysis identifying the pairs of dbOTUs most
463 consistently found with but not without each other (FDR < 0.01 for both co-presence and
464 co-absence), i.e., "exclusively co-occurring". These pairs correspond to either
465 unprecedented examples of synchronized environmental filtering or dependence of at
466 least one dbOTU on another. Thus, the clusters of co-occurring taxa we observed (Fig.
467 6a) represent either convergently evolved niche spaces or webs of interdependent
468 organisms.

469 We found three clusters of exclusively co-occurring microorganisms (Fig. 6a). All
470 the clusters contained bacteria with streamlined genomes and known metabolic
471 dependencies. *Pelagibacteraceae* (SAR11), HIMB59 (formerly included in the SAR11
472 clade), D2472, and pelagic *Methylophilaceae* all have unusually small genomes missing
473 essential functions[89–92]. A further, systematic, genome-based analysis is needed to
474 assess whether the other members of the clusters metabolically complement the
475 streamlined bacteria. At the very least, BACL14 (a genus to which *Methylophilaceae*
476 from clusters 1 and 3 belong) and MB11C04 can provide vitamers that
477 *Pelagibacteraceae* depend on[93]. Importantly, despite potentially providing the resources
478 to *Pelagibacteraceae*, pelagic *Methylophilaceae* usually have streamlined genomes[92]
479 (and all BACL14 representatives in GTDB release 09-RS220[94] have genomes smaller
480 than 1.5 Mbp). Overall, the co-occurrence within the clusters might largely correspond
481 to the dynamics of the "Strong Black Queen Hypothesis"[95]. That is, the clusters might
482 represent metabolically interdependent taxa that need to disperse together to establish
483 themselves in a new location. Consequently, the inability of some species to cross the
484 environmental barriers might affect the whole cluster.

485 The distribution patterns of the exclusive co-occurrence clusters across the
486 salinity barrier were even more pronounced based on abundance (Fig 6b-d) than the
487 presence/absence of dbOTUs (Supplementary Fig. S10), despite the co-occurrence
488 analysis being based on the latter data type. Clusters 1 and 3 showed strong
489 preferences for high and low salinities, respectively. They also consisted almost entirely
490 of bacterial dbOTUs. In contrast, cluster 2 spanned across the entire salinity gradient.
491 However, it was less abundant and incompletely represented in the least saline samples
492 from the northernmost location. The bacteria-to-protists ratio in cluster 2 was 2:1, on par
493 with the ratio for the whole set of dbOTUs considered for this analysis (2.005:1.000).
494 These results further support the notion of a stronger salinity limitation among bacteria
495 than protists. However, if these microorganisms truly are interdependent, the effects
496 might be amplified, as the salinity sensitivity of one dbOTU would affect the others.

497 The exclusive co-occurrence clusters were generally abundant and present
498 together throughout the year (Fig 6b-d). However, in the warmest season, their
499 geographic ranges decreased (Fig 6b-d). Some representatives were absent even in

the preferred salinities. The seasonal disappearance might be imposed by competition for depleted nutrients or enhanced grazing on bacteria during the summer[96–98]. The persistence of other members of the clusters might largely depend on cyanobacteria providing the functions of the missing bacteria. For example, during the summer, picocyanobacteria are the main producers of vitamers[93], on which other bacteria depend.

Overall, our results suggest the existence of groups of potentially interdependent microorganisms that tend to spread together. Their differential presence and abundance highlight that the salinity gradient, limiting the ranges of some microorganisms, also changes the biotic conditions for the rest of the community.

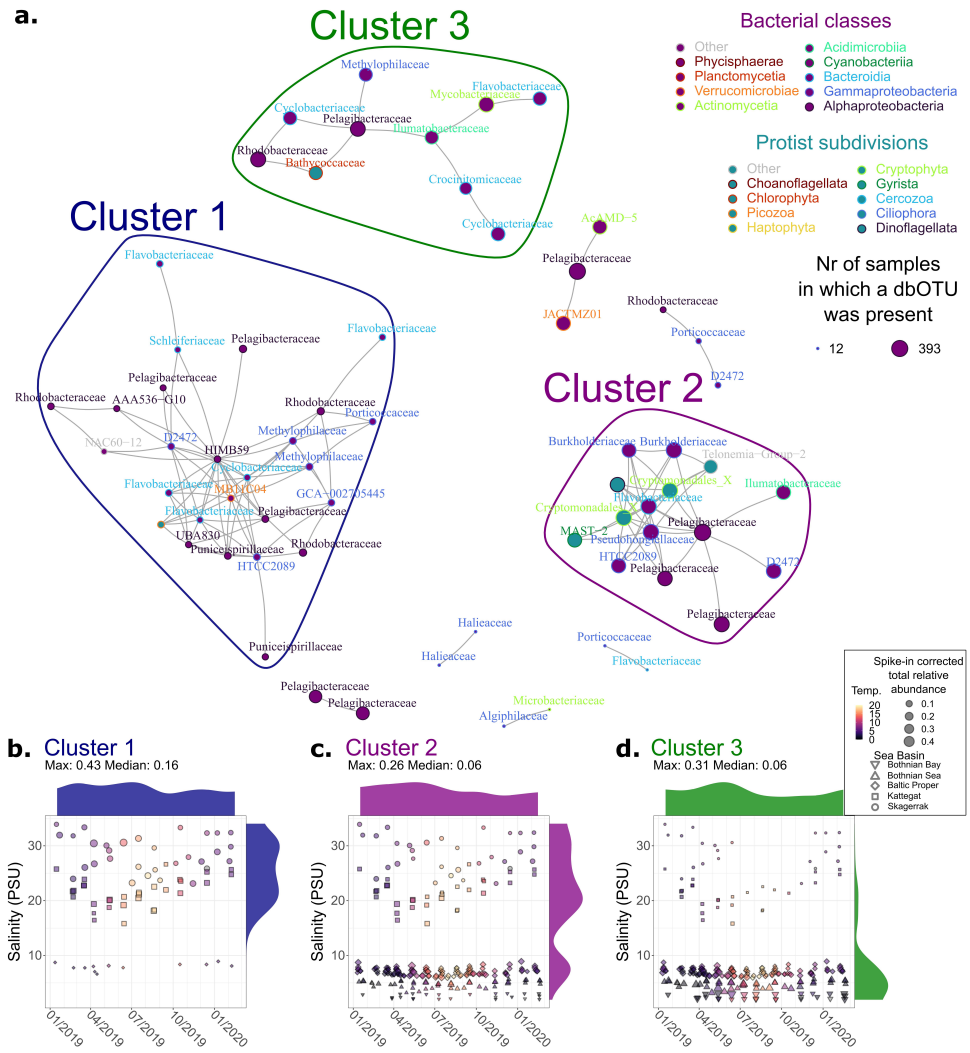


Fig. 6 | Co-occurring bacterial and protist dbOTUs. a. Grey lines link pairs of dbOTUs occurring in and being absent from the same samples (FDR > 0.01 based on beta distribution, corresponding to the less common dbOTU being present in at least 86% of the samples in which the more common one was found and absent from all the samples where the more common one was not found). Each dbOTU is labeled by annotation to bacteria or protists, a class or subdivision (only taxa corresponding to >1% of all reads are labeled), and family (if available). The three largest clusters are distinguished and ordered by the number of dbOTUs belonging to them. Samples from both the 2015-2017 and 2019-2020 datasets were used to build the network. Only dbOTUs present in at least ten samples were considered for the co-occurrence analysis. **b-d.** Relative abundance based on spike-in normalized counts summed over dbOTUs from each cluster across the 2019-2020 dataset. On the top and the right axes of each plot are density plots based on mean spike-in normalized abundance per month (top) and per unit salinity (right). The median given above each plot, unlike the means for the density plots, is based only on samples in which at least one dbOTU from the cluster was present. The same legend (**d.**) applies to panels b-d. Temperature is given in degrees Celsius (ºC). dbOTU - distribution-based operational taxonomic unit.

528 **Bacterial dbOTUs are less likely than protist dbOTUs to occur at both**
529 **high and low brackish salinities**

530 The salinity gradient in the Baltic Sea area contains a characteristic, rapid shift across
531 the Danish straits. Consequently, our data contained geographically proximate samples
532 from lower (<9 PSU) and higher (>15 PSU) brackish salinities (Fig. 1a, Fig. 5a). Our
533 previous analyses showed the separation across this salinity barrier to be the strongest
534 structuring effect on microbial communities, though less so for protists than for bacteria
535 (Fig. 2c-d, Supplementary Fig. S4).
536 We further investigated whether bacteria are, in general, less likely to cross this
537 salinity barrier between lower and higher salinities. First, the proportion of protist
538 dbOTUs crossing the salinity barrier was significantly higher than the proportion of
539 bacterial dbOTUs (Fig. 7a, Supplementary Fig. S11). Based on comparable subsets of
540 samples, the proportion of both bacterial and protist dbOTUs crossing the salinity barrier
541 was higher in the samples from 2016-2017 than from 2019-2020 (Supplementary Fig.
542 S11b-c). This suggests that the above noted differences in alpha diversity (Fig. 4g-h)
543 may be due to a higher sensitivity of the sequencing approach used for 2015-2017 data,
544 not technical artifacts. Still, we based further analyses on the 2019-2020 dataset, since
545 that allowed us to use spike-in normalized counts as an abundance metric comparable
546 between bacteria and protists. Furthermore, to minimize the effects of other
547 environmental filtering effects (especially climate), we chose the same number of
548 stations and samples across the salinity barrier (see Methods for details).
549 The proportion of barrier-crossing bacteria and protists may be affected by
550 detection biases coming from differences in relative abundance, length of vegetative
551 season, and environmental filtering by factors other than salinity. Bacteria were on
552 average more abundant in terms of rRNA gene marker copy numbers (Supplementary
553 Fig. S12a). As protists tend to have higher rRNA copy numbers per cell[128], this
554 difference likely translates to even larger difference in cell numbers, which is likely the
555 most relevant abundance measure in the context of dispersal capabilities and detection
556 probability. Meanwhile protists were present in more samples at one side of the salinity
557 barrier (Supplementary Fig. S12b). Overall, being similarly widespread on one side of
558 the salinity barrier, bacteria were more abundant (Supplementary Fig. S12c-f). Still,
559 across the occupancy values, a higher proportion of protists crossed the salinity barrier
560 (Fig. 7b). The same was true across the abundance values, apart from the most
561 abundant outliers, which were barrier-crossing dbOTUs among bacteria but not protists
562 (Fig. 7c). Generally, across different occupancy and abundance values, a higher
563 proportion of protists crossed the salinity barrier (Fig. 7d, Supplementary Fig. 12g-h).
564 To confirm that protists are more likely to cross the salinity barrier, we build a
565 Bayesian model using a Monte Carlo Markov Chain (MCMC) based on Just Another
566 Gibbs Sampler[99] (JAGS). The model, based on the Bernoulli distribution, calculates the
567 probability of crossing a barrier from logistic regression based on the maximal

occupancy on one side of the barrier, geometric mean abundance, and whether a dbOTU is a protist or a bacterium. After model selection based on the Watanabe-Akaike information criterion[100], we chose a version of the model assuming the probability dependence on the other variables follows a different function for bacteria and for protists (see Methods for details). The model predicted protists to be more likely to cross the salinity barrier across the occupancy and abundance values, both on average (Fig. 7e, Supplementary Fig. S12i-j) and in a majority of MCMC iterations (Fig. 7f). The probability that protists are more likely to cross the salinity barrier was consistently close to one except for dbOTUs with low occupancy and high abundance, in which case detection biases might outweigh the biological signal. Finally, the model consistently showed a positive relationship between occupancy and the probability of crossing the salinity barrier (probability = 1 for both bacteria and protists). Meanwhile, an independent effect of abundance was much less certain (probability of positive effect = 0.526 for protists and 0.614 for bacteria).

Finally, using a similar approach, we built a Bayesian model of total occupancy as a function of the occupancy at the more occupied side of the salinity barrier, based on the Poisson distribution. From this model we calculated the predicted occupancy on the less occupied side of the salinity barrier, and compared it to the experimental values (Fig. 7g). The model and the data showed that protists not only more often crossed the salinity barrier, but also were, in general, more widespread in the less preferable salinity regime.

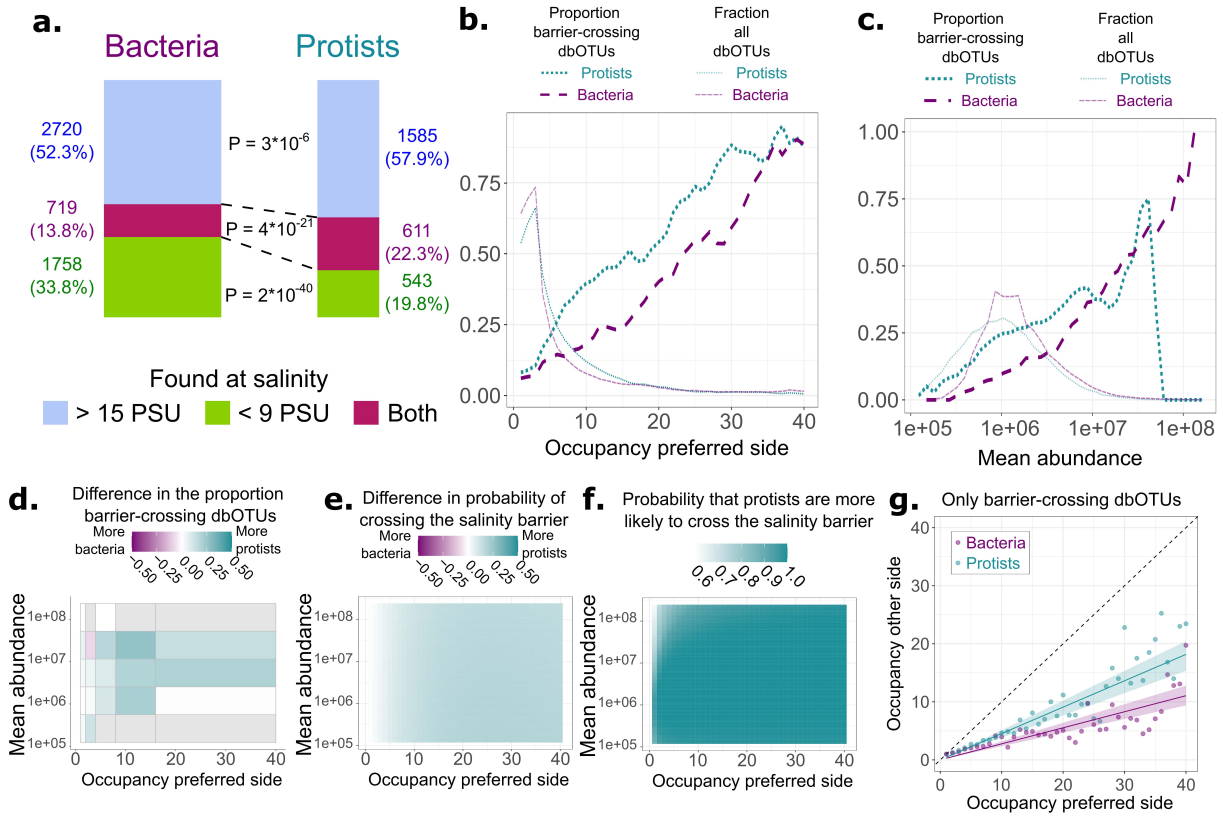


Fig. 7 | Proportion and probability of bacteria and protists (being) found in both high (>15 PSU) and low (<9 PSU) brackish salinity. a. The number and proportion of bacterial and protists dbOTUs found in high (>15 PSU), low (<9 PSU) brackish salinity, and across the salinity barrier. **b-c.** The proportion 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 at the less occupied side of the salinity barrier as a function of the occupancy at 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 rRNA 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 maximum and the minimum occupancy at one of the two sides of the salinity barrier, respectively.

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

615 We propose that compartmentalization makes eukaryotic cells (thus, protists) more
616 acclimatable than prokaryotic ones (thus, bacteria) to a different salinity regime (Fig. 8).
617 In a compartmentalized cell, crucial metabolic processes happen across internal
618 membranes. External ion concentrations and pH do not directly affect these processes.
619 Notably, in eukaryotic cells, the proton motive force used for ATP synthesis is
620 maintained along the inner membranes of mitochondria and plastids. In contrast,
621 prokaryotes usually produce proton motive force across the cellular membrane directly
622 exposed to the external environment. While eukaryotic cytosolic pH is kept at stable,
623 neutral levels[101], marine and brackish waters are buffered and usually slightly alkaline
624 (pH~8)[27,102]. Thus, in the presence of sea salt, some transferred protons are consumed
625 in reactions with (bi-)carbonates or hydroxide. Consequently, many marine bacteria
626 preferably use sodium motive force, i.e., Na^+-gradient across the membrane, for
627 transmembrane transport, motility, and/or ATP production[19,27,103,104]. The competitive
628 advantage of this strategy diminishes with decreasing salinity (in a logarithmic manner,
629 assuming stable intracellular ion concentrations[105]), and so likely does the relative
630 abundance of bacteria using it. Finally, many protists also have additional organelles
631 with osmoregulating functions, especially the contractile vacuole[106].

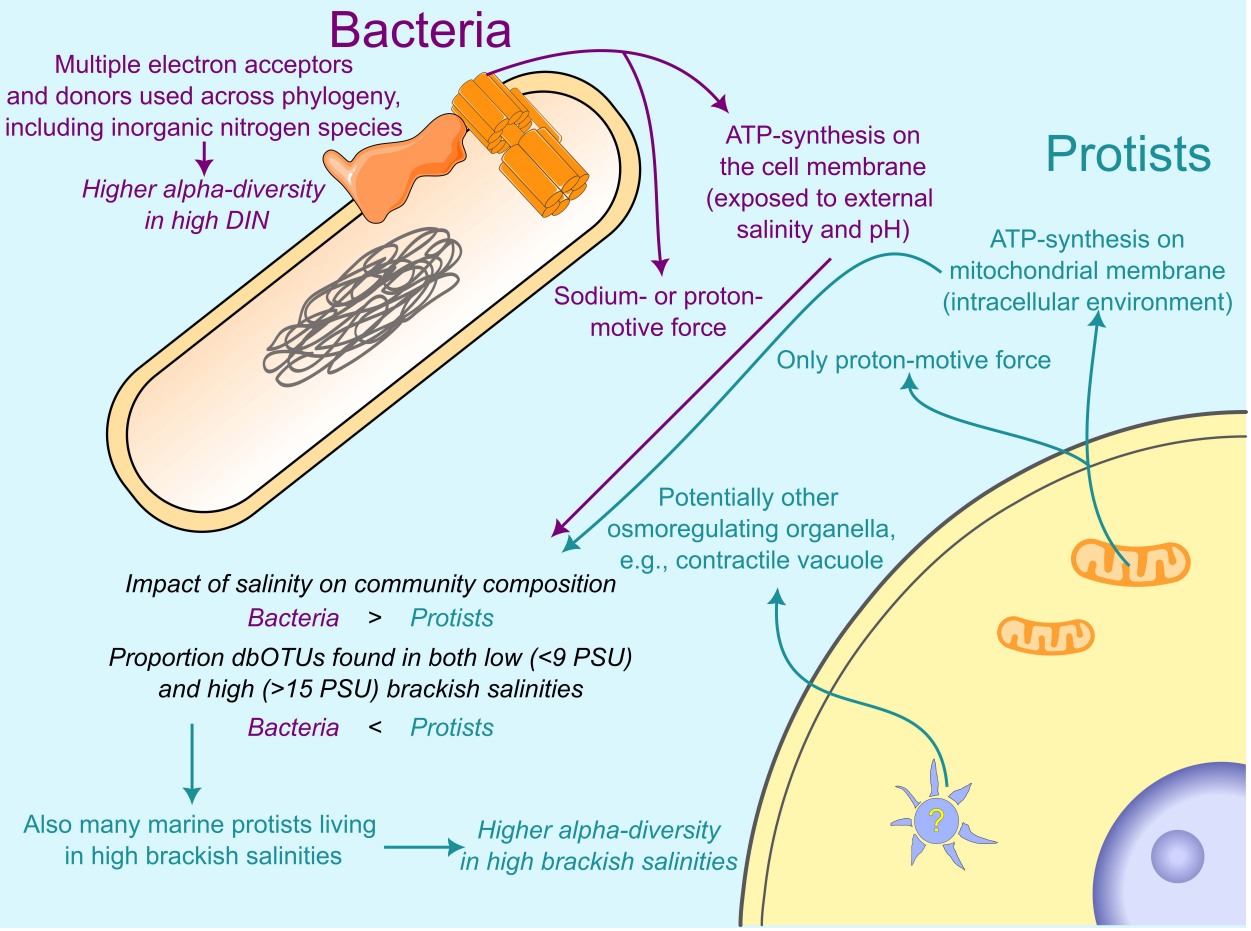


Fig. 8 | The proposed mechanistic explanation for higher ecological sensitivity to salinity for bacteria than for protists. The cells and their components are drawn schematically and do not accurately represent sizes and shapes. All the cell components, except the protist nucleus and bacterial nucleoid, 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[107] 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. We propose that compartmentalization allows protists to acclimatize to a different salinity regime more easily (Fig. 8). Importantly, as further validation is needed, these considerations apply to withstanding different salinities in a competitive environment, and not necessarily to salinity ranges that isolated cultures can withstand. Nor does the proposed explanation apply to adaptation through long-term evolutionary changes. Finally, our theoretical model could be extended to pH, as we already discussed the bioenergetic aspects in this context (Fig.

8). Consistently, in soil, pH has a dominant effect on the community composition of bacteria but not fungi[108].

We found that nitrogen-rich waters host higher numbers of bacterial taxa. Furthermore, we connected this increase in alpha diversity to an overrepresentation of nitrite oxidizers and denitrifiers. As eutrophication ultimately leads to an increase in harmful algal blooms and subsequent nutrient depletion[109], the conclusion should not be that adding nitrogen increases bacterial diversity. Instead, what requires more focus is the nitrogen cycle-related roles of the rare taxa from nitrogen-rich waters. A monitoring scheme sensitive to their presence is needed to understand their susceptibility to environmental disruptions and subsequent impacts. DNA-based methods may well be the only feasible option at present.

We show geographic differences not only in microbial community composition but also in the seasonal dynamics of closely related taxa, especially protists. These differences occur at both broad and more fine-grained taxonomic levels. Thus, time-series results from multiple and diverse locations should be integrated for generalizability and proper interpretation. We compared a publicly available DNA metabarcoding dataset with newly sequenced samples from an earlier monitoring effort. We observed interannual stability of community composition relative to the environmental conditions. However, minor technical differences between the datasets confounded alpha diversity analysis. Thus, a standardized DNA-based scheme is needed to monitor rare microbial taxa.

Using DNA metabarcoding, we could connect bacterial and protist diversity to the ecosystem processes in the Baltic Sea. Apart from detecting major already-known dynamics, we highlight overlooked taxa that may have crucial roles in nutrient cycling in brackish waters. Moreover, we show how salinity and inorganic nitrogen availability affect microbial diversity, with conclusions generalizable beyond the coastal waters. Still, DNA metabarcoding allows only the monitoring of phylogenetic diversity at a limited resolution. Metagenomics and other omics methods can uncover full-scope genetic diversity and its functional potential.

Ultimately, as Earth's environments undergo rapid change, microorganisms from diverse, often understudied groups will be impacted and/or consequential for the future trajectory of coastal and other ecosystems. DNA-based monitoring covers the phylogenetic scope needed to understand and respond to those unprecedented changes.

Methods:

Data collection and sequencing

The 2019-2020 data comes from Latz et al. 2024[39]. The collection and sequencing procedures are described in detail therein. In brief, samples were collected monthly or every second week, depending on the location, as a part of the Swedish National Marine Monitoring Program. The samples for metabarcoding were collected using a depth-integrating hose, covering 0-10m (except 0-5m at RÅNEÅ-1, and 0-20m at B1 and BY31). 500 ml of water from the hose sampling was filtered using 0.22 µm pore size. DNA was extracted using the Zymobiomics™ DNA miniprep kit following a modified protocol[110], and spike-in DNA was added. The DNA was sequenced using Illumina™ MiSeq flow cells. For 16S rRNA gene-based metabarcoding of prokaryotes, primers targeting the hypervariable V3-V4 regions[17], and for 18S metabarcoding of eukaryotes, primers targeting the V4 region[111], were used. Phased primers[112] were used for both 18S (on the forward primer) and 16S (on both primers) metabarcoding. Sequences of the primers and the details of the phasing strategy and sample indexing through a second PCR can be found in Latz et al. 2024[39]. Physico-chemical and geographic data has been obtained from the Swedish Meteorology and Hydrology Insitute's service SHARKweb: https://sharkweb.smhi.se/hamta-data/

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

1. Volumes varied from 200ml to 800ml. According to our previous tests run in the same setting, volumes ≥200ml should capture the microbial community structure equally well[39].
2. DNA extraction was performed using Qiagen™ DNeasy PowerWater Kit, and not ZymoBIOMICS™ DNA Miniprep Kit with a small modification[110] as for the 2019-2020 dataset.
3. No spike-in DNA was added.
4. Illumina™ NextSeq, rather than MiSeq system, was used.
5. For 18S, the initial use of 20 cycles for the first PCR amplification was insufficient, leading to many library preparation failures. It was then re-run with 24 cycles, and the results of both runs were merged after processing the tables by summing up the counts in the output count table.
6. For 18S, reverse primers were also phased, following a sequence (CTACGA)CTTTCGTTCTTGATYRR, with versions of the primer starting from any base in the brackets included in the primer mix.

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

722 Processing of sequencing data

723 The data processing for 2019-2020 samples has been described in detail in Latz et al.
724 2024[39]. In brief, the phased primer sequences were removed from the obtained reads
725 using a modified version of a cutadapt[113]-based pipeline (original pipeline taken from:
726 https://github.com/biodiversitydata-se/amplicon-multi-cutadapt). DADA2[114] version
727 1.18.0 was used to denoise the data, infer amplicon sequence variants (ASVs), and
728 taxonomically annotate them. For taxonomic annotation of 16S ASVs, we used 16S
729 sequences from GTDB[115] version R06-RS202-1 corrected for mislabeled sequences
730 using SATIVA[116,117]. Annotation was additionally conducted using the SILVA
731 database[118] version 138.1 to identify plastids and mitochondria. The 18S ASVs were
732 annotated using the PR2 database[75] version 5.0.1.
733 The 2015-2017 data was analyzed the same way, with the only difference being
734 that the denoising and ASV inference with DADA2 was performed using the nf-
735 core/ampliseq pipeline[119] version 2.7.0 instead of an in-house R script. We kept the
736 DADA2 parameters as in Latz et al. 2024[39].
737 We further combined the 2015-2017 and 2019-2020 data. We matched ASVs
738 using exact matching. On the merged ASV-sequence list and count table, we ran an
739 additional chimera removal step using UCHIME v1[120] and clustered the remaining ASVs
740 based on their simultaneous genetic and distribution similarity using dbOTU3[40]. Both
741 these steps were performed using an in-house pipeline available at:
742 https://github.com/kjurdzinski/chimera_dbOTU_pipeline. Default parameters were used
743 for UCHIME, and for dbOTU#, only the maximum genetic distance for 16S has been
744 changed to 3% (instead of the default 10%). The procedure should minimize the
745 influence of intra-specific variation on our data and further reduce the effects of PCR
746 amplification and sequencing errors. We called thus obtained taxonomic units dbOTUs
747 (distance-based operational taxonomic units, as inferred using dbOTU3[40])
748 For both 16S and 18S, we removed dbOTUs with no annotation at the
749 phylum/supergroup taxonomic level. From 16S dbOTUs, *Archaea*, plastids, and
750 mitochondria were removed. Thus, the remaining ASVs should correspond to bacteria.
751 From 18S ASVs, animals (*Metazoa*), *Fungi*, land plants (*Embryophyceae*), and
752 macroalgae (Rhodophyta, *Phaeophyceae*, *Cyanidiales*) were removed. Thus, the
753 remaining ASVs were primarily protists.
754 We kept spike-in dbOTUs in a separate table and used them only to obtain spike-
755 in corrected abundance, i.e. the inferred number of rRNA copies per liter. The rRNA
756 copier per liter were calculated according to the equation:

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

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

For comparability of results for protists and bacteria, we used only the samples with data available from both 16S and 18S amplicon sequencing (excluding the under- and over-sequenced samples) for all the further analyses. Additionally, in two samples (station RÅNEÅ-2 on 2019-01-29 (sample ID: 20190129_263856_1) and B7 on 2019-09-17 (sample ID: 20190917_264450_1)) 18S spike-in addition failed. Those samples were excluded from analyses in which spike-in normalization was used, except for bacteria-only focused analyses in Fig. 1b and Fig. 3a-b, g-i.

Abundance, alpha and beta diversity spatiotemporal patterns

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

All the statistical analyses and data visualization was performed using R version 4.2.2[121]. Inkscape[122] was used to assemble multi-panel figures and to make Fig. 8.

The multivariate analyses were performed using vegan R package[123], and all the functions in this paragraph can be found there unless another package is explicitly stated. Bray–Curtis distances between communities were obtained using rarefaction via the avgdist function. Those distances were used for Principal Coordinate Analysis (PCoA) and distance-based Redundancy Analysis (dbRDA), performed with, respectively, capscale (without setting constraints) and dbrda functions. Variance partitioning was performed using rdacca.hp R package[124]. Families were positioned on the ordination plots by calculating weighted averages (sppscores function). Only the families deviating (Bonferroni corrected P-value < 0.05) from normal distribution around

the center of the dbRDA coordinate system were shown on the plots. To find the corresponding two-dimensional ellipsoid confidence intervals, aspace[125] package was used. Whenever only samples were visualized, their ordination scores were scaled using the eigenvalues, and symmetric scaling was used if both the families and the samples were visualized (using, respectively, “sites” and “symmetric” scaling options of the plotting functions respectively). For dbRDA and variance partitioning, we used a subset of 228 (out of 246) samples with all the environmental variables of interest available. For more informative visualization, to keep the sample structuring visible despite very high family scores, we divided the family scores in half.

dbOTU richness was calculated as the mean number of dbOTUs detected in a sample over 100 rarefaction iterations. Variance partitioning was performed using rdacca.hp R package[124]. dbOTUs differentially present in high/low DIN/salinity were identified using the Wilcoxon signed-rank test based on presence/absence data. The families overrepresented among these dbOTUs were found using a hypergeometric test (Fisher’s exact test).

Co-occurrence analysis

dbOTUs found in at least ten samples across the 2015-2017 and 2019-2020 datasets were considered for the co-occurrence analysis. Their presence/absence data was used to calculate the proportion of samples in which they were co-present relative to the number of samples in which the more widespread dbOTU was found:

$$co_presence_{i,j} = \frac{N_{i\,\&\,j}}{max(N_i, N_j)}$$

Where *i* and *j* stand for two different dbOTUs, $N_{i\,\&\,j}$ is the number of samples in which both dbOTUs were found, and N_i and N_j are the total numbers of samples in which each dbOTU was present.

Additionally, a co-absence value was calculated, also relative to the more widespread dbOTU:

$$co_absence_{i,j} = \frac{N_{\neg i\,\&\neg\,j}}{min(N_{\neg i}, N_{\neg j})}$$

where $N_{\neg i\,\&\neg\,j}$ is the number of samples in which neither dbOTUs was found, and $N_{\neg i}$ and $N_{\neg j}$ are the total numbers of samples in which each dbOTU was absent.

We used the beta-distribution-based test (pbeta in R) to identify the co-occurrence and co-absence values significantly different from an average pair of dbOTUs (FDR < 0.01 for each of the two values, in this case corresponding to the less common dbOTU being present in at least 86% of the samples in which the more common one was found and absent from all the samples where the more common one

was not found). The co-occurrence network was visualized using the igraph R package[126].

Proportion salinity barrier crossing dbOTUs and Bayesian modelling

Any dbOTU found among included samples with salinity >15 PSU and among samples with salinity <9 PSU was regarded as barrier crossing. For balancing out the number of samples on both sides of the salinity barrier, apart from the four stations with salinity >15 PSU (Å17, SLÄGGÖ), four out of twelve stations with salinity <9 PSU: BY2 ARKONA, BY5 BORNHOLMSDJ, BY15 GOTLANDSDJ, REF M1V1 (see Fig. 1a for locations of the stations). First, we chose the four most saline stations from among those with salinity <9 PSU. Then, we switched BCS III-10 to REF M1V1 (see Fig. 1), since the latter had more samples, and, unlike other chosen stations from lower salinities, was located close to the mainland coastline. For each station but the least sampled (REF M1V1), the samples were downsampled, choosing samples closest in time to the ones collected at the least sampled stations. Whenever available, those samples should correspond to the same monitoring cruises.

The Bayesian models were built and implemented in JAGS [99], using the R package rjags[127]. The parameters, probabilities, and predictions were calculated based on every fifth out of 10000 MCMC iterations.

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

$$if_crossed_barrier_i \sim Bernoulli(p_i)$$

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

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

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[100], 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 the grant number NV-03728-17. AFA was additionally funded by a research grant (2021-05563) from the Swedish Research Council VR.

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.

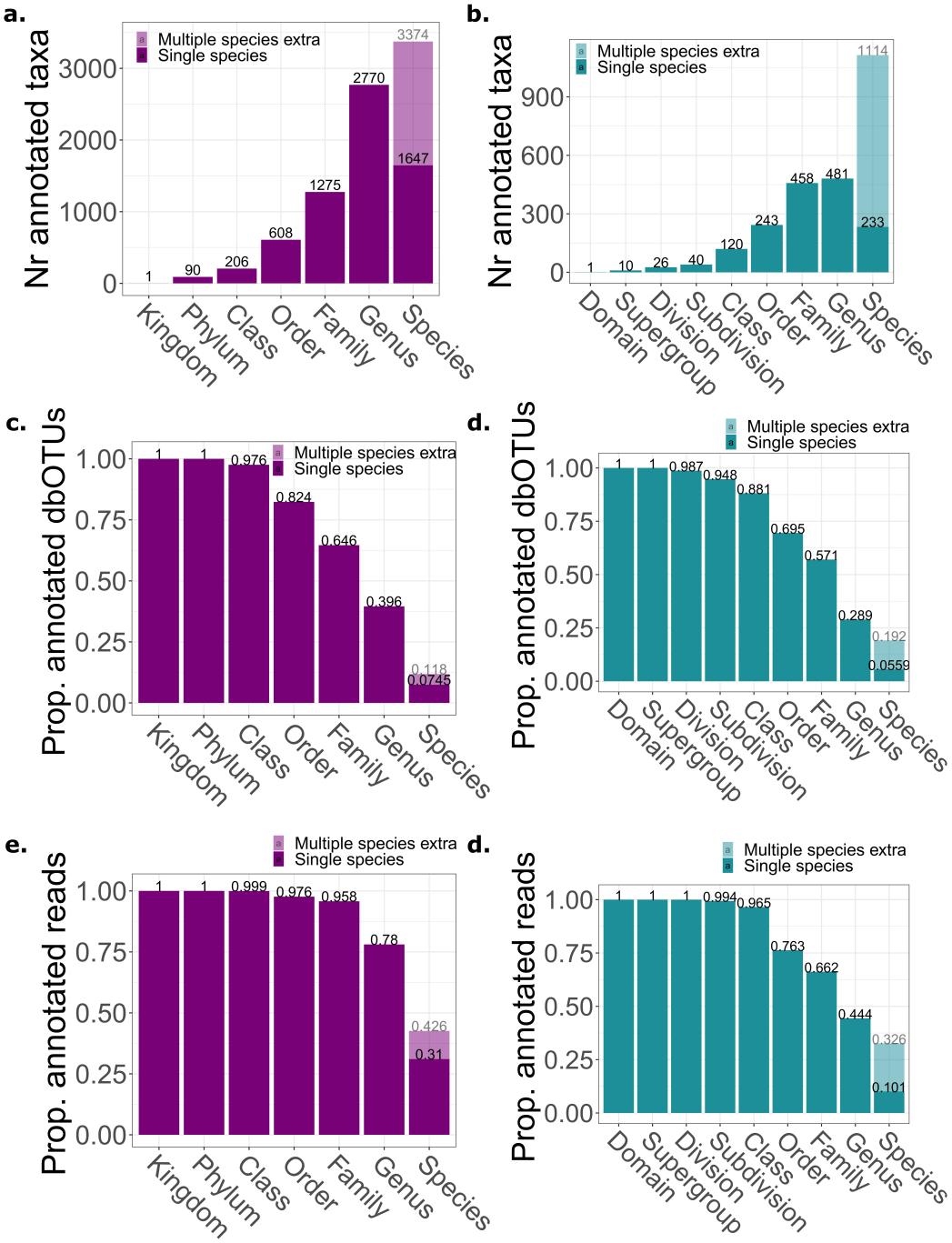
Author contributions:

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.

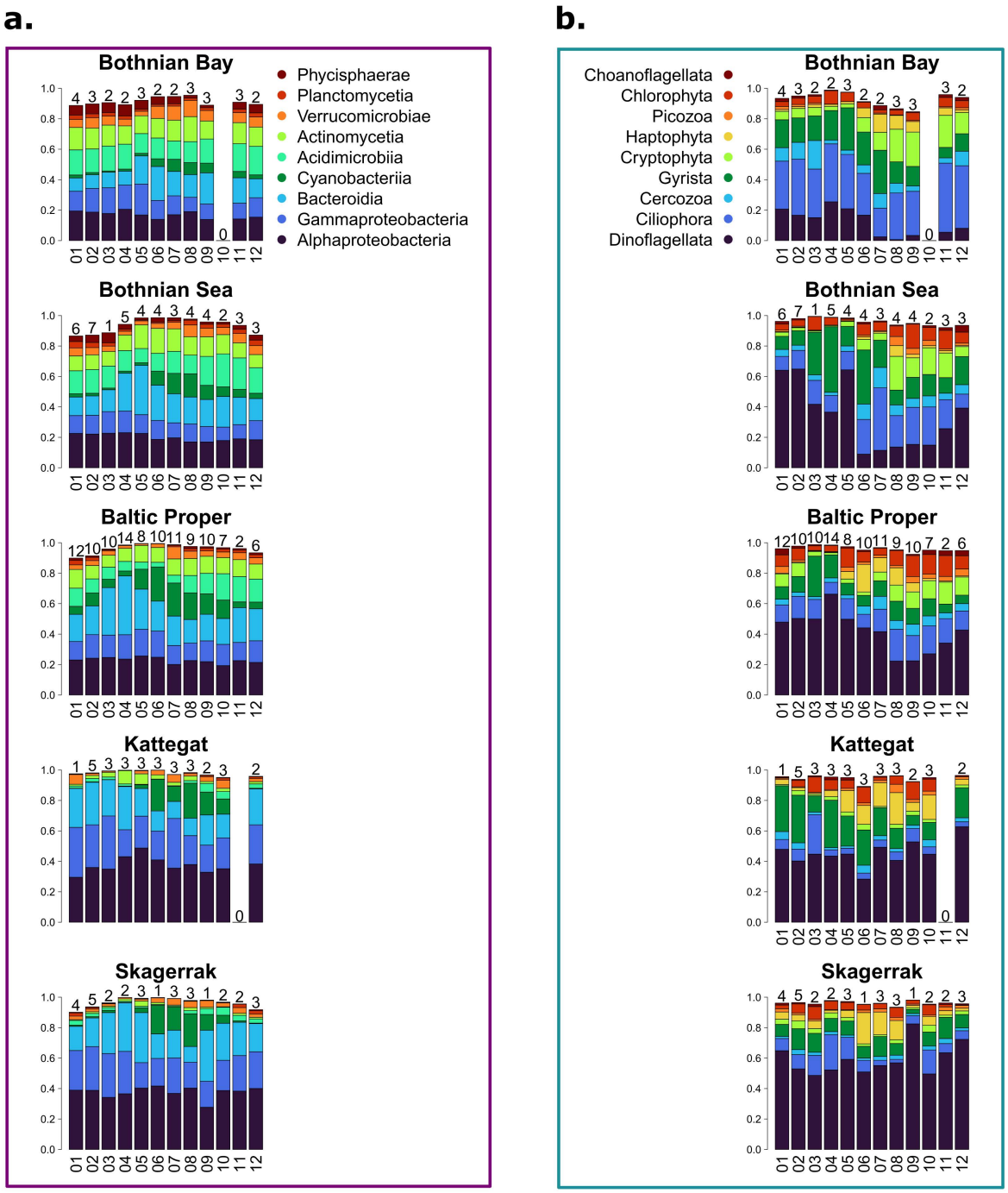
Data and code availability:

This is a preliminary version of the manuscript. We are still working on making the data and code publicly available. The 2019-2020 data (reference) is available at the European Nucleotide Archive (ENA) under the study accession number https://identifiers.org/ena.embl:PRJEB55296.

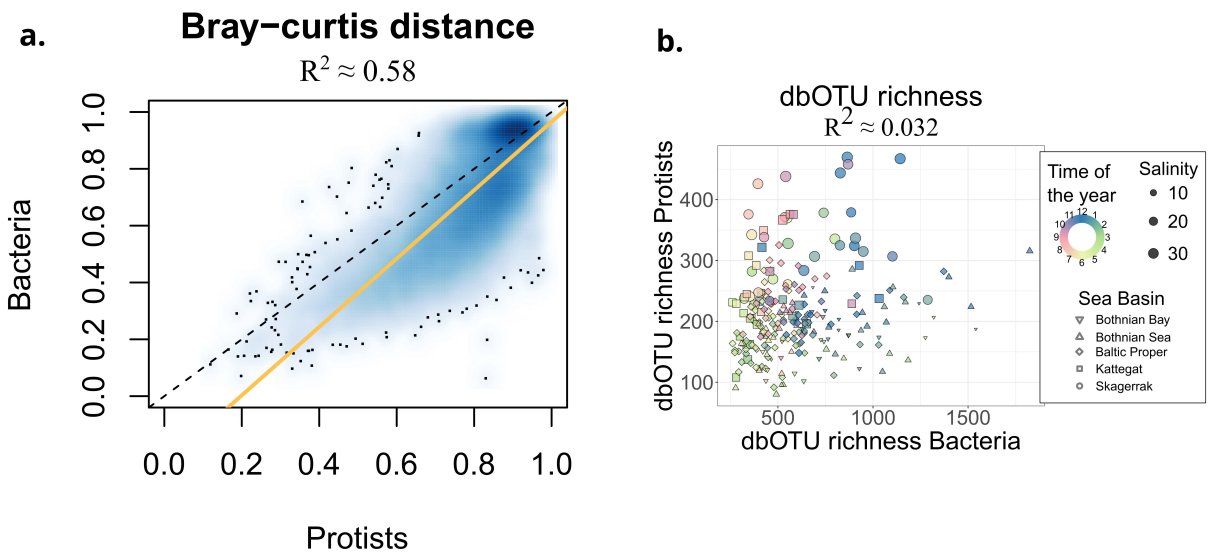
Supplementary figures:



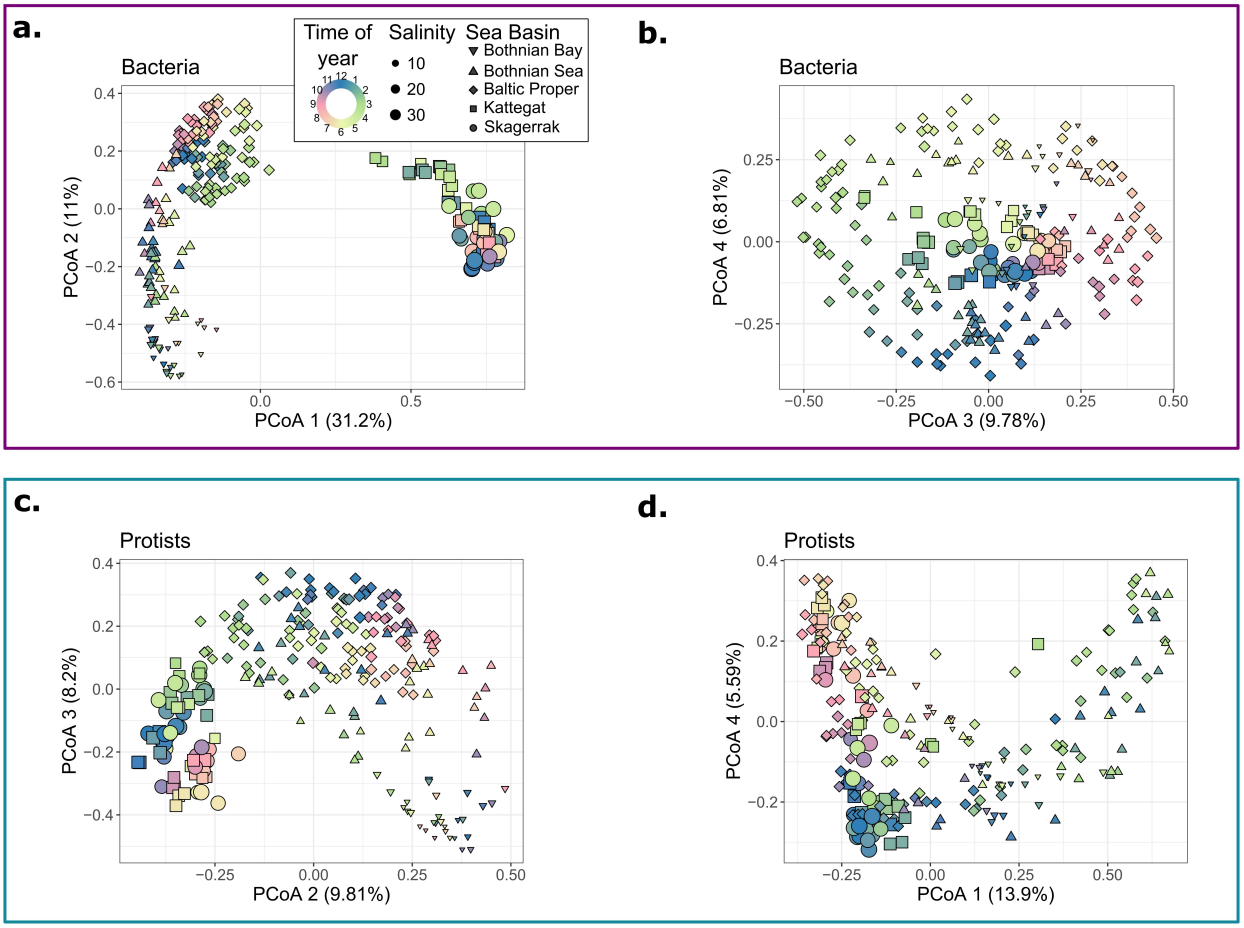
Supplementary Fig. S1 | Taxonomic annotation statistics a-b. Number of different bacterial (**a.**) and protist (**b.**) taxa found at different taxonomic levels. **c-d**. Proportions of bacterial (**c.**) and protist (**d.**) taxa annotated at different taxonomic levels. **e-f**. Proportions of bacterial (**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.



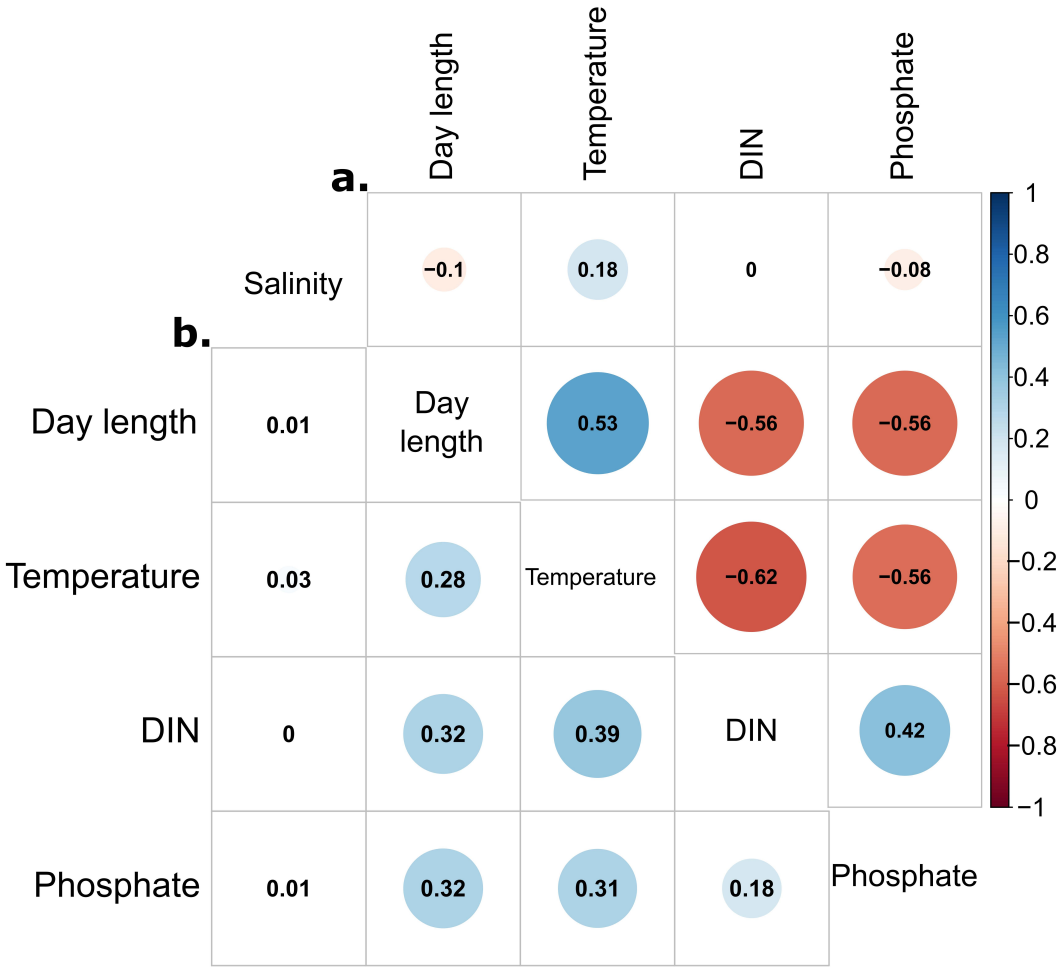
Supplementary Fig. S2 | Spatiotemporal changes in abundance of major microbial taxa, 2019-2020 a-b. The proportion of rRNA marker gene reads annotated to **a.** bacterial classes and **b.** protist subdivisions averaged over each month and Baltic Sea area basin. Only the classes/subdivisions which, on 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.



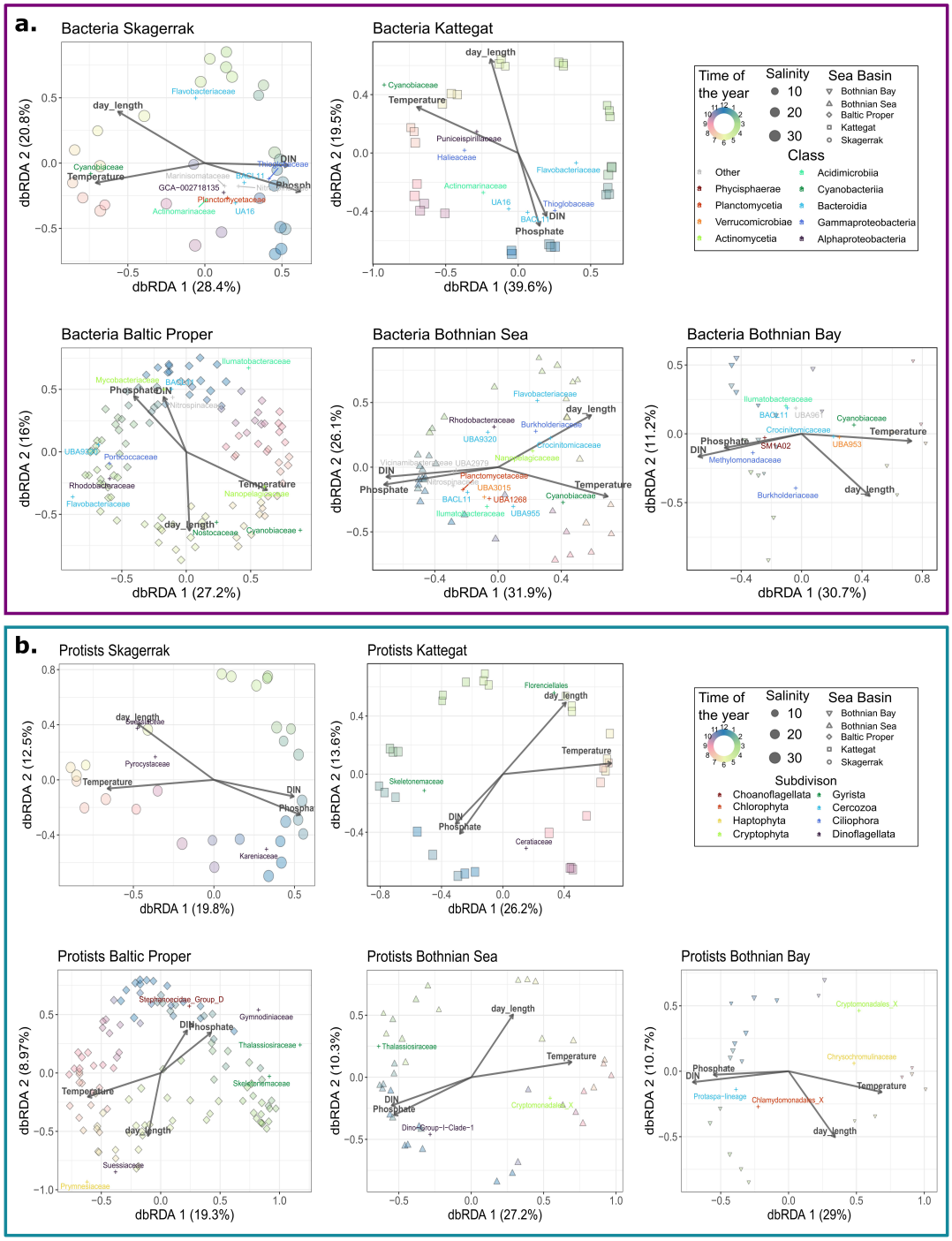
Supplementary Fig. S3 | Dissimilarity of diversity metrics between bacteria and protist, 2019-2020 a. Beta diversity calculated as Bray-Curtis distances. The golden line represents the best linear fit, while the dashed black line is the identity line. **b.** Number of detected dbOTUs (alpha diversity).



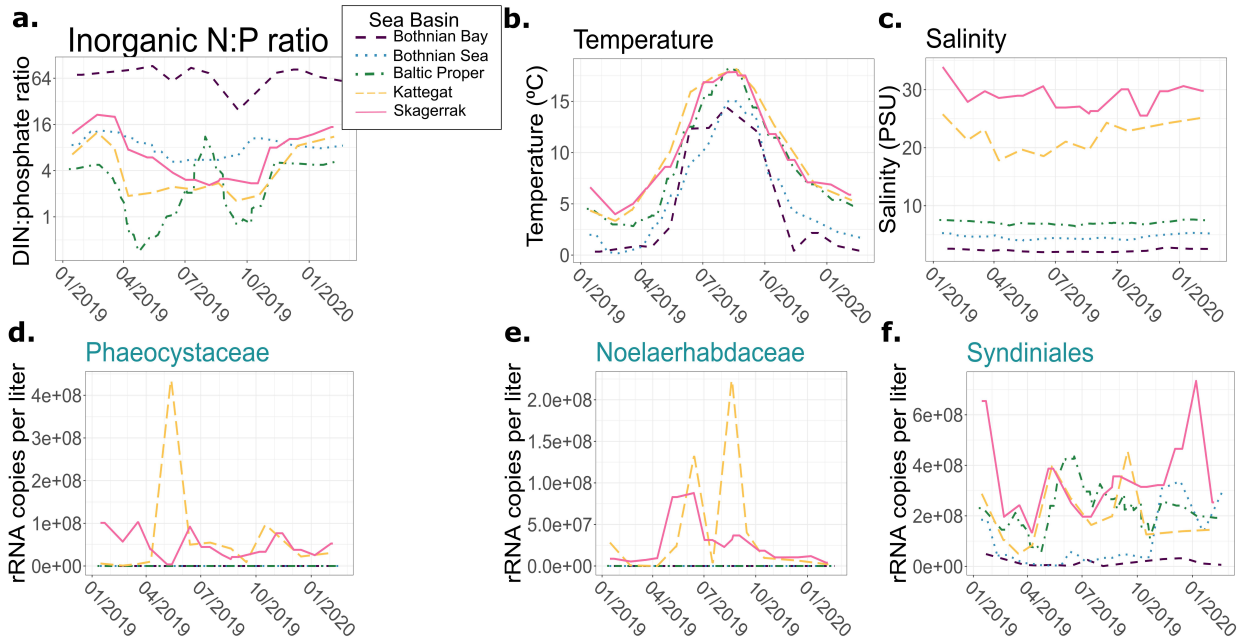
Supplementary Fig. S4 | Principal coordinate analysis (PCoA) 2019-2020. The first four ordination axes are based on Bray-Curtis distances between bacterial (**a-b**) and protist (**c-d**) communities. Note: The axes are grouped and ordered by which ones, after visual examination, correspond to geographic 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.



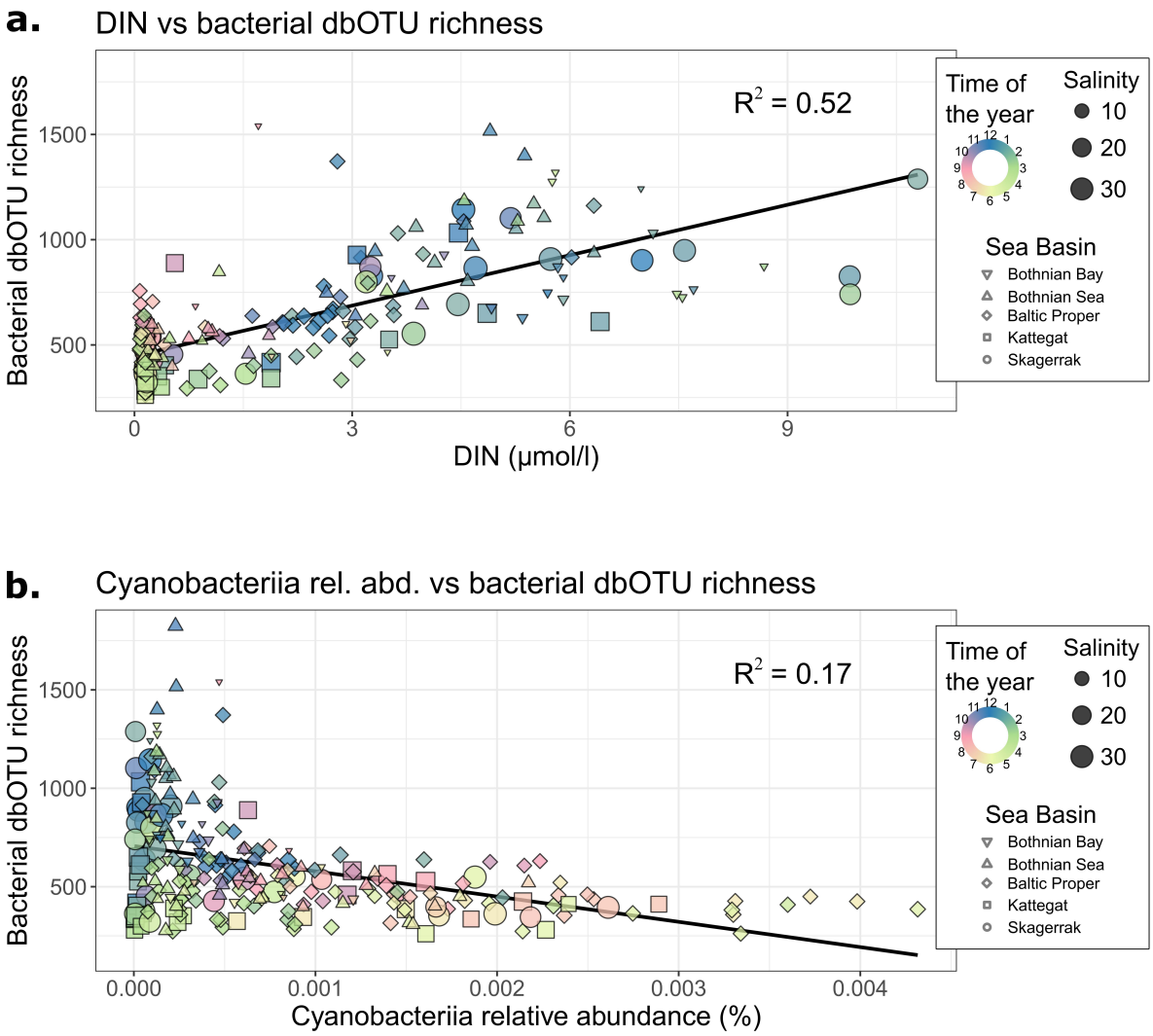
Supplementary Fig. S5 | Correlations between chosen environmental variables for the 2019-2020 dataset a. Pearson correlation coefficients (ρ) on an upper-triangle correlation plot. **b.** Coefficients of determination (R^2) on a lower-triangle correlation plot.



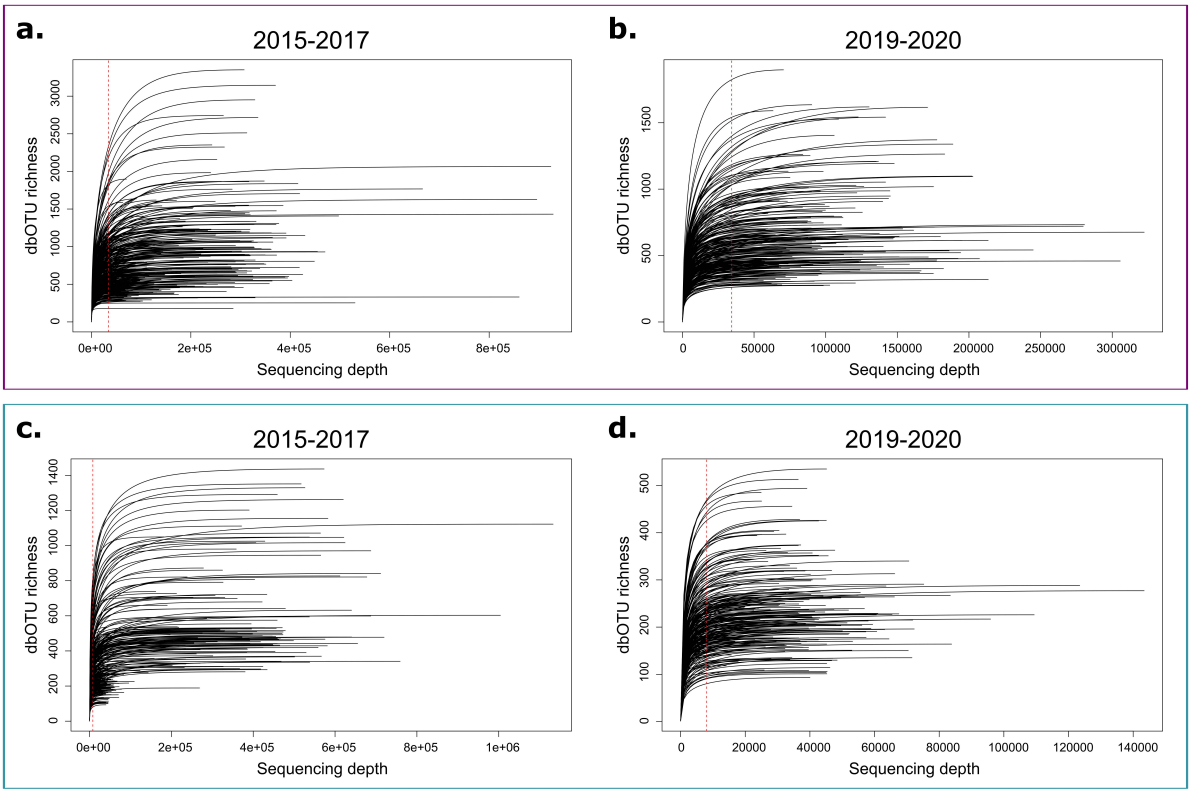
Supplementary Fig. S6 | Seasonal changes in community composition within distinguished basins of the Baltic Sea area. distance-based Redundancy Analysis (dbRDA) based on Bray-Curtis distances between the bacterial (**a.**) and protist (**b.**) communities within each of the distinguished basins of the 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.



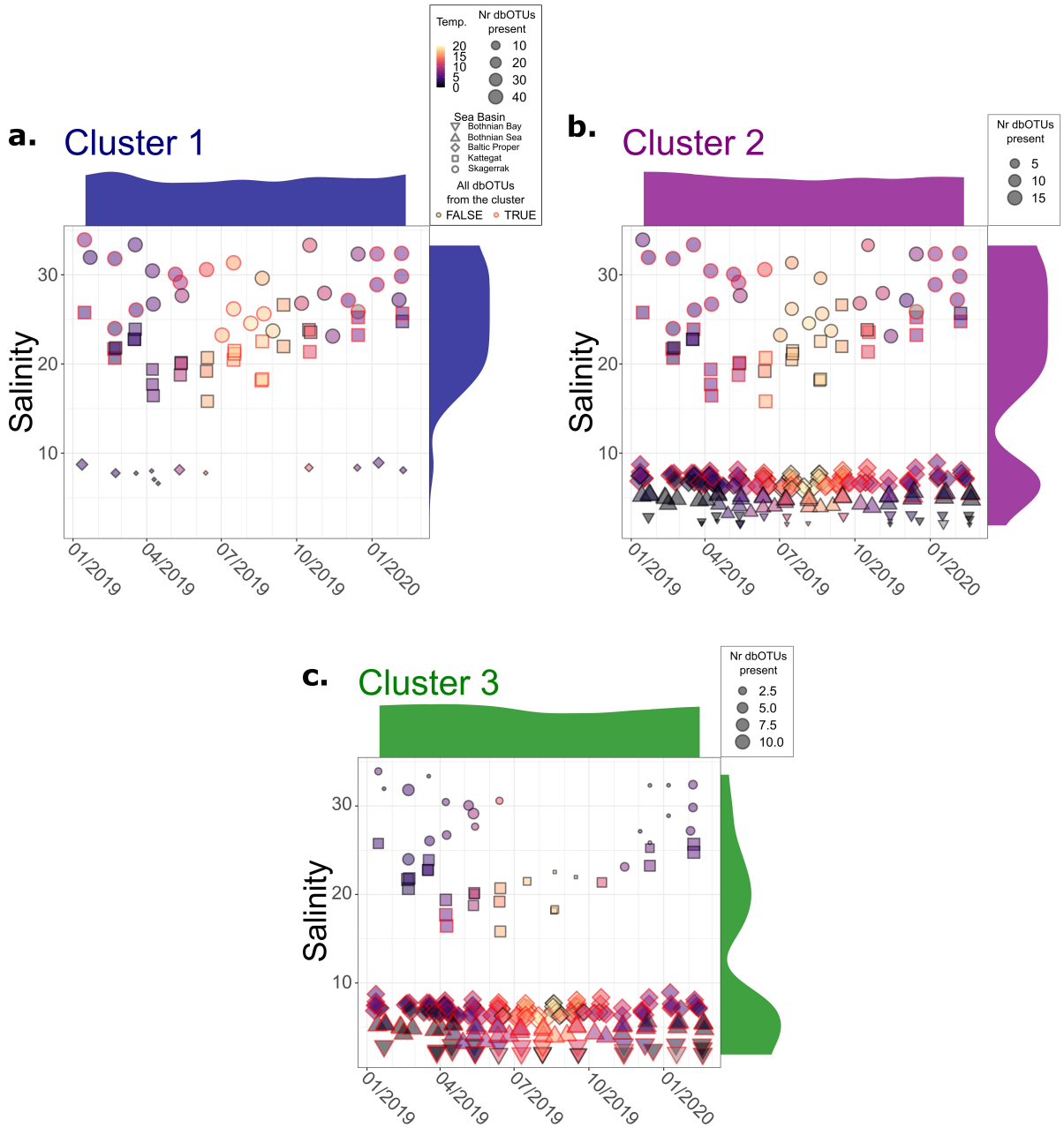
Supplementary Fig. S7 | The seasonal dynamics of inorganic N:P ratio, salinity, and additional protist taxa. a. DIN (dissolved inorganic nitrogen) to phosphate concentration ratio across the 2019-2020 samples. **b-c.** Temperature (**b.**) and salinity (**c.**) across the 2019-2020 samples. **d-e**. Spike-in normalized abundance of the haptophyte families *Phaeocystaceae* and *Noelaerhabdaceae*. **f.** Spike-in normalized abundance of the dinoflagellate class *Sindiniales*. The lines correspond to the rolling mean, with values averaged for measurements two weeks before and after each date. The same legend (a.) applies to all the panels, with basins ordered from lowest to highest salinity (top to bottom). PSU - practical salinity units



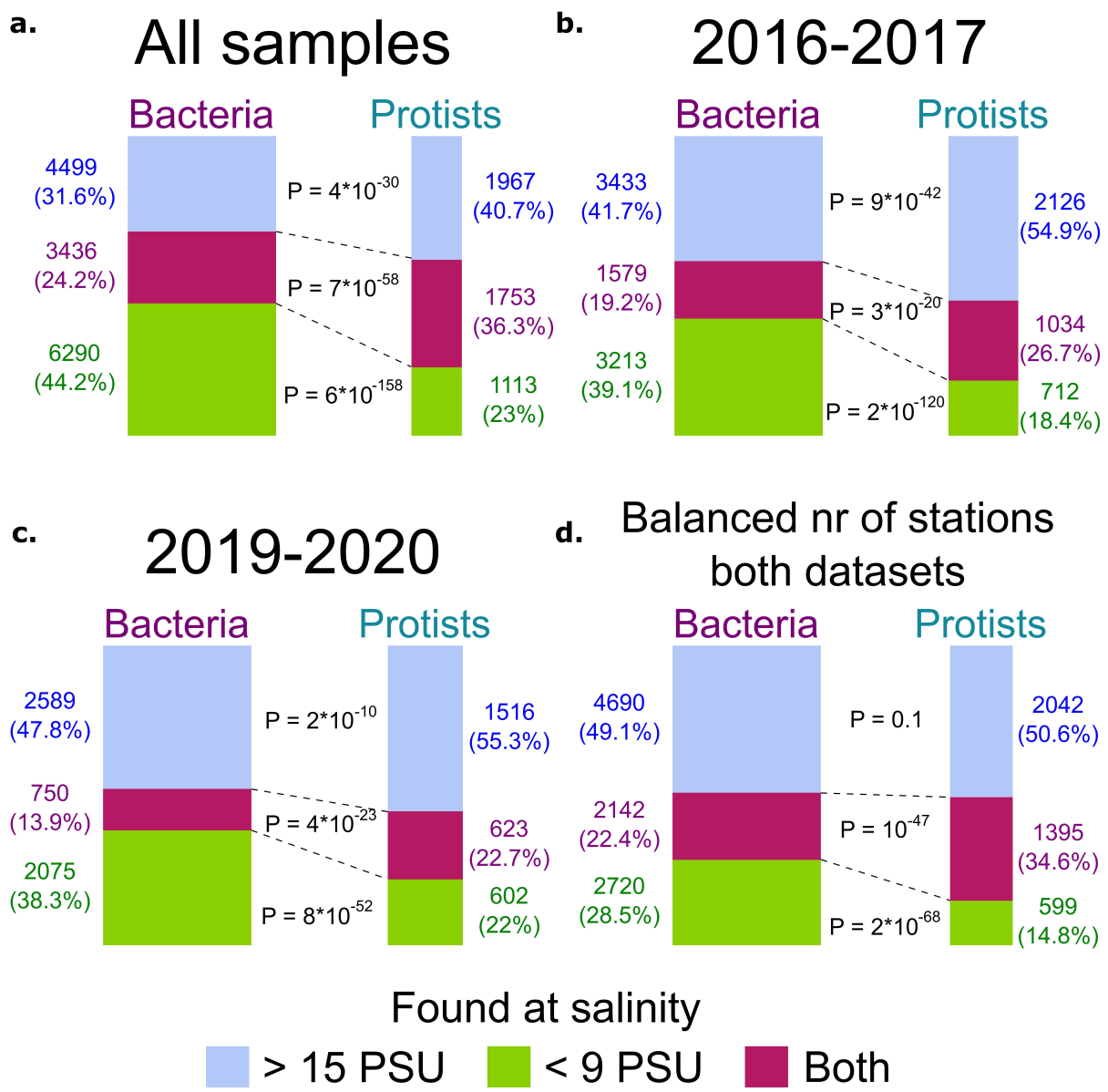
Supplementary Fig S8 | Correlation of bacterial db OTU richness a. with dissolved inorganic nitrogen (DIN) **b.** relative abundance of cyanobacteria.



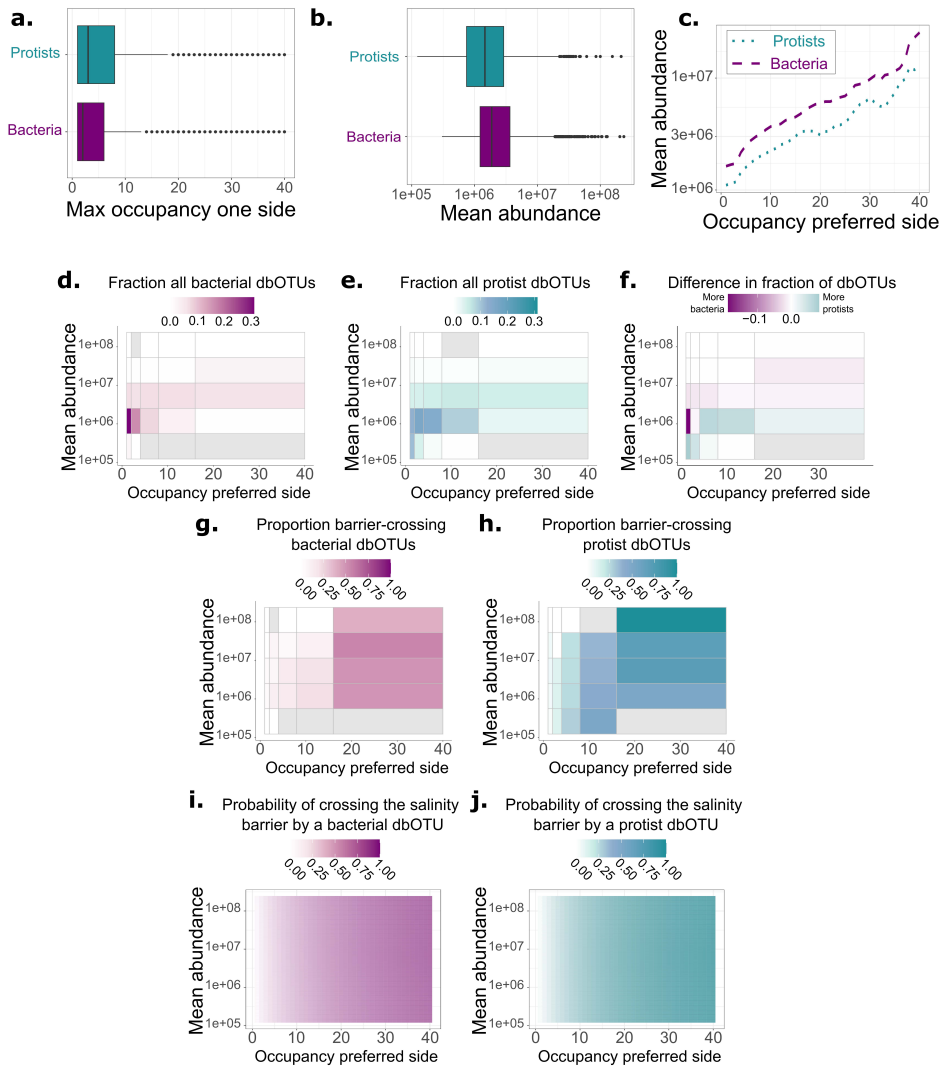
Supplementary Fig. S9 | Refraction curves a-b. Based on 16S filtered (only bacteria) read from the 2015-2017 (**a.**) and 2019-2020 (**b.**) datests. **c-d**. Based on 18S filtered (only protists) read from the 2015-2017 (**c.**) and 2019-2020 (**d.**) datests. dbOTU - distribution-based operational taxonomic unit.



Supplementary Fig. S10 | Number of dbOTUs from co-occurring clusters (Fig. 6a) found in each sample. The samples in which all the dbOTUs from a cluster were present are outlined in red. On the top and the right axes of each plot are density plots based on mean number of dbOTUs per month (top) and per unit salinity (right). The same legend (**a.**) applies to all the panels except for the size-coding corresponding to the number of dbOTUs from a cluster.



Supplementary Fig. S11 | The number and proportion of dbOTUs found only in high (>15 PSU), low (<9 PSU) brackish salinity, and across the salinity barrier. **a.** Proportions across all samples. **b-c.** Proportions for the samples from 2016-2017 (**b.**) and 2019-2020 (**c.**). Included samples from the two datasets 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. **d.** Proportions across the two datasets with a balanced number of stations and observations across the salinity barrier. Only the four stations with salinity >15 PSU, and four out of the most saline stations with salinity <9 PSU were chosen. The data for each station was downsampled to the same number of observations. Samples closest in time were chosen. The same legend applies to all the panels.



Supplementary Fig. S12 | 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 maximal occupancy at one side of the salinity barrier (**a.**) and mean abundance (**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 rRNA 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. Lundsør, E. *et al.* Marine phytoplankton community data and corresponding environmental properties from eastern Norway, 1896-2020. *Sci Data* **9**, 767 (2022).
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. SHARK. SHARK - National marine environmental monitoring of Phytoplankton in Sweden since 1983. The Swedish Meteorological and Hydrological Institute https://doi.org/10.15468/S3YG8Z (2022).
16. 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).
17. 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).
18. Hu, Y. O. O., Karlson, B., Charvet, S. & Andersson, A. F. Diversity of Pico- to Mesoplankton along the 2000 km Salinity Gradient of the Baltic Sea. *Front. Microbiol.* **7**, 679 (2016).
19. Dupont, C. L. *et al.* Functional tradeoffs underpin salinity-driven divergence in microbial community composition. *PLoS One* **9**, e89549 (2014).
20. Hugerth, L. W. *et al.* Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biol.* **16**, 279 (2015).
21. 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).
22. Wu, Q. L., Zwart, G., Schauer, M., Kamst-van Agterveld, M. P. & Hahn, M. W. Bacterioplankton community composition along a salinity gradient of sixteen high-mountain lakes located on the Tibetan Plateau, China. *Appl. Environ. Microbiol.* **72**, 5478–5485

1053 (2006).
1054 23. Lozupone, C. A. & Knight, R. Global patterns in bacterial diversity. *Proc. Natl. Acad. Sci. U.*
1055 *S. A.* **104**, 11436–11440 (2007).
1056 24. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of
1057 sequences per sample. *Proc. Natl. Acad. Sci. U. S. A.* **108 Suppl 1**, 4516–4522 (2011).
1058 25. Singer, D. *et al.* Protist taxonomic and functional diversity in soil, freshwater and marine
1059 ecosystems. *Environ. Int.* **146**, 106262 (2021).
1060 26. Logares, R. *et al.* Infrequent marine-freshwater transitions in the microbial world. *Trends*
1061 *Microbiol.* **17**, 414–422 (2009).
1062 27. Walsh, D. A., Lafontaine, J. & Grossart, H.-P. On the Eco-Evolutionary Relationships of
1063 Fresh and Salt Water Bacteria and the Role of Gene Transfer in Their Adaptation. in *Lateral*
1064 *Gene Transfer in Evolution* (ed. Gophna, U.) 55–77 (Springer New York, New York, NY,
1065 2013).
1066 28. Burki, F., Sandin, M. M. & Jamy, M. Diversity and ecology of protists revealed by
1067 metabarcoding. *Curr. Biol.* **31**, R1267–R1280 (2021).
1068 29. Jurdzinski, K. T. *et al.* Large-scale phylogenomics of aquatic bacteria reveal molecular
1069 mechanisms for adaptation to salinity. *Sci Adv* **9**, eadg2059 (2023).
1070 30. Telesh, I. V., Schubert, H. & Skarlato, S. O. Revisiting Remane's concept: evidence for high
1071 plankton diversity and a protistan species maximum in the horohalinicum of the Baltic Sea.
1072 *Mar. Ecol. Prog. Ser.* **421**, 1–11 (2011).
1073 31. Olli, K., Ptacnik, R., Klais, R. & Tamminen, T. Phytoplankton Species Richness along
1074 Coastal and Estuarine Salinity Continua. *Am. Nat.* **194**, E41–E51 (2019).
1075 32. Olofsson, M., Hagan, J. G., Karlson, B. & Gamfeldt, L. Large seasonal and spatial variation
1076 in nano- and microphytoplankton diversity along a Baltic Sea-North Sea salinity gradient.
1077 *Sci. Rep.* **10**, 17666 (2020).
1078 33. Herlemann, D. P. R., Lundin, D., Andersson, A. F., Labrenz, M. & Jürgens, K. Phylogenetic

Signals of Salinity and Season in Bacterial Community Composition Across the Salinity Gradient of the Baltic Sea. *Front. Microbiol.* **7**, 1883 (2016).

34. Remane, A. Die Brackwasserfauna : mit besonderer Berücksichtigung der Ostsee. *Zool. Anz.* **Suppl. 7**, (1934).
35. Lindh, M. V. *et al.* Disentangling seasonal bacterioplankton population dynamics by high-frequency sampling. *Environ. Microbiol.* **17**, 2459–2476 (2015).
36. Gilbert, J. A. *et al.* Defining seasonal marine microbial community dynamics. *ISME J.* **6**, 298–308 (2012).
37. Ladau, J. *et al.* Global marine bacterial diversity peaks at high latitudes in winter. *ISME J.* **7**, 1669–1677 (2013).
38. Raes, E. J. *et al.* Seasonal patterns of microbial diversity across the world oceans. *Limnol. Oceanogr. Lett.* (2024) doi:10.1002/lol2.10422.
39. Latz, M. A. C. *et al.* A comprehensive dataset on spatiotemporal variation of microbial plankton communities in the Baltic Sea. *Sci Data* **11**, 18 (2024).
40. Olesen, S. W., Duvallet, C. & Alm, E. J. dbOTU3: A new implementation of distribution-based OTU calling. *PLoS One* **12**, e0176335 (2017).
41. Goldford, J. E. *et al.* Emergent simplicity in microbial community assembly. *Science* **361**, 469–474 (2018).
42. Tourlousse, D. M. *et al.* Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. *Nucleic Acids Res.* **45**, e23 (2017).
43. Lloréns-Rico, V., Vieira-Silva, S., Gonçalves, P. J., Falony, G. & Raes, J. Benchmarking microbiome transformations favors experimental quantitative approaches to address compositionality and sampling depth biases. *Nat. Commun.* **12**, 3562 (2021).
44. Wesslander, K., Viktorsson, L., Thor, P., Nilsson, M. & Skjevik, A.-T. *The Swedish National Marine Monitoring Programme 2019 : Hydrography Nutrients Phytoplankton*. 59 (2020).
45. Zettler, M. L., Kremp, A. & Dutz, J. Biological assessment of the Baltic Sea 2019.

Meereswiss. Ber. , Warnemünde **115**, (2020).

46. Fridolfsson, E. *et al.* Multiyear analysis uncovers coordinated seasonality in stocks and composition of the planktonic food web in the Baltic Sea proper. *Sci. Rep.* **13**, 11865 (2023).
47. Logares, R. *et al.* Disentangling the mechanisms shaping the surface ocean microbiota. *Microbiome* **8**, 55 (2020).
48. Farjalla, V. F. *et al.* Ecological determinism increases with organism size. *Ecology* **93**, 1752–1759 (2012).
49. Wu, W. *et al.* Contrasting the relative importance of species sorting and dispersal limitation in shaping marine bacterial versus protist communities. *ISME J.* **12**, 485–494 (2018).
50. Neuenschwander, S. M., Ghai, R., Pernthaler, J. & Salcher, M. M. Microdiversification in genome-streamlined ubiquitous freshwater Actinobacteria. *ISME J.* **12**, 185–198 (2018).
51. Mehrshad, M. *et al.* Hidden in plain sight—highly abundant and diverse planktonic freshwater Chloroflexi. *Microbiome* **6**, 176 (2018).
52. 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).
53. Morris, R. M. & Spietz, R. L. The Physiology and Biogeochemistry of SUP05. *Ann. Rev. Mar. Sci.* **14**, 261–275 (2022).
54. 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).
55. Karlson, A. M. L. *et al.* Nitrogen fixation by cyanobacteria stimulates production in Baltic food webs. *Ambio* **44 Suppl 3**, 413–426 (2015).
56. Aguilera, A. *et al.* Ecophysiological analysis reveals distinct environmental preferences in closely related Baltic Sea picocyanobacteria. *Environ. Microbiol.* **25**, 1674–1695 (2023).

1131 57. Karlson, B. *et al.* Harmful algal blooms and their effects in coastal seas of Northern Europe.
1132 *Harmful Algae* **102**, 101989 (2021).
1133 58. Egge, E. S. *et al.* Seasonal diversity and dynamics of haptophytes in the Skagerrak,
1134 Norway, explored by high-throughput sequencing. *Mol. Ecol.* **24**, 3026–3042 (2015).
1135 59. Rost, B. & Riebesell, U. Coccolithophores and the biological pump: responses to
1136 environmental changes. in *Coccolithophores: From Molecular Processes to Global Impact*
1137 (eds. Thierstein, H. R. & Young, J. R.) 99–125 (Springer Berlin Heidelberg, Berlin,
1138 Heidelberg, 2004).
1139 60. Jordan, R. W. & Chamberlain, A. H. L. Biodiversity among haptophyte algae. *Biodiversity &*
1140 *Conservation* **6**, 131–152 (1997).
1141 61. Lücker, S. & Daims, H. The Family Nitrospinaceae. in *The Prokaryotes: Deltaproteobacteria*
1142 *and Epsilonproteobacteria* (eds. Rosenberg, E., DeLong, E. F., Lory, S., Stackebrandt, E. &
1143 Thompson, F.) 231–237 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2014).
1144 62. Lidbury, I. D. E. A. *et al.* Niche-adaptation in plant-associated Bacteroidetes favours
1145 specialisation in organic phosphorus mineralisation. *ISME J.* **15**, 1040–1055 (2021).
1146 63. Lidbury, I. D. E. A. *et al.* A widely distributed phosphate-insensitive phosphatase presents a
1147 route for rapid organophosphorus remineralization in the biosphere. *Proc. Natl. Acad. Sci.*
1148 *U. S. A.* **119**, (2022).
1149 64. Teeling, H. *et al.* Substrate-controlled succession of marine bacterioplankton populations
1150 induced by a phytoplankton bloom. *Science* **336**, 608–611 (2012).
1151 65. Wiegand, S. *et al.* Taxonomic Re-Classification and Expansion of the Phylum Chloroflexota
1152 Based on over 5000 Genomes and Metagenome-Assembled Genomes. *Microorganisms*
1153 **11**, (2023).
1154 66. Baldani, J. I. *et al.* The Family Rhodospirillaceae. in *The Prokaryotes: Alphaproteobacteria*
1155 *and Betaproteobacteria* (eds. Rosenberg, E., DeLong, E. F., Lory, S., Stackebrandt, E. &
1156 Thompson, F.) 533–618 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2014).

1157 67. Urios, L., Michotey, V., Intertaglia, L., Lesongeur, F. & Lebaron, P. Nisaea denitrificans gen.
1158 nov., sp. nov. and Nisaea nitritireducens sp. nov., two novel members of the class
1159 Alphaproteobacteria from the Mediterranean Sea. *Int. J. Syst. Evol. Microbiol.* **58**, 2336–
1160 2341 (2008).
1161 68. Hawley, A. K. *et al.* Diverse Marinimicrobia bacteria may mediate coupled biogeochemical
1162 cycles along eco-thermodynamic gradients. *Nat. Commun.* **8**, 1507 (2017).
1163 69. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421
1164 (2009).
1165 70. Alldredge, A. L. & Cohen, Y. Can microscale chemical patches persist in the sea?
1166 Microelectrode study of marine snow, fecal pellets. *Science* **235**, 689–691 (1987).
1167 71. Klawonn, I., Bonaglia, S., Brüchert, V. & Ploug, H. Aerobic and anaerobic nitrogen
1168 transformation processes in N2-fixing cyanobacterial aggregates. *ISME J.* **9**, 1456–1466
1169 (2015).
1170 72. Louca, S. The rates of global bacterial and archaeal dispersal. *ISME J.* **16**, 159–167 (2022).
1171 73. de Vargas, C. *et al.* Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean.
1172 *Science* **348**, 1261605 (2015).
1173 74. Guillou, L. *et al.* Widespread occurrence and genetic diversity of marine parasitoids
1174 belonging to Syndiniales (Alveolata). *Environ. Microbiol.* **10**, 3349–3365 (2008).
1175 75. Guillou, L. *et al.* The Protist Ribosomal Reference database (PR2): a catalog of unicellular
1176 eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41**,
1177 D597–604 (2013).
1178 76. Massana, R., del Campo, J., Sieracki, M. E., Audic, S. & Logares, R. Exploring the
1179 uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within
1180 stramenopiles. *ISME J.* **8**, 854–866 (2014).
1181 77. Eikrem, W. *et al.* Florenciella parvula gen. et sp. nov. (Dictyochophyceae,
1182 Heterokontophyta), a small flagellate isolated from the English Channel. *Phycologia* **43**,

658–668 (2004).

78. Eckford-Soper, L. & Daugbjerg, N. The ichthyotoxic genus Pseudochattonella (Dictyochophyceae): Distribution, toxicity, enumeration, ecological impact, succession and life history - A review. *Harmful Algae* **58**, 51–58 (2016).

79. Schoemann, V., Becquevort, S., Stefels, J., Rousseau, V. & Lancelot, C. Phaeocystis blooms in the global ocean and their controlling mechanisms: a review. *J. Sea Res.* **53**, 43–66 (2005).

80. Kaczmarska, I., Beaton, M., Benoit, A. C. & Medlin, L. K. MOLECULAR PHYLOGENY OF SELECTED MEMBERS OF THE ORDER THALASSIOSIRALES (BACILLARIOPHYTA) AND EVOLUTION OF THE FULTOPORTULA[1]. *J. Phycol.* **42**, 121–138 (2006).

81. Alverson, A. J., Jansen, R. K. & Theriot, E. C. Bridging the Rubicon: phylogenetic analysis reveals repeated colonizations of marine and fresh waters by thalassiosiroid diatoms. *Mol. Phylogenet. Evol.* **45**, 193–210 (2007).

82. 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).

83. Yeh, Y.-C. & Fuhrman, J. A. Contrasting diversity patterns of prokaryotes and protists over time and depth at the San-Pedro Ocean Time series. *ISME Commun* **2**, 36 (2022).

84. 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).

85. 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).

86. HELCOM. Baltic Sea Action Plan – 2021 update. Preprint at (2021).

87. 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.*

1209 *Acad. Sci. U. S. A.* **120**, e2211796120 (2023).
1210 88. Schloss, P. D. Rarefaction is currently the best approach to control for uneven sequencing
1211 effort in amplicon sequence analyses. *mSphere* **9**, e0035423 (2024).
1212 89. Giovannoni, S. J. *et al.* Genome streamlining in a cosmopolitan oceanic bacterium. *Science*
1213 **309**, 1242–1245 (2005).
1214 90. Grote, J. *et al.* Streamlining and core genome conservation among highly divergent
1215 members of the SAR11 clade. *MBio* **3**, (2012).
1216 91. Roda-Garcia, J. J., Haro-Moreno, J. M., Rodriguez-Valera, F., Almagro-Moreno, S. &
1217 López-Pérez, M. Single-amplified genomes reveal most streamlined free-living marine
1218 bacteria. *Environ. Microbiol.* **25**, 1136–1154 (2023).
1219 92. Salcher, M. M., Schaefle, D., Kaspar, M., Neuenschwander, S. M. & Ghai, R. Evolution in
1220 action: habitat transition from sediment to the pelagial leads to genome streamlining in
1221 Methylophilaceae. *ISME J.* **13**, 2764–2777 (2019).
1222 93. Bittner, M. J. *et al.* New chemical and microbial perspectives on vitamin B1 and vitamer
1223 dynamics of a coastal system. *ISME Commun* **4**, ycad016 (2024).
1224 94. Parks, D. H. *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a
1225 phylogenetically consistent, rank normalized and complete genome-based taxonomy.
1226 *Nucleic Acids Res.* **50**, D785–D794 (2022).
1227 95. Fullmer, M. S., Soucy, S. M. & Gogarten, J. P. The pan-genome as a shared genomic
1228 resource: mutual cheating, cooperation and the black queen hypothesis. *Front. Microbiol.* **6**,
1229 728 (2015).
1230 96. Kuosa, H. Picoplanktonic algae in the northern Baltic Sea: seasonal dynamics and
1231 flagellate grazing. *Mar. Ecol. Prog. Ser.* **73**, 269–276 (1991).
1232 97. Weitere, M., Dahlmann, J., Viergutz, C. & Arndt, H. Differential grazer-mediated effects of
1233 high summer temperatures on pico- and nanoplankton communities. *Limnol. Oceanogr.* **53**,
1234 477–486 (2008).

98. Novotny, A. *et al.* DNA metabarcoding highlights cyanobacteria as the main source of primary production in a pelagic food web model. *Sci Adv* **9**, eadg1096 (2023).

99. Plummer, M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria* (2003).

100. Watanabe, S. A widely applicable Bayesian information criterion. *J. Mach. Learn. Res.* **14**, 867–897 (2012).

101. Casey, J. R., Grinstein, S. & Orlowski, J. Sensors and regulators of intracellular pH. *Nat. Rev. Mol. Cell Biol.* **11**, 50–61 (2010).

102. Schneider, B. *et al.* Biogeochemical cycles. in *Biological Oceanography of the Baltic Sea* (eds. Snoeijs-Leijonmalm, P., Schubert, H. & Radziejewska, T.) 87–122 (Springer Netherlands, Dordrecht, 2017).

103. Kogure, K. Bioenergetics of marine bacteria. *Curr. Opin. Biotechnol.* **9**, 278–282 (1998).

104. Zhang, H. *et al.* Repeated evolutionary transitions of flavobacteria from marine to non-marine habitats. *Environ. Microbiol.* **21**, 648–666 (2019).

105. Benarroch, J. M. & Asally, M. The Microbiologist's Guide to Membrane Potential Dynamics. *Trends Microbiol.* **28**, 304–314 (2020).

106. More, K. J., Kaur, H., Simpson, A. G. B., Spiegel, F. W. & Dacks, J. B. Contractile vacuoles: a rapidly expanding (and occasionally diminishing?) understanding. *Eur. J. Protistol.* **94**, 126078 (2024).

107. Kawano, S., Ono, H., Takagi, T. & Bono, H. Tutorial videos of bioinformatics resources: online distribution trial in Japan named TogoTV. *Brief. Bioinform.* **13**, 258–268 (2012).

108. Rousk, J. *et al.* Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J.* **4**, 1340–1351 (2010).

109. Heisler, J. *et al.* Eutrophication and Harmful Algal Blooms: A Scientific Consensus. *Harmful Algae* **8**, 3–13 (2008).

1261 110. Agneta, A. *et al.* DNA extraction protocol for DNA-metabarcoding of marine phytoplankton
1262 using Zymobiomics DNA minprep kit (Zymo Research; D4300) v1. (2021)
1263 doi:10.17504/protocols.io.bucjnsun.
1264 111. Balzano, S., Abs, E. & Leterme, S. C. Protist diversity along a salinity gradient in a coastal
1265 lagoon. *Aquat. Microb. Ecol.* **74**, 263–277 (2015).
1266 112. Wu, L. *et al.* Phasing amplicon sequencing on Illumina Miseq for robust environmental
1267 microbial community analysis. *BMC Microbiol.* **15**, 125 (2015).
1268 113. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
1269 *EMBnet.journal* **17**, 10–12 (2011).
1270 114. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data.
1271 *Nat. Methods* **13**, 581–583 (2016).
1272 115. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny
1273 substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
1274 116. Kozlov, A. M., Zhang, J., Yilmaz, P., Glöckner, F. O. & Stamatakis, A. Phylogeny-aware
1275 identification and correction of taxonomically mislabeled sequences. *Nucleic Acids Res.* **44**,
1276 5022–5033 (2016).
1277 117. Lundin, D. & Andersson, A. SBDI Sativa curated 16S GTDB database. *SciLifeLab*
1278 *https://doi. org/10. 17044/scilifelab* **14869077**, v3 (2021).
1279 118. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data
1280 processing and web-based tools. *Nucleic Acids Res.* **41**, D590–6 (2013).
1281 119. Straub, D. *et al.* Interpretations of Environmental Microbial Community Studies Are Biased
1282 by the Selected 16S rRNA (Gene) Amplicon Sequencing Pipeline. *Front. Microbiol.* **11**,
1283 550420 (2020).
1284 120. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves
1285 sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).
1286 121. R Core Team. R: A Language and Environment for Statistical Computing. Preprint at

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

122. Inkscape Project. *Inkscape*. (2020).

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

124. 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).

125. 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).

126. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal* vol. Complex Systems 1695 Preprint at https://igraph.org (2006).

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

128. Lavrinienko, A., Jernfors, T., Koskimäki, J. J., Pirttilä, A. M. & Watts, P. C. Does Intraspecific Variation in rDNA Copy Number Affect Analysis of Microbial Communities? Trends Microbiol. 29, 19–27 (2021).