



Microbial Interactions – Underexplored Links Between Public Health Relevant Bacteria and Protozoa in Coastal Environments

Karolina I. A. Eriksson¹, Johanna Thelaus², Agneta Andersson^{1,3} and Jon Ahlinder^{2*}

¹ Department of Ecology and Environmental Sciences, Faculty of Science and Technology, Umeå University, Umeå, Sweden, ² Division of CBRN Defence and Security, Swedish Defence Research Agency (FOI), Umeå, Sweden, ³ Umeå Marine Sciences Centre, Umeå University, Hörnefors, Sweden

OPEN ACCESS

Edited by:

Diane McDougald,
University of Technology Sydney,
Australia

Reviewed by:

Neil Ross McEwan,
Robert Gordon University,
United Kingdom
Yolanda Moreno,
Universitat Politècnica de València,
Spain

*Correspondence:

Jon Ahlinder
jon.hallander@gmail.com

Specialty section:

This article was submitted to
Aquatic Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 16 February 2022

Accepted: 23 May 2022

Published: 13 June 2022

Citation:

Eriksson KIA, Thelaus J,
Andersson A and Ahlinder J (2022)
Microbial Interactions –
Underexplored Links Between Public
Health Relevant Bacteria
and Protozoa in Coastal
Environments.
Front. Microbiol. 13:877483.
doi: 10.3389/fmicb.2022.877483

The co-existence of bacteria and protozoa in aquatic environments has led to the evolution of predation defense mechanisms by the bacteria. Some of the predation-resistant bacteria (PRB) are also pathogenic to humans and other mammals. The links between PRB and protozoa in natural aquatic systems are poorly known, but they are important in predicting outbreaks and determining the long-term consequences of a contamination event. To elucidate co-occurrence patterns between PRB (16S rRNA) and bacterivorous protozoa (18S rRNA), we performed a field study in a coastal area in the northern Baltic Sea. Interactions between bacteria and protozoa were explored by using two complementary statistical tools. We found co-occurrence patterns between specific PRB and protozoa, such as *Legionella* and Ciliophora, and we also found that the interactions are genotype-specific as, for example, *Rickettsia*. The PRB sequence diversity was larger in bays and freshwater inlets compared to offshore sites, indicating local adaptations. Considering the PRB diversity in the freshwater in combination with the large spring floods in the area, freshwater influxes should be considered a potential source of PRB in the coastal northern Baltic Sea. These findings are relevant for the knowledge of survival and dispersal of potential pathogens in the environment.

Keywords: bacteria, protozoa, predation resistance, biotic interactions, aquatic microbiology, co-evolution, joint species distribution model, direct acyclic graph (DAG)

INTRODUCTION

Opportunistic bacterial pathogens often hide in the environment before infecting humans (Matz and Kjelleberg, 2005; Winstanley et al., 2016; Amaro and Martín-González, 2021). They survive in the environment, as they master ecological niches where grazing from protozoa is evaded. Bacteria and protozoa have co-existed in the aquatic ecosystems for a long period, even before the evolution of mammals. As a response to the long-term predation pressure by protozoa, the bacteria have developed defense mechanisms, for example, by changing their size, forming biofilms, excreting toxic compounds, or preventing degradation once engulfed by the protozoa

(Matz and Kjelleberg, 2005). Simply described, there are two major forms of predation-resistant bacteria (PRB): those having an extracellular lifestyle (EPRB) and those having an intracellular lifestyle (IPRB) (Weitere et al., 2005; Thelaus et al., 2009).

Some traits that render bacteria resistant to predation also provide tools for infecting mammals, and these are classified as virulence factors (Amaro and Martín-González, 2021). Protozoa, such as ciliates, flagellates, and amoebas, drive the evolution of bacteria as a selective force. Bacterial interactions with protozoa thus play an important role in the development and persistence of virulence traits (Sun et al., 2018). In nature, examples of both potentially pathogenic EPRB and IPRB can be found (Von Reyn et al., 1993; Miyasaka et al., 2006; Kao et al., 2013; Pappa et al., 2017). Despite the importance of associations within the microbial food web, little has been documented regarding PRB–protozoa interactions in the natural aquatic environments (Thelaus et al., 2008). The challenges lie in understanding interactions in complex and dynamic ecosystems, and further disentangling abiotic and biotic factors in large datasets.

Biotic interactions have been recognized as important for understanding the influence and functions of a community. Unfortunately, estimating such interactions in highly dimensional metabarcoding data is difficult, as the number of interactions grows exponentially with the number of taxa. Several efforts have been made to overcome this challenge, mostly based on defining the distance between assemblages, such as ordination or network methods (Faust and Raes, 2012; Weiss et al., 2016). However, these methods often include subjective choices and difficulties in interpreting the results. An appealing alternative is to *a priori* visualize a network for PRB based on direct acyclic graphs (DAG), where dependencies between taxa can be ordered in a topological sequence, which can be interpreted as taxa influencing other community members (e.g., A→B, taxa A leads to taxa B). In addition, recent advances within hierarchical modeling tools have paved the way toward using a joint model for abundance on taxa (underpinned by the generalized linear latent variable model generalized linear latent variable model (GLLVM); Warton et al., 2015; Niku et al., 2017; Ovaskainen et al., 2017). This multivariate tool can incorporate several aspects in a single analysis, such as assessing the impact of environmental factors on species abundance and finding a correlation between taxa. Still, a few studies have applied these tools for correlations in microbial ecology (e.g., Metcalf et al., 2016; Niku et al., 2017; Andersson et al., 2018).

To improve the understanding of the ecological niche and specific interactions between PRB and protozoa in aquatic systems, we performed a field study on the northern Baltic Sea coast. PRB were defined as bacteria that have been shown to resist amoeba degradation with varying ecological strategies (amoeba-associated and free-living). The target bacteria are based on the PRB identified by isolation from amoebas and bacteria shown to resist phagocytosis (Bertelli and Greub, 2012), listed in **Table 1**. We sampled four bays representing offshore water and river inflow during the productive season in the period from May to September. We subsequently combined the use of two statistical methods (DAG and GLLVM) in order to identify possible interactions between the phagotrophic protozoa and

TABLE 1 | Predation-resistant bacteria (PRB) identified by isolation from amoebas and bacteria shown to resist phagocytosis (Bertelli and Greub, 2012).

Phyla	Genus	PRB
<i>α</i> proteobacteria	<i>Afiplia</i>	<i>Afiplia felis</i> , <i>Afiplia broomae</i>
<i>α</i> proteobacteria	<i>Bosea</i>	<i>Bosea</i> spp.
<i>α</i> proteobacteria	<i>Rickettsia</i>	<i>Rickettsia bellii</i>
<i>β</i> proteobacteria	<i>Burkholderia</i>	<i>Burkholderia cepacia</i> , <i>Burkholderia pseudomallei</i>
<i>γ</i> proteobacteria	<i>Coxiella</i>	<i>Coxiella burnetii</i>
<i>γ</i> proteobacteria	<i>Francisella</i>	<i>Francisella tularensis</i>
<i>γ</i> proteobacteria	<i>Pseudomonas</i>	<i>Pseudomonas aeruginosa</i>
<i>γ</i> -proteobacteria	<i>Legionella</i>	<i>Legionella drancourtii</i> , <i>Legionella longbeachae</i> , <i>Legionella pneumophila</i>
<i>Chlamydiae</i>	<i>Parachlamydia</i>	<i>Parachlamydia acanthamoebae</i> , <i>Protochlamydia amoebophila</i>
<i>Chlamydiae</i>	<i>Simkania</i>	<i>Simkania negevensis</i>
<i>Actinobacteria</i>	<i>Mycobacterium</i>	<i>Mycobacterium tuberculosis</i> , <i>Mycobacterium bovis</i> , <i>Mycobacterium avium</i> , <i>Mycobacterium smegmatis</i>
<i>Chlamydiae</i>	<i>Waddlia</i>	<i>Waddlia chondrophila</i>
<i>Bacteroidetes</i>	<i>Candidatus</i> <i>Amoebophilus</i>	<i>Amoebophilus asiaticus</i>

PRB. In addition, in order to identify the source for PRB, we used the source tracking method Random Forest (RF).

We hypothesized that in spring the freshwater inflow to the bays would seed and promote opportunistic EPRB, such as *Pseudomonas*. We expected IPRB, such as *Legionella*, to show strong links and dependency on certain protozoan hosts due to their intracellular lifestyle. These would increase in abundance later in the season, as protozoa often show maxima in summer.

RESULTS

Environmental Conditions

The river inflow to the coastal area peaked in May due to the spring snowmelt in the catchment area (**Supplementary Figure 1C**). It was 35-fold higher in spring than in summer. This decreased the salinity in the bays, which had an influence on the bacterial composition in May (**Supplementary Figure 1**). The mean seawater temperature in the study bays [Valviken (VA), Kalvarskatan (KA), Stadsviken (ST), and Ångerån (AN)] and the offshore water (HÖRNE and ÖRE) was 10–12°C in spring, peaked in July (23–24°C), and decreased toward autumn (14°C) (**Supplementary Figure 2**). The salinity was lowest during the spring flood, ~2.5 psu, and increased to 3.5–4 psu during the summer. The dissolved organic carbon (DOC) showed maximum values in spring and decreased as the season progressed (**Supplementary Figure 2**).

The chlorophyll *a* concentration and phytoplankton primary production peaked in spring and decreased in summer (**Supplementary Figure 2**). A similar trend was also observed for the protozoan biomass; however, occasional blooms of dinoflagellates were observed in summer (**Supplementary Figure 3**). The bacterial biomass was more stable over the study

period (**Supplementary Figure 4**), showing slightly higher values in summer than in spring and autumn. In general, the bacterial production was higher at the innermost sampling site (position 1) than at the seaward sites in the bays (positions 2 and 3) (**Supplementary Figure 2**). The protozoan and bacterial biomass also tended to have higher values at the innermost position in the bays, decreasing in the seaward direction (**Supplementary Figures 3, 4**). Bacterial biomass showed lower values at the offshore sampling sites.

Principal component analysis (PCA) of environmental variables resulted in monthly clusters (**Figure 1C**), indicating that the temporal variation was larger than the spatial variation in the bays. The influence of DOC and humic substances were marked in May, while salinity influenced the distribution pattern in summer.

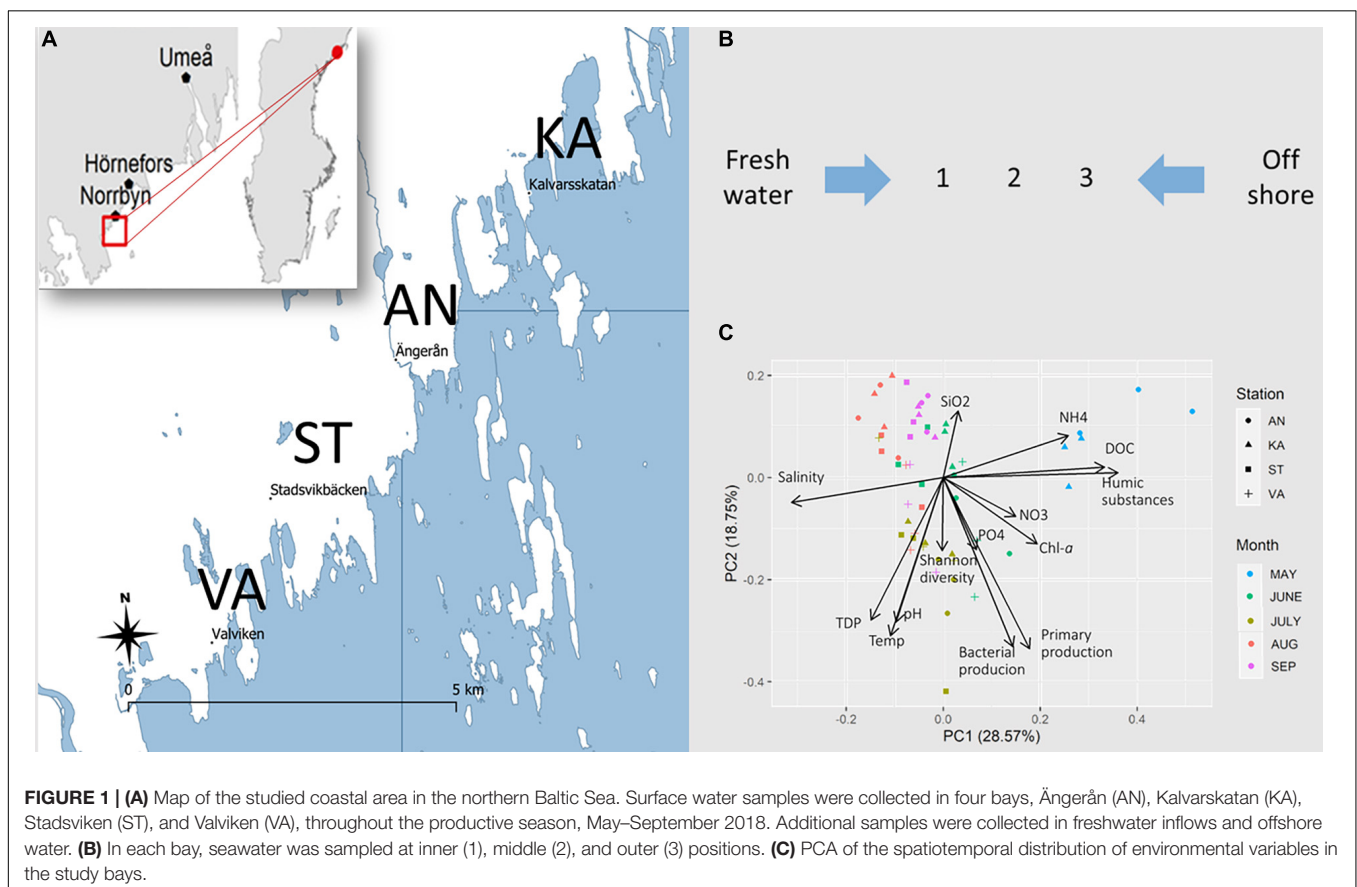
Occurrence of Predation Resistant Bacteria (PRB)

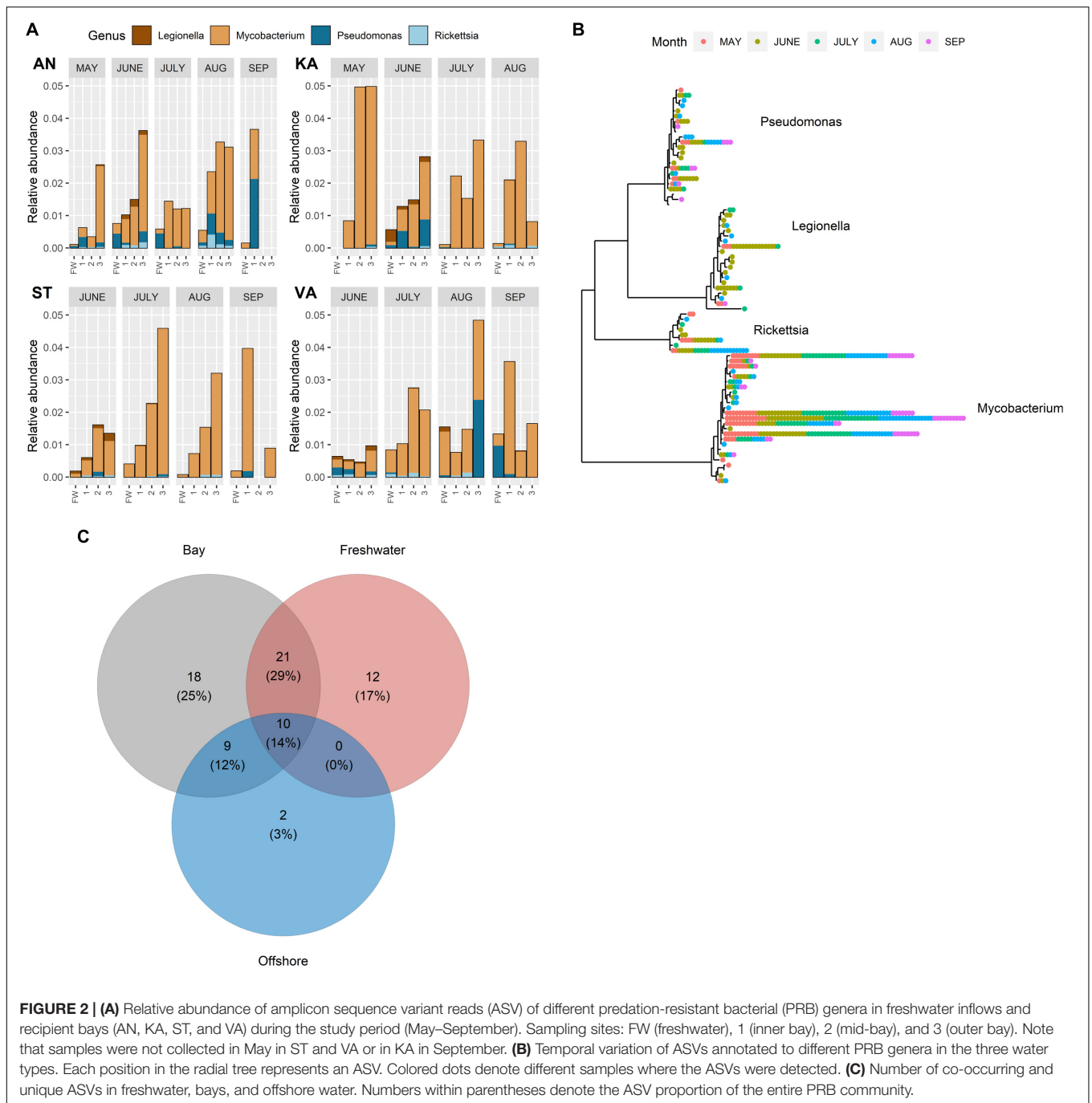
Predation-resistant bacteria were detected in all the collected samples, that is, in freshwater, bays, and offshore water throughout the study period (**Figures 2A, 3**). A total of 76 amplicon sequence variants (ASVs) were identified, which were distributed to four PRB genera. Several families that contain PRB were also detected, however, without further taxonomic information (**Supplementary Figure 5**). The majority of the PRB genera corresponded to *Mycobacterium* (25

ASVs), followed by *Pseudomonas* (23 ASVs), *Legionella* (20 ASVs), and *Rickettsia* (8 ASVs) (**Figure 2B**). The relative abundance of the different genera approximately followed the PRB ASV distribution, where *Mycobacterium* was the most abundant PRB and *Rickettsia* was the least abundant (**Figures 2A, 3A**).

Some seasonal trends of the PRB abundances were observed. *Legionella* was detected mainly in June, when it showed widespread occurrence in freshwater, the bays, and offshore water (**Figures 2, 3**). *Rickettsia* occurred scarcely and seem close to the detection limit in September (**Figures 2, 3**). Within a month, the PRB composition showed a somewhat similar distribution pattern from the inner to the outer parts of the bays, even though variations were also found (**Figure 2A**). Seasonal trends of different ASVs within the same genus were also found (**Figure 2B**). For example, some ASVs of *Pseudomonas*, *Mycobacteria*, and *Rickettsia* occurred only in May (**Figure 2B**).

The number of ASVs, that is, the sequence diversity, was about twice as high in the bays and freshwater than observed in the offshore water (**Figure 2C**). Approximately, 30% of the ASVs were shared among freshwater and bays, while only 10% were shared among all the three habitats. A total of 25, 17, and 3% of the ASVs were unique for the bays, freshwater, and offshore water, respectively. The number of *Legionella* ASVs was higher in freshwater and in the bays compared to

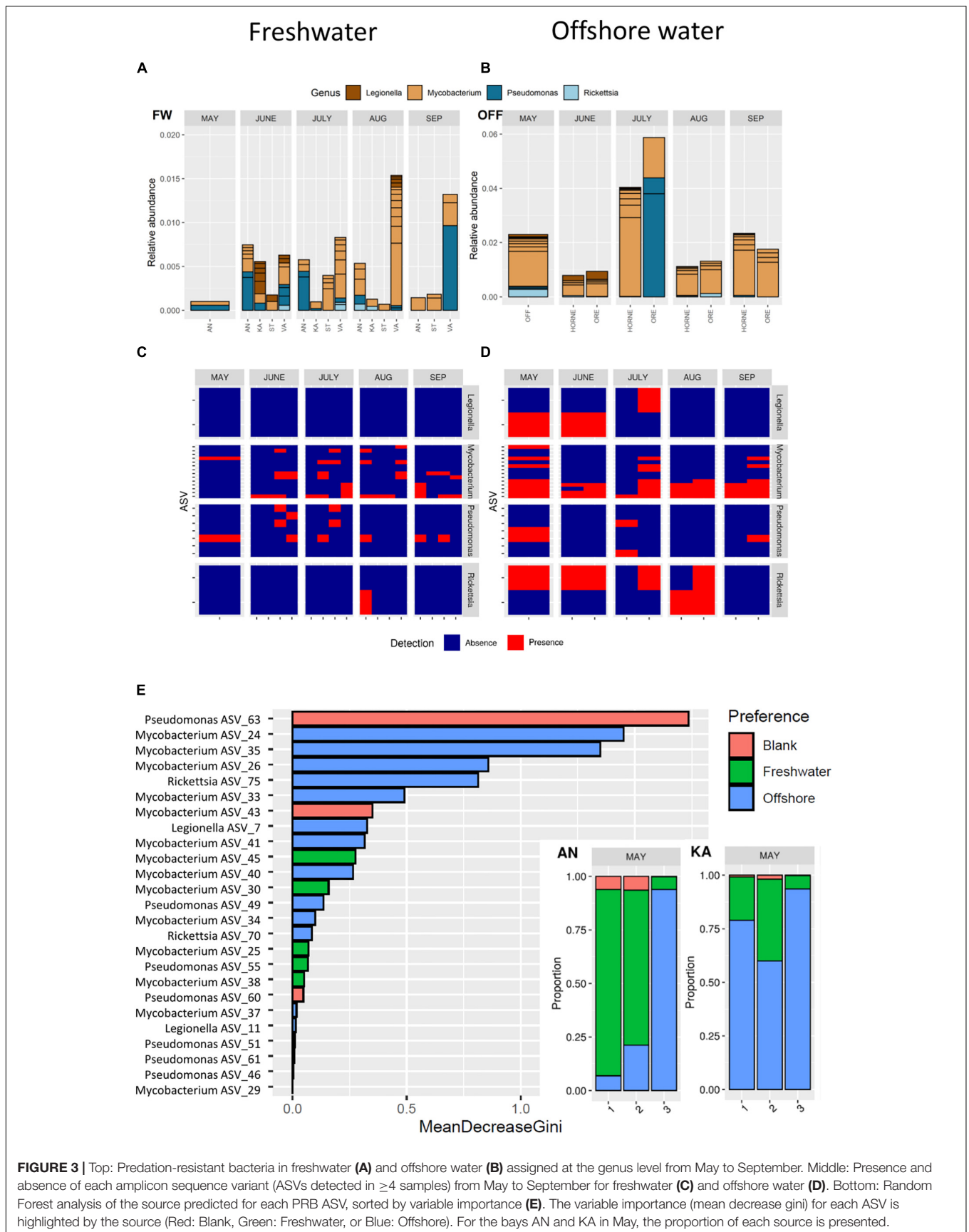




offshore water (**Supplementary Table 1**). For the other PRB genera, no difference in the number of ASVs could be detected between the habitats.

Analysis of the absence/presence of widespread ASVs detected in ≥ 4 samples showed that *Mycobacterium* and *Pseudomonas* occurred in both freshwater and offshore water all through the study period (**Figures 3C,D**). In contrast, ASVs of *Legionella* and *Rickettsia* mainly occurred in offshore water during the spring–summer period and not in autumn (September) (**Figures 3C,D**).

Random Forest analysis indicated that specific ASVs belonging to *Mycobacterium*, *Rickettsia*, *Legionella*, and *Pseudomonas* originate from offshore water, while specific ASVs belonging to *Mycobacterium*, *Rickettsia*, and *Pseudomonas* originate from freshwater (**Figure 3E** and **Supplementary Figure 8**). In spring, as much as 90% of the PRB ASVs at the innermost station of AN were predicted to originate from freshwater, while the freshwater influence decreased markedly in a seaward direction. In the bay receiving less river inflow, the analysis indicated that ASVs at all sampling sites in KA to



a higher degree originated from the offshore water. Thus, the potential freshwater origin of many PRB was notable, such as for certain ASVs of *Mycobacterium* and *Pseudomonas* (Figure 3E). Thus, the spring river inflow was identified as significant source habitat for many of the PRB ASVs.

Phagotrophic Protozoa Resolved at the Species Level

We designed a novel set of primers specific for 18S rRNA amplicon sequencing, based on the V6F and V8R regions. Based on our *in silico* primer predictions, this design has the benefit of both increasing the taxonomic resolution and covering much of the diversity. The benefits of this region are shown in detail by Latz et al. (2022). Unlike the 18S rRNA V9 region, this longer 18S rRNA V6-V8 region was able to assign the majority of the taxa to species level (>50%). For the relative abundance of protozoans (Supplementary Figure 7, 30 most abundant taxa), Ciliophora (ciliates) was most abundant over the season. The relative abundance for dinoflagellates was the highest in May in all bays. Telonemia mostly occurred in August and September, while Ochrophyta was detected all through the season. Similar to PRB, we found a taxonomic difference over the season, as well as between the bays, for example, Apicomplexa were only detected in Valviken. However, the variability in composition was great between the bays and freshwater in most cases, with species of the family Strobilidiidae only present in freshwater. A highly similar composition of protozoa was observed for all sampling positions (1, 2, and 3) in each bay, indicating little within-bay variation of protozoa and good reproducibility.

DAG Analysis Identifies PRB That Are Dependent on Protozoans

To infer interactions between bacteria and protozoa in this coastal area, we merged three datasets (based on environmental data found in Figure 1C and relative abundance data found in Figure 2A and Supplementary Figure 7) to construct a DAG network. Ecotypes within genus were identified by clustering the ASVs into groups (Group 1 and Group 2), based on the relative abundance in the samples. According to the model (Figure 4), the most important connection in the PRB network (according to influence score statistics) is the dependency of bacterial production on dissolved organic carbon (Supplementary Table 2). Overall, in the inferred network, the PRB are directly dependent on the protozoans. PRB ASVs assigned to *Rickettsia*, *Legionella*, and *Pseudomonas* are dependent on the relative abundance of Ciliophora ASVs, which suggests that these organisms interact with each other.

GLLVMs Analysis Identifies Interactions Between PRB and Protozoa

To better understand co-occurrence patterns, we need to disentangle the environmental factors and interactions identified by DAG, while correcting for spatiotemporal effects (Figures 1C, 2A). The independent contribution of each independent variable in the model is found in Figure 5, Left, as much as 69.2% and the full model explains 85.9% of the total co-variation in the PRB and protozoan groups. For

the full model, combination effects of the month, position (1, 2, and 3), and bay (i.e., station) were included to account for specific conditions. Other environmental variables used in the model were pH, NH₄, PO₄, DOC, bacterial production, and salinity. After the inclusion of the predictors in the full model, the seasonal variations were captured, allowing predictions of biotic interactions (Supplementary Figure 10B). Significant parameters for each PRB group are found in the coefficient plot in Supplementary Figure 5 and Supplementary Table 5. For both *Pseudomonas* and *Rickettsia*, strong differences between the within-genus PRB groups were inferred, with opposite effects of most predictors (in 22 and 33 out of total 41 cases, respectively), suggesting niche differentiation within these genera. This effect was not as pronounced for *Mycobacterium*, suggesting that this group is more homogeneous. Furthermore, many combinations of predictor months, bay, and position were significantly pointing toward strong local adaptation, at least for *Pseudomonas* and *Rickettsia*. *Legionella* was significantly overrepresented in June and in bay VA.

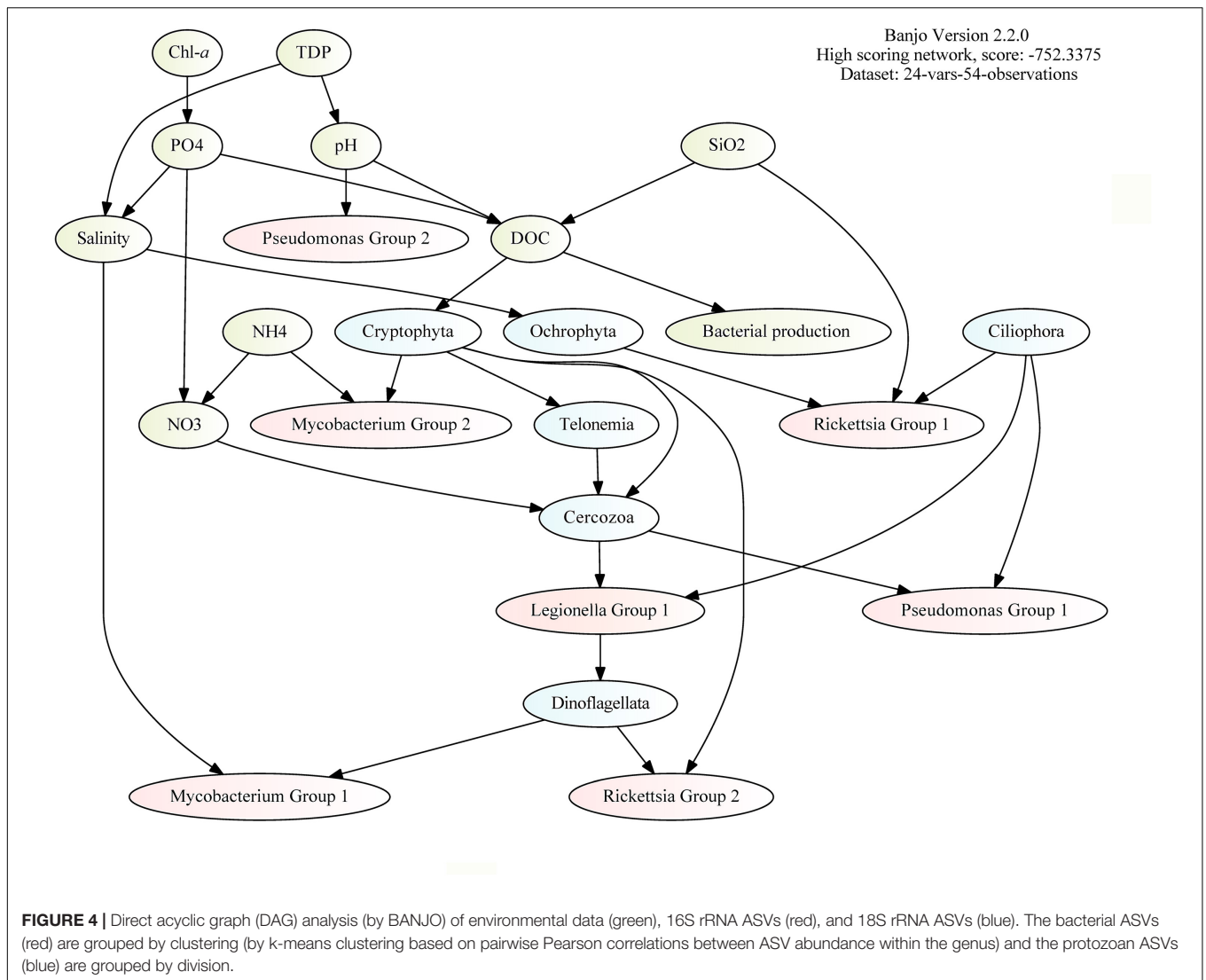
The remaining 14.1% of the co-variation in the model not explained after the inclusion of predictors was likely to be explained by biological interactions. After accounting for environmental and spatiotemporal effects, the GLLVMs identify interactions between PRB and protozoa, such as between *Legionella* and Ciliophora, and between *Rickettsia* and Ciliophora (Figure 5, Right). *Mycobacterium* and Dinoflagellata show a strong positive correlation, while *Legionella* and Dinoflagellata show a strong negative correlation. Members of *Pseudomonas* Group 1 show a strong correlation with all phagotrophic protozoa, either positively or negatively. PRB–protozoa interactions are therefore an important source of variation.

DISCUSSION

In this study, ecological niches and specific interactions between PRB and protozoa were identified in a coastal area in the northern Baltic Sea. Four public health-relevant PRB genera, *Mycobacterium*, *Legionella*, *Pseudomonas*, and *Rickettsia*, were found to occur throughout the productive season in the study area. Information about protozoan diversity was obtained by designing and applying a new set of V6–V8 18S rRNA-specific primers for amplicon sequencing. Specific ASVs of the PRB co-occurred with certain phagotrophic protozoa in the bays, freshwater discharge, and offshore water. The PRB occurrence and PRB–protozoa associations depended strongly on the time and place of sampling. By applying GLLVMs and direct acyclic graphs (DAGs), we were able to examine biological relationships beyond temporal and local effects. One important result from the analysis was the identification of a link between *Legionella* and ciliates (Ciliophora) in June, possibly due to an endosymbiotic relationship as *Legionella* has an intracellular lifestyle (Gomez-Valero and Buchrieser, 2019).

PRB Occurrence Is Influenced by Strong Spatiotemporal Effects

In general, a strong spatiotemporal variation of PRB abundance was inferred by our model, where the month of sampling and



interacting variables between month, station, and position was most important in explaining the PRB occurrence (**Figure 5** left and **Supplementary Figure 5**). Interestingly, the location was of importance even though the geographic scale was relatively small (~8.5 km between bays KA and VA), indicating local adaptations and habitat fragmentation. Further, we found unique ASVs at different sampling sites (**Figure 2B**), which also support PRB local adaptation.

Amplicon sequence variants specific to *Legionella*, in particular, displayed a strong seasonality as they occurred more or less exclusively during June in all bays. Previous studies have suggested that temperatures above 20°C promote the relative abundance of *Legionellales* in freshwater (Graells et al., 2018). Herein, we observed a peak of *Legionella* in June when the water temperature was around 12°C, which did not differ significantly from May (**Supplementary Figure 2**). Generally, for the Baltic Sea, the south–north differences in temperature and the salinity gradient make the northern parts more similar to freshwater. During an inventory of bacterial 16S rRNA amplicon sequences

along the salinity gradient of the Baltic Sea, Herlemann et al. (2011), identified marked differences in the bacterial community compositions of brackish water compared to freshwater and marine counterparts (Herlemann et al., 2011). Taken together, the northern parts hold an easier habitat to exploit by species adapted to freshwater, such as *Legionella*. The low salinity in the study area may explain why *Vibrio* was not detected in this study (Baker-Austin et al., 2012), even though this genus has been reported in the Baltic Proper (Gyraite et al., 2019). Considering climate change, and the predicted effect on both temperature and salinity, the Baltic Sea may face different challenges in the future (Andersson et al., 2015).

Freshwater is considered the major reservoir for the pathogenic *Legionella pneumophila* (Boamah et al., 2017). However, *Legionella* is often studied in man-made water systems, such as cooling towers (Brigmon et al., 2020), due to the thermophily of the pathogen *L. pneumophila* (Lesnik et al., 2015). Even though it is known that *Legionella* exist in freshwater and have cold-adapted variants (Lesnik et al., 2015), few studies



have reported their occurrence in natural aquatic environments, such as in lakes, rivers, and estuaries (Carvalho et al., 2007; Parthuisot et al., 2010; Shimada et al., 2020). However, a large meta-analysis of the obligate intracellular family Legionellales (also including *Rickettsia*) shows that these bacteria exist in half of the freshwater and marine water samples globally (Graells et al., 2018). Importantly, that study found that both cold and marine waters contain many uncultured variants, as confirmed by our study. To summarize, we could not identify temperature as a driver for the occurrence of *Legionella* in this area. Rather, the occurrence was inferred by biotic interactions.

DAG and GLLVM Identify PRB Preferences to Protozoa

In order to infer biotic interactions based on environmental parameters and sequence data, it is necessary to disentangle factors causing spatial and temporal variations. By using two complementary inference methods, interactions were predicted between specific PRB and protozoa, such as between *Mycobacterium* and Dinoflagellata, between *Legionella* and Ciliophora, and between *Rickettsia* and both Ciliophora and Cercozoa (Figures 4, 5). Since the statistical models are a simplification of a complex dynamic ecosystem, it is worth noting that missing significant predictors can have an effect on the results. Therefore, it is possible that not all correlations represent true biotic interactions. This study focused on bacteria that have previously shown resistance to degradation by a protozoan, which does not guarantee resistance to other protozoans. The mentioned interactions are,

however, supported by previous works, which have reported endosymbiosis of *L. pneumophila* in the ciliate *Paramecium caudatum* (Watanabe and Watarai, 2017; Nishida et al., 2018; Watanabe et al., 2018) and association of a *Rickettsia*-like bacterium with another common ciliate (Vannini et al., 2005; Ferrantini et al., 2009). In addition, Szokoli et al. (2016), reported the association of two *Rickettsiales* symbionts with the ciliate protist *Paramecium biaurelia* via the cytoplasm. *Rickettsiales* members have also been associated with Cercozoa species, where Hess et al. (2016) discovered endosymbionts in an ameboflagellate. Also, several species of *Mycobacterium* have previously been shown to replicate inside the dinoflagellate *Dictyostelium discoideum* (Solomon et al., 2003; Steinert and Heuner, 2005).

The mentioned studies isolated the bacteria and protozoa to identify their interactions. This requires that the organisms can be cultivated, which is often a major challenge considering intracellular bacteria (Singh et al., 2013). As an alternative approach, Lamrabet et al. (2012) used genomic information to search for protein homologs within *Mycobacterium* and three protozoans (*D. discoideum* included), which support the exchange of genetic material. For inferring novel interactions, we show that our approach identified potentially similar interactions at the community level in the natural environments of both PRBs and protozoa. Thus, our approach could be a viable option to detect novel interacting microbes in their natural habitats using culture-independent methods. Isolation and single-cell genomics could then be used to further investigate potential interactions (Boscaro et al., 2022).

Managing Rare Sequence Variants and Avoiding Bias

The PRB identified occurred at relatively low abundances in the sequence data, constituting a maximum of 3–5% of the entire bacterial community. A common pattern for *Legionella*, *Mycobacteria*, and *Rickettsia* ASVs was the detection of rare sequence variants that only occurred on a few occasions or few sampling sites (≤ 4 samples). In line with previous studies (Niku et al., 2017; Roguet et al., 2018), the rare sequence variants were not included in the models (DAG, GLLVM, and Random Forest), since they increase the noise and risk for false positives. Thus, 54 rare sequence variants were excluded (out of 76 in total), of which many occurred in freshwater (Supplementary Table 4). The resulting model that includes the widespread ASVs shows a generally increased detection frequency of these ASVs in offshore waters (Figures 3A–D). This was supported in the Random Forest analysis that predicted the source for each ASV (Figure 3E), where offshore water was the dominating habitat, except during the spring flood in May for AN that was receiving the largest amount of river inflow (Figure 3E). To summarize, even after the exclusion of rare sequence variants, Random Forest was able to identify a gradient without the knowledge of the study design.

PRB Diverse in Freshwater – Abundant in Offshore Water

In general, the sequential diversity of PRB was higher in freshwater than in the marine system, where each ASV was detected in fewer freshwater samples compared to the ASVs found in the offshore samples. However, the relative abundance of the widespread ASVs was higher in the offshore water. *Pseudomonas* and *Mycobacterium* were abundant during the spring river flush in the AN bay which has the highest freshwater inflow. The number of *Pseudomonas* ASVs detected in freshwater was larger when compared to offshore water, but the difference was not statistically significant ($n = 9$ for freshwater and $n = 4$ for offshore). The GLLVM analysis indicated that *Pseudomonas* and *Mycobacterium* were adapted to several local environments (Supplementary Figure 9 and Supplementary Table 5), and based on the source tracking analysis, different ASVs seemed to originate from different sources (Figure 3E). The model predicted that certain *Pseudomonas* and *Mycobacterium* ASVs originated from freshwater, especially in AN bay during spring (Figure 3E). Importantly, it seems possible for *Pseudomonas* to originate from the spring flood and survive during the season, as could be the case for one of the ASVs seen in Figure 3C.

Pseudomonas is a ubiquitous genus occurring worldwide in many environments, for example, in soils, freshwater, estuaries, and marine systems (Hutalle-Schmelzer and Grossart, 2009; Lindh et al., 2015; Pent et al., 2017; Qiao et al., 2018). *Pseudomonas* is opportunistic and able to benefit from disturbed or polluted conditions (Andersson et al., 2018; Rizzo et al., 2019). These bacteria have previously been described in the Baltic Sea, and some can cause infections in fish and other animals (Lönnstrom et al., 1994; Wiklund, 2016; Sonne et al., 2020). Alarming detection of *Pseudomonas* has been previously described after disturbing events, such as flooding (Shankar et al.,

2021), and in wastewater effluents (Luczkiewicz et al., 2015). Our results suggest that the spring flood was a disturbance for the bacterial community (Supplementary Figure 1) and indicate that *Pseudomonas* profited by quickly exploiting resources during the spring colonization of the bays.

PRB: A Diverse Group With Varying Co-occurrences

The model clustering of the ASVs indicated that the bacteria constituted different PRB ecotypes. We identified different protozoan host preferences within the PRB genus, wherein two particular ASVs assigned to *Rickettsia* demonstrated varying co-occurrences (Figure 5, Right). These conflicting patterns apply to both the protozoan interactions and association with location and environmental variables, indicating different niche preferences. In addition, the two groups of *Pseudomonas* showed many strong negative correlations with protozoans (Figure 5, Right). These findings are consistent with previous studies where highly resolved sequence clusters showed opposite environmental preferences across a gradient of disturbance (Newton and McLellan, 2015) and over a gradient of salinity (Herlemann et al., 2011). Both *Pseudomonas* and *Rickettsia* are heterogeneous groups of bacteria including not only human, animal, and plant intracellular pathogens, but also non-pathogenic free-living relatives. In addition, variation in traits, such as biofilm formation, is expected in these genera. Several strains of *Pseudomonas* are known to produce toxins (Matz et al., 2004), some of which can kill protozoans, such as amoebas (Matz et al., 2008). Therefore, the many strong negative correlations between *Pseudomonas* Group 1 and, for example, Ciliophora and Cercozoa could be interpreted as excretion of toxins harmful to potential grazers. Thus, when studying interactions between organisms in the genus containing potential human pathogens, it is important to account for genetic variability within the genus and host preference.

CONCLUSION

We conclude that PRB constitute a group of highly diverse bacteria with many unique sequence variants that are hard to predict. Our data suggest that freshwater, coastal water, and offshore water pose a risk of harboring potentially pathogenic bacteria. We observe strong spatial and temporal variation in the occurrence of PRB. Amplicon sequence variants of *Pseudomonas*, *Rickettsia*, *Legionella*, and *Mycobacterium* were more diverse in freshwater, while more abundant in seawater, and occurred at specific periods throughout the productive season. These dependencies prevent investigators from inferring specific interactions that depend on biotic interactions, such as with protozoans. By applying two statistical models, we were able to disentangle interactions in a complex and dynamic ecosystem. We showed that the PRB have varying co-occurrences and host preferences. For example, our model suggests that the maximum population of *Legionella* recorded in June was dependent on the occurrence of Ciliophora. Thus, this work provides a basis for identifying sources and drivers for potential human pathogens that survive in the environment.

MATERIALS AND METHODS

Study Sites

From May to September 2018, monthly water samples were collected at 0.5 m depth at three sampling points in four bays along the northern Baltic coast: Ångerån (AN), Kalvarsskatan (KA), Stadsviken (ST), and Valviken (VA) (**Figure 1A**). Additionally, samples were taken from the freshwater inlets (FW) and offshore waters: Örefjärden (ORE), Hörnefors (HORNE), and Degerfjärden (OFF). The first sampling coincided with the spring snowmelt, and the bays were thus exposed to relatively high freshwater inflow in May than during later samplings. Due to the unstable ice situation, we were not able to sample Stadsviken and Valviken bays in May. Ångerån (AN) received the largest river discharge among the sites and was the brownest in water color (**Figure 1C**), while Kalvarsskatan (KA) had no river inflow and was relatively a clear water bay. Stadsviken (ST) and Valviken (VA) received intermediate freshwater inputs compared to AN and KA. In total, 80 water samples were collected. Therefore, our sites covered a gradient in freshwater supply and associated inputs of terrestrial dissolved organic matter and nutrients. The magnitude of this gradient varies temporally. Procedures for measurement of physicochemical variables, primary production, bacterial production, and microscopic analysis of phytoplankton are presented in **Supplementary Materials and Methods**.

Bacterial DNA Sampling and Extraction

At each sampling occasion, 200–500 ml of water was gently filtered (≤ 20 kPa) onto 0.2- μm filters (Pall Corporation sterilized filters, Supor[®] 0.2 μm , 47 mm, S-pack white gridded). For freshwater samples, < 500 ml was filtered due to limitations associated with the filtering of turbid water. A total of 80 water samples were collected. The filters were then folded using cleaned tweezers and placed in 2-ml Eppendorf tubes. The samples were stored at -80°C until the DNA extraction. Filters were thawed and placed into PowerWater bead beating tubes. The DNA was extracted using a DNeasy[®] PowerWater kit[®] (Qiagen, Hilden, Germany) according to a modified DNeasy PowerWater protocol. The samples were treated with an additional heating step (horizontal water bath for 30 min, 65°C) and with a bead beating step in each direction of 20 Hz for 3×3 min with a Tissuelyser II (Qiagen, Hilden, Germany). Two filter blanks were used as a negative control during DNA extraction. When the DNA extracts were obtained from the DNeasy[®] PowerWater[®] kit, the samples were prepared for PCR analysis by Illustra columns (MicroSpin S-200 HR, GE Healthcare). The final DNA was frozen before subjecting to PCR analysis. Sample preparation, thermal cycling, and PCR product preparation were performed in separate rooms.

18S rRNA Primer Design

In order to design a suitable candidate PCR primer pair to target conserved regions in the 18S rRNA (Hugert et al., 2014),

which covers a large taxonomic coverage, we used the program Primer Prospector (Walters et al., 2011). The primer pairs V6F: 5'-AATTYGAHTCAACRCGGG-3' (Hadziavdic et al., 2014) and V8R: 5'-GACRGGCGGTGTGNACAA-3' (Edgcomb et al., 2011) were evaluated, using higher degeneracies than earlier publications. To predict the taxonomic coverage, PR2 version 4.12.0 (Guillou et al., 2012) and SILVA (Pruesse et al., 2007) databases were used. All major taxonomic divisions were covered (**Supplementary Figure 11**). The forward and reverse primers were modified to incorporate a 12-bp Golay error-correcting barcode that enables sample multiplexing (Caporaso et al., 2012). The primers were optimized and tested for secondary structures (Gibbs free energy: mean, -11 kcal/mol).

Amplicon Preparation and Sequencing

For the 16S rRNA amplicon preparation, DNA was amplified and sequenced as previously described (Hägglund et al., 2018), apart from using the PCR purification kit. In short, the DNA was amplified using the No. 5 Hot Mastermix 2.5x kit (5 PRIME, Qiagen, Hilden, Germany) with bacteria/archaeal primers 515F/806R specific for the hypervariable V4 region of the 16S rRNA gene. The forward and reverse primers were modified to incorporate a 12-bp Golay error-correcting barcode that enables sample multiplexing (Caporaso et al., 2012). All samples were amplified in triplets and pooled after PCR amplification (94°C for 3 min, followed by 35 cycles at 94°C for 45 s, 50°C for 1 min, 72°C for 1.5 min, and finally 10 min rest to finish).

For the 18S rRNA amplicon preparation, DNA was amplified using the No. 5 Hot Mastermix 2.5x kit (5 PRIME, Qiagen, Hilden, Germany) with eukaryotic primers V6F/V8R (V6F: 5'-AATTYGAHTCAACRCGGG-3' and V8R: 5'-GACRGGCGGTGTGNACAA-3') specific for the hypervariable V6–V8 region of the 18S rRNA gene. All samples were amplified in triplets and pooled after PCR amplification (94°C for 3 min, followed by 35 cycles at 94°C for 45 s, 57°C for 1 min, 72°C for 1.5 min, and finally, 10 min rest to finish).

For PCR, three template blank reactions were used by using water instead of template. The PCR product was run on a 1% agarose gel, and the DNA concentration was estimated with a Qubit fluorometer (Invitrogen, Carlsbad, CA, United States). The amplicons were pooled at equimolar concentrations and purified with the QIAquick PCR purification kit (Qiagen, Hilden, Germany) following the supplier's instructions. The DNA concentration of the pooled amplicon product was measured with a Qubit fluorometer and adjusted to 2 nM. The library was denatured and diluted as described by Illumina (MiSeq System User Guide, Part # 15027617 Rev. C), before it was loaded onto a MiSeq cartridge (Illumina, San Diego, CA, United States) and sequenced using a 2×300 bp paired-end sequencing protocol (Hägglund et al., 2018).

Quality Control and Raw Data Processing

Demultiplexing was done using deML (Renaud et al., 2015). The Quantitative Insights Into Microbial Ecology (QIIME2)

pipeline, version 2019.1, was used for processing raw sequence data (Bolyen et al., 2019). Greengenes version 13.8 (McDonald et al., 2012) and PR2 version 4.12.0 (Guillou et al., 2012) databases were used for bacteria and eukaryote taxonomic assignment. Quality filtering was done using dada2 default parameter values. For the 16S rRNA sequence raw data processing, the reads were trimmed to 220 bp for the forward read and to 65 bp for the reverse read. For the 18S rRNA sequence raw data processing, the reads were trimmed to 280 bp for the forward read and to 170 bp for the reverse read, allowing a maximum of 5 and 8 expected errors per read, respectively.

Target Organisms

Since a diverse range of bacteria has been shown to use various strategies for predation resistance, the dataset was first analyzed toward a set of known PRBs (Bertelli and Greub, 2012; **Table 1**). Phylogenies were estimated based on FastTree 2 (Price et al., 2010). To test whether a relationship existed between sampling environment (i.e., bays, offshore, and freshwater) and PRB detection frequency, we made use of Fisher's exact test (**Supplementary Table 1**) implemented in the R statistical software.

In order to include protozoa in the analysis based on knowledge of feeding style, including heterotrophs and mixotrophs, other phylogenetic groups were excluded from the eukaryotic dataset (Olenina et al., 2006). The supergroup Archaeplastida was excluded. For the supergroup Opisthokonta, Choanoflagellates were retained as they are filter feeders (Richter and Nitsche, 2017), while the Division of Metazoa and Fungi were excluded. For the supergroup Stramenopiles, heterotrophic and mixotrophic Chrysophyceae were included, since they have phagotrophic feeding styles. From the Division Ochrophyta, classes Chrysophyceae, Dictyochophyceae, and NA were retained, while non-relevant classes were excluded (Bolidophyceae, Bacillariophyta, MOCH-2, Phaeothamniophyceae, MOCH-5 Raphidophyceae, Synurophyceae, Eustigmatophyceae, Phaeophyceae, and Xanthophyceae). The sequence data were visualized using the R package phyloseq (McMurdie and Holmes, 2013) and ggplot2 (Wickham, 2016).

Microbial Source Tracking of PRB

The dataset with PRB bacterial taxa and phagotrophic protists was adjusted to the requirements of the analysis. To find the source of specific PRBs, we created a subset by filtering out taxa that did not meet the criteria of detection in a minimum of four samples. To assign PRB ASVs to their source environments, the Random Forest (RF) classification method implemented in the RandomForest R package was used (Liaw and Wiener, 2002). The source environments (i.e., training data) were defined as offshore ($n = 9$ samples), freshwater of river inflow ($n = 17$), and template-free PCR water ($n = 7$), with presence/absence data of PRB ASVs in all samples included. RF models were tuned using 10-fold cross-validation with a grid search of parameters to optimize the

predictive performance: ntree 500, 1,000, 2,000, 5,000 (number of trees in the forest), mtry 1–10 (number of candidates drawn to feed the algorithm), and maxnodes 5–15 (set the maximum amount of terminal nodes in the forest). The tuning procedure was performed with R packages caret (Kuhn, 2008) and e1071 (Meyer et al., 2020).

Direct Acyclic Graphs—BANJO

Dynamic Bayesian Network (DBN) was used to model interactions between PRB bacterial taxa and phagotrophic protists. The dataset with PRB bacterial taxa and phagotrophic protists with detection in a minimum of four samples was adjusted to the requirements of the analysis. First, the data were normalized such that the count of each sample was equal to the data (to compensate for the difference in the total number of reads between the samples). To account for within-group variation (such as niche differences), a subset of PRB and phagotrophic protozoa were gathered, respectively (**Supplementary Table 3, 4**). The DBN was created by collapsing the sequence datasets to the highest taxonomic resolution possible, 16S rRNA ASV table to genus, and 18S rRNA ASV data to species. The data were also collapsed by filtering out taxa that did not meet the criteria of detection in a minimum of four samples. Selected environmental variables were also included to infer dependencies (edges) between bacteria, protists, and environmental factors (nodes), and thus, potential drivers of community structure. Then, all features (i.e., bacteria and protists groups, and environmental variables) were normalized such that the mean value was centered at zero and the variation was transformed to one to be able to compare across datasets. For phagotrophic protists, the top 30 most abundant ASVs were then grouped to division. PRB ASVs were clustered using k-means clustering in R based on pairwise Pearson correlations between ASV abundance within the genus. For each genus, two partitions were chosen with the exception of the genus *Legionella*, which did not show any conflicting ASV abundance pattern.

The DBN was inferred (Smith et al., 2002) using BANJO version 2.2.0¹ (for an application of the use of BANJO in the community, see Metcalf et al., 2016) using the following analysis conditions: discretization = 2 intervals, max parents = 3, min lag = 0, max lag = 0, and Simulated Annealing search algorithm. These parameter choices were based on preliminary runs minimizing the negative log-likelihood score (i.e., resulted in best goodness-of-fit to the data). To aid the best fit and a logic network, environmental nodes were only permitted as parent nodes. Also, no edges between the bacteria were permitted. For other settings, default parameters were used.

Generalized Linear Latent Variable Models (GLLMs)

A generalized linear latent variable model was used to model interactions between PRB bacterial taxa and phagotrophic protists. The dataset was adjusted to the requirements of the

¹<http://www.cs.duke.edu/~amink/software/banjo/>

analysis in the same way as for the DBN analysis with both 16S rRNA ASV clusters within genus and 18S rRNA ASVs clustered at the species levels comprising the outcome data. The GLLVM was constructed using the R package *gllvm* (Niku et al., 2019). Gaussian distribution function (i.e., identity link) for the responses was selected. The following environmental variables were included in the model as continuous predictors: pH, DOC, phosphate (PO₄), salinity, chlorophyll a concentration (Chl-a), and ammonium (NH₄). To avoid including highly correlated variables as predictors in the model, a pre-selection step was performed based on the PCA score plot of the environmental variables (i.e., only one of a group of tightly clustered variables was selected). The selected variables were normalized so that the mean value and standard deviation were set to zero and one, respectively, for each included predictor. Design variables included were as follows: station ($i = 1, \dots, 4$), month of sampling ($j = 5, \dots, 9$), and the position of the sampling point in each bay ($k = 1, 2, 3$). Second-order interactions between all design variables were included, as these interactions greatly improved the goodness-of-fit of the model to the data. To improve convergence, jitter variance for starting values of latent variables was set to 0.5, and the number of initial runs was set to five. Three latent variables were selected to capture the residual variation following recommendations suggested in a previous study (Warton et al., 2015). Otherwise, default values of the GLLVMs function were used.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**. Supplementary data related to this article can be accessed at the Short Read Archive (<https://www.ncbi.nlm.nih.gov/sra>), project accession: PRJNA796828, ID: 796828.

REFERENCES

- Amaro, F., and Martín-González, A. (2021). Microbial warfare in the wild—the impact of protists on the evolution and virulence of bacterial pathogens. *Int. Microbiol.* 24, 559–571. doi: 10.1007/s10123-021-00192-y
- Andersson, A., Ahlinder, J., Mathisen, P., Hägglund, M., Bäckman, S., Nilsson, E., et al. (2018). Predators and nutrient availability favor protozoa-resisting bacteria in aquatic systems. *Sci. Rep.* 8:8415. doi: 10.1038/s41598-018-26422-4
- Andersson, A., Meier, H. E. M., Ripszám, M., Rowe, O., Wikner, J., Hägglund, P., et al. (2015). Projected future climate change and Baltic Sea ecosystem management. *Ambio* 44, 345–356. doi: 10.1007/s13280-015-0654-8
- Baker-Austin, C., Trinanes, J. A., Taylor, N. G. H., Hartnell, R., Siitonen, A., and Martínez-Urtaza, J. (2012). Emerging *Vibrio* risk at high latitudes in response to ocean warming. *Nat. Clim. Change* 3, 73–77.
- Bertelli, C., and Greub, G. (2012). Lateral gene exchanges shape the genomes of amoeba-resisting microorganisms. *Front. Cell. Infect. Microbiol.* 2:110. doi: 10.3389/fcimb.2012.00110
- Boamah, D. K., Zhou, G., Ensminger, A. W., and O'Connor, T. J. (2017). From many hosts, one accidental pathogen: the diverse protozoan hosts of *Legionella*. *Front. Cell. Infect. Microbiol.* 7:477. doi: 10.3389/fcimb.2017.00477
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857.

AUTHOR CONTRIBUTIONS

KE processed the samples for sequencing. KE and JA processed the data after sequencing and conducted data analyses. KE wrote the manuscript together with all co-authors. AA, JT, and JA organized the data collection and supervised the project. AA supervised the data collection. All authors contributed to the article and approved the submitted version.

FUNDING

This project was financed by the Industrial Doctoral School at Umeå University, the marine research environment EcoChange (FORMAS) and the Swedish Ministry of Defence (Grant No: A4001).

ACKNOWLEDGMENTS

Sonia Brugel and the field team are gratefully acknowledged for field samplings and sample analyses. We thank Umeå Marine Sciences Center for chemical analyses; Anders Andersson and Meike Latz for the guidance on 18S rRNA primer design; Stina Bäckman and Emelie Näslund Salomonsson for help with the sequencing; and Pia Haecky and Nick Blackburn for language improvement.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcimb.2022.877483/full#supplementary-material>

- Boscaro, V., Manassero, V., Keeling, P. J., and Vannini, C. (2022). Single-cell microbiomics unveils distribution and patterns of microbial symbioses in the natural environment. *Microb. Ecol.* [Online ahead of print]. doi: 10.1007/s00248-021-01938-x
- Brigmon, R. L., Turick, C. E., Knox, A. S., and Burckhalter, C. E. (2020). The impact of storms on *Legionella pneumophila* in cooling tower water, implications for human health. *Front. Microbiol.* 11:543589. doi: 10.3389/fcimb.2020.543589
- Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi: 10.1038/ismej.2012.8
- Carvalho, F. R. S., Vazoller, R. F., Foronda, A. S., and Pellizari, V. H. (2007). Phylogenetic study of legionella species in pristine and polluted aquatic samples from a tropical atlantic forest ecosystem. *Curr. Microbiol.* 55, 288–293. doi: 10.1007/s00284-006-0589-1
- Edgcomb, V., Orsi, W., Bunge, J., Jeon, S., Christen, R., Leslin, C., et al. (2011). Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *ISME J.* 5, 1344–1356. doi: 10.1038/ismej.2011.6
- Faust, K., and Raes, J. (2012). Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* 10, 538–550. doi: 10.1038/nrmicro2832
- Ferrantini, F., Fokin, S. I., Modeo, L., Andreoli, I., Dini, F., Görtz, H. D., et al. (2009). *Candidatus Cryptoprodotis polytropus*, A novel *Rickettsia*-like

- organism in the ciliated protist *Pseudomicrothorax dubius* (ciliophora, nassophorea). *J. Eukaryot. Microbiol.* 56, 119–129. doi: 10.1111/j.1550-7408.2008.00377.x
- Gomez-Valero, L., and Buchrieser, C. (2019). Intracellular parasitism, the driving force of evolution of *Legionella pneumophila* and the genus *Legionella*. *Genes Immun.* 20, 394–402.
- Graells, T., Ishak, H., Larsson, M., and Guy, L. (2018). The all-intracellular order *Legionellales* is unexpectedly diverse, globally distributed and lowly abundant. *FEMS Microbiol. Ecol.* 94:fy185.
- Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., et al. (2012). The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* 41, D597–D604. doi: 10.1093/nar/gks1160
- Gyraite, G., Katarzyte, M., and Schernewski, G. (2019). First findings of potentially human pathogenic bacteria *Vibrio* in the south-eastern Baltic Sea coastal and transitional bathing waters. *Mar. Pollut. Bull.* 149:110546. doi: 10.1016/j.marpolbul.2019.110546
- Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E. M., and Troedsson, C. (2014). Characterization of the 18S rRNA Gene for designing universal eukaryote specific primers. *PLoS One* 9:e87624. doi: 10.1371/journal.pone.0087624
- Hägglund, M., Bäckman, S., Macellaro, A., Lindgren, P., Borgmästars, E., Jacobsson, K., et al. (2018). Accounting for bacterial overlap between raw water communities and contaminating sources improves the accuracy of signature-based microbial source tracking. *Front. Microbiol.* 9:2364. doi: 10.3389/fmicb.2018.02364
- Herlemann, D. P. R., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J. J., and Andersson, A. F. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* 5, 1571–1579. doi: 10.1038/ismej.2011.41
- Hess, S., Suthaus, A., and Melkonian, M. (2016). “*Candidatus Finniella*” (*Rickettsiales, Alphaproteobacteria*), novel endosymbionts of viridiplanorid amoeboflagellates (*Cercozoa, Rhizaria*). *Appl. Environ. Microbiol.* 82, 659–670. doi: 10.1128/AEM.02680-15
- Hugerth, L. W., Muller, E. E. L., Hu, Y. O. O., Lebrun, L. A. M., Roume, H., Lundin, D., et al. (2014). Systematic design of 18S rRNA gene primers for determining eukaryotic diversity in microbial consortia. *PLoS One* 9:e95567. doi: 10.1371/journal.pone.0095567
- Hutalle-Schmelzer, K. M. L., and Grossart, H. P. (2009). Changes in the bacterioplankton community of oligotrophic Lake Stechlin (northeastern Germany) after humic matter addition. *Aquat. Microb. Ecol.* 55, 155–167.
- Kao, P.-M., Tung, M.-C., Hsu, B.-M., Hsu, S.-Y., Huang, J.-T., Liu, J.-H., et al. (2013). Differential *Legionella* spp. survival between intracellular and extracellular forms in thermal spring environments. *Environ. Sci. Pollut. Res.* 20, 3098–3106. doi: 10.1007/s11356-012-1159-7
- Kuhn, M. (2008). Building predictive models in R using the caret package. *J. Statist. Softw.* 28, 1–26.
- Lamrabet, O., Merhej, V., Pontarotti, P., Raoult, D., and Drancourt, M. (2012). The genealogical tree of mycobacteria reveals a long-standing sympatric life into free-living protozoa. *PLoS One* 7:e34754. doi: 10.1371/journal.pone.0034754
- Latz, M. A. C., Grujic, V., Brugel, S., Lycken, J., John, U., Karlson, B., et al. (2022). Short- and long-read metabarcoding of the eukaryotic rRNA operon: evaluation of primers and comparison to shotgun metagenomics sequencing. *Mol. Ecol. Resour.* [Online ahead of print]. doi: 10.1111/1755-0998.13623
- Lesnik, R., Brettar, I., and Höfle, M. G. (2015). *Legionella* species diversity and dynamics from surface reservoir to tap water: from cold adaptation to thermophily. *ISME J.* 10, 1064–1080. doi: 10.1038/ismej.2015.199
- Liaw, A., and Wiener, M. (2002). Classification and Regression by randomForest. *R News* 2, 18–22.
- Lindh, M. V., Figueroa, D., Sjöstedt, J., Baltar, F., Lundin, D., Andersson, A., et al. (2015). Transplant experiments uncover Baltic Sea basin-specific responses in bacterioplankton community composition and metabolic activities. *Front. Microbiol.* 6:223. doi: 10.3389/fmicb.2015.00223
- Lönnstrom, L., Wiklund, T., and Bylund, G. (1994). *Pseudomonas anguilliseptica* isolated from Baltic herring *Clupea harengus* membras with eye lesions. *Dis. Aquat. Organ.* 18, 143–147.
- Luczkiewicz, A., Kotlarska, E., Artichowicz, W., Tarasewicz, K., and Fudala-Ksiazek, S. (2015). Antimicrobial resistance of *Pseudomonas* spp. isolated from wastewater and wastewater-impacted marine coastal zone. *Environ. Sci. Pollut. Res.* 22, 19823–19834. doi: 10.1007/s11356-015-5098-y
- Matz, C., Bergfeld, T., Rice, S. A., and Kjelleberg, S. (2004). Microcolonies, quorum sensing and cytotoxicity determine the survival of *Pseudomonas aeruginosa* biofilms exposed to protozoan grazing. *Environ. Microbiol.* 6, 218–226. doi: 10.1111/j.1462-2920.2004.00556.x
- Matz, C., and Kjelleberg, S. (2005). Off the hook – how bacteria survive protozoan grazing. *Trends Microbiol.* 13, 302–307. doi: 10.1016/j.tim.2005.05.009
- Matz, C., Moreno, A. M., Alhede, M., Manfield, M., Hauser, A. R., Givskov, M., et al. (2008). *Pseudomonas aeruginosa* uses type III secretion system to kill biofilm-associated amoebae. *ISME J.* 2, 843–852. doi: 10.1038/ismej.2008.47
- McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., et al. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 6, 610–618. doi: 10.1038/ismej.2011.139
- McMurdie, P. J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. doi: 10.1371/journal.pone.0061217
- Metcalf, J. L., Xu, Z. Z., Weiss, S., Lax, S., Van Treuren, W., Hyde, E. R., et al. (2016). Microbial community assembly and metabolic function during mammalian corpse decomposition. *Science* 351, 158–162. doi: 10.1126/science.aad2646
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. (2020). *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-4*. Available online at: <https://cran.r-project.org/package=e1071> [accessed on September 16, 2021].
- Miyasaka, J., Yahiro, S., Arahira, Y., Tokunaga, H., Katsuki, K., and Hara-Kudo, Y. (2006). Isolation of *Vibrio parahaemolyticus* and *Vibrio vulnificus* from wild aquatic birds in Japan. *Epidemiol. Infect.* 134, 780–785. doi: 10.1017/S0950268805005674
- Newton, R. J., and McLellan, S. L. (2015). A unique assemblage of cosmopolitan freshwater bacteria and higher community diversity differentiate an urbanized estuary from oligotrophic Lake Michigan. *Front. Microbiol.* 6:1028. doi: 10.3389/fmicb.2015.01028
- Niku, J., Hui, F. K. C., Taskinen, S., and Warton, D. I. (2019). gllvm: fast analysis of multivariate abundance data with generalized linear latent variable models in R. *Methods Ecol. Evol.* 10, 2173–2182.
- Niku, J., Warton, D. I., Hui, F. K. C., and Taskinen, S. (2017). Generalized linear latent variable models for multivariate count and biomass data in ecology. *J. Agric. Biol. Environ. Statist.* 22, 498–522. doi: 10.1371/journal.pone.0216129
- Nishida, T., Hara, N., Watanabe, K., Shimizu, T., Fujishima, M., and Watarai, M. (2018). Crucial role of *Legionella pneumophila* ToxC in the inhibition of cellular trafficking in the protistan host *Paramecium tetraurelia*. *Front. Microbiol.* 9:800. doi: 10.3389/fmicb.2018.00800
- Olenina, I., Hajdu, S., Edler, L., Andersson, A., Wasmund, N., Busch, S., et al. (2006). Biovolumes and size-classes of phytoplankton in the Baltic Sea HELCOM. *Balt. Sea Environ. Proc.* 106, 1–144.
- Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., et al. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecol. Lett.* 20, 561–576. doi: 10.1111/ele.12757
- Pappa, O., Beloukas, A., Vantarakis, A., Mavridou, A., Kefala, A.-M., and Galanis, A. (2017). Molecular characterization and phylogenetic analysis of *Pseudomonas aeruginosa* isolates recovered from greek aquatic habitats implementing the double-locus sequence typing scheme. *Microb. Ecol.* 74, 78–88. doi: 10.1007/s00248-016-0920-8
- Parthuisot, N., West, N. J., Lebaron, P., and Baudart, J. (2010). High diversity and abundance of *Legionella* spp. in a pristine river and impact of seasonal and anthropogenic effects. *Appl. Environ. Microbiol.* 76, 8201–8210. doi: 10.1128/AEM.00188-10
- Pent, M., Pöldmaa, K., and Bahram, M. (2017). Bacterial communities in boreal forest mushrooms are shaped both by soil parameters and host identity. *Front. Microbiol.* 8:836. doi: 10.3389/fmicb.2017.00836
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490

- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., et al. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196. doi: 10.1093/nar/gkm864
- Qiao, Y., Liu, J., Zhao, M., and Zhang, X. H. (2018). Sediment depth-dependent spatial variations of bacterial communities in mud deposits of the Eastern China marginal seas. *Front. Microbiol.* 9:1128. doi: 10.3389/fmicb.2018.01128
- Renaud, G., Stenzel, U., Maricic, T., Wiebe, V., and Kelso, J. (2015). deML: robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics* 31, 770–772. doi: 10.1093/bioinformatics/btu719
- Richter, D. J., and Nitsche, F. (2017). *Choanoflagellata. Pages 1479–1496 Handbook of the Protists*, 2nd Edn. Berlin: Springer International Publishing.
- Rizzo, C., Malavenda, R., Gerçe, B., Papale, M., Sylđatk, C., Hausmann, R., et al. (2019). Effects of a simulated acute oil spillage on bacterial communities from arctic and antarctic marine sediments. *Microorganisms* 7:632. doi: 10.3390/microorganisms7120632
- Roguet, A., Eren, A. M., Newton, R. J., and McLellan, S. L. (2018). Fecal source identification using random forest. *Microbiome* 6:185. doi: 10.1186/s40168-018-0568-3
- Shankar, A., Jagajeedas, D., Radhakrishnan, M. P., Paul, M., Narendrakumar, L., Suryaletha, K., et al. (2021). Elucidation of health risks using metataxonomic and antibiotic resistance profiles of microbes in flood affected waterbodies, Kerala 2018. *J. Flood Risk Manage.* 14:e12673.
- Shimada, S., Nakai, R., Aoki, K., Shimoeda, N., Ohno, G., Kudoh, S., et al. (2020). Chasing waterborne pathogens in antarctic human-made and natural environments, with special reference to *Legionella* spp. *Appl. Environ. Microbiol.* 87:e02247-20. doi: 10.1128/AEM.02247-20
- Singh, S., Eldin, C., Kowalczywska, M., and Raoult, D. (2013). Axenic culture of fastidious and intracellular bacteria. *Trends Microbiol.* 21, 92–99. doi: 10.1016/j.tim.2012.10.007
- Smith, V. A., Jarvis, E. D., and Hartemink, A. J. (2002). Evaluating functional network inference using simulations of complex biological systems. *Bioinformatics* 18, S216–S224. doi: 10.1093/bioinformatics/18.suppl_1.s216
- Solomon, J. M., Leung, G. S., and Isberg, R. R. (2003). Intracellular replication of *Mycobacterium marinum* within *Dictyostelium discoideum*: efficient replication in the absence of host coronin. *Infect. Immun.* 71, 3578–3586. doi: 10.1128/IAI.71.6.3578-3586.2003
- Sonne, C., Lakemeyer, J., Desforgues, J.-P., Eulaers, I., Persson, S., Stokholm, I., et al. (2020). A review of pathogens in selected Baltic Sea indicator species. *Environ. Int.* 137:105565. doi: 10.1016/j.envint.2020.105565
- Steinert, M., and Heuner, K. (2005). *Dictyostelium* as host model for pathogenesis. *Cell. Microbiol.* 7, 307–314.
- Sun, S., Noorian, P., and McDougald, D. (2018). Dual role of mechanisms involved in resistance to predation by protozoa and virulence to humans. *Front. Microbiol.* 9:1017. doi: 10.3389/fmicb.2018.01017
- Szokoli, F., Castelli, M., Sabaneyeva, E., Schrallhammer, M., Krenek, S., Doak, T. G., et al. (2016). Disentangling the taxonomy of *Rickettsiales* and description of two novel symbionts (“*Candidatus* Bealeia paramacronuclearis” and “*Candidatus* Fokinia cryptica”) sharing the cytoplasm of the ciliate protist *Paramecium biaurelia*. *Appl. Environ. Microbiol.* 82, 7236–7247. doi: 10.1128/AEM.02284-16
- Thelaus, J., Andersson, A., Mathisen, P., Forslund, A.-L., Noppa, L., and Forsman, M. (2009). Influence of nutrient status and grazing pressure on the fate of *Francisella tularensis* in lake water. *Fems Microbiol. Ecol.* 67, 69–80. doi: 10.1111/j.1574-6941.2008.00612.x
- Thelaus, J., Forsman, M., and Andersson, A. (2008). Role of productivity and protozoan abundance for the occurrence of predation-resistant bacteria in aquatic systems. *Microb. Ecol.* 56, 18–28. doi: 10.1007/s00248-007-9320-4
- Vannini, C., Petroni, G., Verni, F., and Rosati, G. (2005). A bacterium belonging to the *Rickettsiaceae* family inhabits the cytoplasm of the marine ciliate *Diophrys appendiculata* (Ciliophora, Hypotrichia). *Microb. Ecol.* 49, 434–442. doi: 10.1007/s00248-004-0055-1
- Von Reyn, C. F., Waddell, R. D., Eaton, T., Arbeit, R. D., Maslow, J. N., Barber, T. W., et al. (1993). Isolation of *Mycobacterium avium* complex from water in the United States, Finland, Zaire, and Kenya. *J. Clin. Microbiol.* 31, 3227–3230. doi: 10.1128/jcm.31.12.3227-3230.1993
- Walters, W. A., Caporaso, J. G., Lauber, C. L., Berg-Lyons, D., Fierer, N., and Knight, R. (2011). PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics* 27, 1159–1161. doi: 10.1093/bioinformatics/btr087
- Warton, D. I., Blanchet, F. G., O’Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., et al. (2015). So many variables: joint modeling in community ecology. *Trends Ecol. Evolut.* 30, 766–779. doi: 10.1016/j.tree.2015.09.007
- Watanabe, K., Suzuki, H., Nishida, T., Mishima, M., Tachibana, M., Fujishima, M., et al. (2018). Identification of novel *Legionella* genes required for endosymbiosis in *Paramecium* based on comparative genome analysis with *Holospira* spp. *FEMS Microbiol. Ecol.* 94:fy162. doi: 10.1093/femsec/fy162
- Watanabe, K., and Watarai, M. (2017). *Legionella* and the protist host. *Yamaguchi J. Vet. Med.* 1–8.
- Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., et al. (2016). Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J.* 10, 1669–1681. doi: 10.1038/ismej.2015.235
- Weitere, M., Bergfeld, T., Rice, S. A., Matz, C., and Kjelleberg, S. (2005). Grazing resistance of *Pseudomonas aeruginosa* biofilms depends on type of protective mechanism, developmental stage and protozoan feeding mode. *Environ. Microbiol.* 7, 1593–1601. doi: 10.1111/j.1462-2920.2005.00851.x
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Berlin: Springer International Publishing.
- Wiklund, T. (2016). *Pseudomonas anguilliseptica* infection as a threat to wild and farmed fish in the Baltic Sea. *Microbiol. Aust.* 37, 135–136.
- Winstanley, C., O’Brien, S., and Brockhurst, M. A. (2016). *Pseudomonas aeruginosa* evolutionary adaptation and diversification in cystic fibrosis chronic lung infections. *Trends Microbiol.* 24, 327–337. doi: 10.1016/j.tim.2016.01.008

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Eriksson, Thelaus, Andersson and Ahlinder. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.