

Microbial metagenomics in the Baltic Sea: Recent advancements and prospects for environmental monitoring

Karolina Ininbergs, Birgitta Bergman,
John Larsson, Martin Ekman

Abstract Metagenomics refers to the analysis of DNA from a whole community. Metagenomic sequencing of environmental DNA has greatly improved our knowledge of the identity and function of microorganisms in aquatic, terrestrial, and human biomes. Although open oceans have been the primary focus of studies on aquatic microbes, coastal and brackish ecosystems are now being surveyed. Here, we review so far published studies on microbes in the Baltic Sea, one of the world's largest brackish water bodies, using high throughput sequencing of environmental DNA and RNA. Collectively the data illustrate that Baltic Sea microbes are unique and highly diverse, and well adapted to this brackish-water ecosystem, findings that represent a novel base-line knowledge necessary for monitoring purposes and a sustainable management. More specifically, the data relate to environmental drivers for microbial community composition and function, assessments of the microbial biodiversity, adaptations and role of microbes in the nitrogen cycle, and microbial genome assembly from metagenomic sequences. With these discoveries as background, prospects of using metagenomics for Baltic Sea environmental monitoring are discussed.

Keywords Microbial metagenomics · Baltic Sea · Environmental monitoring · Brackish microbial communities · Bacterial diversity

INTRODUCTION

Abundant and ubiquitous environmental microorganisms are important drivers of global biogeochemical cycles, and understanding factors controlling their abundance, activities, and diversity is therefore a major area of research. Studying microbes in situ, however, is challenging.

Microbial communities are composed of mixed and diverse assemblages, often reside in hard to sample habitats, and a great majority is un-cultivable in the laboratory. The development of high throughput sequencing (HTS) technologies during the last decade has therefore in many ways revolutionized the study of natural communities of microbes, now characterized as the “unseen majority.” This technology allows, in a cost-effective way, the analysis of diversity, metabolic functions, and biological interactions in complex, uncultured microbial communities. Extraction of DNA from mixed communities of microbes, followed by HTS, has greatly increased our understanding of the fundamental roles played by terrestrial, aquatic, and human-associated microbiota (Fierer et al. 2007; Zinger et al. 2011; Huttenhower et al. 2012). Pioneering work analyzing sequenced environmental DNA (eDNA) from marine bacterial communities led to the discovery that much of the microbial diversity in the world's oceans, from surface to deep waters, had been severely underestimated (Venter et al. 2004; DeLong et al. 2006). Important findings from HTS of eDNA or RNA include the discovery of novel genes, proteins, and microbial species (Yooseph et al. 2007; Gilbert et al. 2008), and findings related to the role of microbes in global biogeochemical cycles of carbon and nitrogen (Frias-Lopez et al. 2008). More recent analyses have also shown that genomic plasticity and metabolic versatility of microorganisms are the basis for bacterial adaptation in marine ecosystems (Konstantinidis et al. 2009; Lauro et al. 2009; Yooseph et al. 2010). Hence, the development and introduction of these potent technologies have enabled a much more detailed view of microbial communities and their functions in natural settings and have profoundly changed our perception of microbial life, genome evolution, and minimal requirements for life (Karl 2007; Giovannoni et al. 2014).

HTS technology-based analyses of natural microbial communities can be divided into two sub-classes. The more commonly used “HTS-signature-gene” approach surveys eDNA for single marker gene distributions and abundances and often uses the conserved 16S rRNA gene, encoding the small subunit of the prokaryotic ribosome. In contrast, the metagenomic shotgun approach potentially exposes the majority of genes/genomes present in eDNA extracted from natural microbial populations. Due to the non-targeted nature of the metagenomic approach, substantial amounts of diverse genetic data and information on functional potential of entire microbial communities may be obtained. Metagenomics, however, holds its own biases. As a result of differences in genome sizes and genetic sequence (e.g., high or low GC content), DNA from different genes and organisms are not uniformly covered during sequencing, nor can all sequences be correctly annotated (identified) due to lack of experimental evidence for protein-coding sequences and the still limited number of sequenced microbial genomes available in databases. Metagenomics is increasingly often combined with metatranscriptomics for which all RNA molecules in a natural sample are targeted, expanding the scope of metagenomics by also providing information on gene expression.

Most early metagenomic studies focused on oceans, while smaller seas, freshwater systems, and, in particular, brackish-water transitions until recently have remained poorly investigated using HTS. The Baltic Sea is the world’s second largest body of brackish water and represents one of the most intensely researched and monitored aquatic environments, with a time series of hydrographic data measured routinely for over 100 years (Fonselius and Valderrama 2003). The contemporary Baltic Sea is, since the turn of the twentieth century, negatively affected by anthropogenic disturbances, specifically urban and agricultural derived eutrophication, fueling phytoplankton blooms causing increased anoxia and hypoxia in deep waters (Savage et al. 2010; Carstensen et al. 2014). In addition, the Baltic Sea offers steep gradients in salinity and key-nutrient concentrations. These gradients are semi-constant over time but change dramatically over a short geographical distance, giving rise to a challenging environment for many marine and freshwater organisms. Paired with geographic isolation, this has resulted in low species and genetic diversity among metazoans (e.g., fish, seals) and macrophytes (e.g., macro-algae) (Johannesson and Andre 2006). The diversity and biogeography of microbes in the Baltic Sea and associated waters have, however, attracted considerably less research.

Here, we provide an updated account on how the introduction of HTS-based analyses into Baltic Sea microbial research has contributed to the understanding of Baltic Sea microbes. Recent data collectively illustrate that Baltic Sea

microbes are both unique and highly diverse, and well adapted to this brackish water ecosystem. To widen the scope, the Baltic Sea findings are placed in the context of metagenomic microbial findings in oceans in general. Finally, the ecological significance of microbes in any environment suggests a need for implementation of HTS data into future environmental monitoring programs, prospects of which are discussed here in a Baltic Sea perspective.

ENVIRONMENTAL SEQUENCING OF BALTIC SEA MICROBES

Earlier reports on Baltic Sea microorganisms focused largely on quantifying abundance and activity in relation to physicochemical parameters (see e.g., Hagström and Larsson 1984; Gast and Gocke 1988; Rheinheimer et al. 1989). The introduction of molecular-based approaches expanded our knowledge of microbial community composition, seasonal succession, and phylogenetic diversity, revealing for instance temporal patterns for specific bacterial phylotypes (Pinhassi et al. 1997; Pinhassi and Hagström 2000) and strong influence of freshwater phyla on brackish water communities (Riemann et al. 2008). A pre-HTS-era metagenomic analysis (using a cloning approach) of Baltic Sea sediment microbial communities was published in 2007 (Hårdeman and Sjöling 2007), while the first HTS-signature-gene-based (pyro-sequencing) study appeared in 2010 (Andersson et al. 2010). This study targeted bacterioplankton at the Landsort Deep, the deepest (459 m) location, and long-term monitoring site in the Baltic Sea (Baltic proper). Subsequently, a number of HTS-based studies of the Baltic Sea microbial life have followed. As seen in Table 1, six studies are based on the HTS-signature-gene approach and four on random metagenomic sequencing of all genes. The work of Feike et al. (2012) stands out by being the only purely metatranscriptomic analysis, and yet another recent study describes genome assembly based on metagenomic sequences (Herlemann et al. 2013).

BACTERIAL BIOGEOGRAPHY AND DIVERSITY

In addition to steep horizontal and vertical concentration gradients related to salinity and key-nutrients, such as nitrogen and phosphorous, there is also a pronounced seasonal variation in both nutrient concentrations and temperature in the Baltic Sea. The effects of these variables were investigated by Andersson et al. (2010) using a 16S rRNA HTS approach to study bacterioplankton communities at Landsort Deep. A pronounced influence of both phosphorous and temperature on the microbial community,

Table 1 Summary of HTS studies of microorganisms in the Baltic Sea

No. of sampling locations/ time points	Sequencing platform	Target DNA/RNA—reads/genes	Number of sequenced reads	Reference
21/1	454 Pyro-sequencing	DNA all genes and—16S rRNA	~20 900 000 1 247 371 16S rRNA	Dupont et al. (2014)
21/1	454 Pyro-sequencing & Illumina	DNA—all cyanobacterial genes	698 865 (454 reads) 7 458 747 084 (Illumina reads)	Larsson et al. (2014)
60/Mesocosm	454 Pyro-sequencing	DNA—16S rRNA	135 037	Herlemann et al. (2014)
3/1	454 Pyro-sequencing	DNA—all genes	1 205 630	Thureborn et al. (2013)
16/Mesocosm	454 Pyro-sequencing	DNA—16S rRNA	97 582	Dinasquet et al. (2013)
215/1	454 Pyro-sequencing	DNA—all genes (genome binning)	37 658 923	Herlemann et al. (2013)
10/12	454 Pyro-sequencing	DNA/RNA— <i>nifH</i>	79 090	Farnelid et al. (2013)
5 Sampling procedures/1	454 Pyro-sequencing	RNA—all genes	190 262	Feike et al. (2012)
213/1	454 Pyro-sequencing	DNA—16S rRNA	224 076 ^a	Herlemann et al. (2011)
11/1	454 Pyro-sequencing	DNA—16S rRNA	36 108	Koskinen et al. (2011)
1/8	454 Pyro-sequencing	DNA—16S rRNA	162 256	Andersson et al. (2010)

^a Calculated based on sample average

notably composed of characteristic freshwater bacteria such as actinobacteria, betaproteobacteria, and verrucomicrobia, was found. While some early studies suggested that salinity may influence the Baltic Sea bacterial growth and biogeography (Väättänen 1980; Heinanen 1991), the extent of this effect was not realized until the first large-scale 16S rRNA gene inventory along the entire Baltic Sea salinity gradient was performed (Herlemann et al. 2011). A subsequent comprehensive metagenomic survey substantiated the strong structuring of the bacterial community composition along the Baltic salinity gradient and also included the freshwater Lake Torne Träsk and the marine waters off the Swedish west coast as additional reference points (Dupont et al. 2014). Among eubacteria, a clear dominance of actinobacteria was apparent in the low-salinity Bothnian Bay in the northern Baltic Sea, while a shift toward dominance of proteobacteria (mainly alpha and gamma) was apparent at higher salinities (including at the Swedish west coast). As in most global oceans, the dominant bacteria in the Baltic Sea were alphaproteobacteria of the SAR11 clade (Morris et al. 2002; Dupont et al. 2014). Overall the community compositions at the phylum level in the two studies were largely in agreement and both reported a unique autochthonous brackish bacterial population present at intermediate salinity stations, including strains of SAR11 and picocyanobacteria (Herlemann et al. 2011; Dupont et al. 2014).

The variation in bacterial community composition seen in the Baltic Sea along the salinity transect is considerably more dramatic than in most oceanic habitats (Dupont et al. 2014). A comparative network-analysis of metagenomes

(both with respect to taxonomy and functional potential) collected at various depths (from surface to anoxic sediments) in the Baltic Sea and 27 metagenomes from 11 sites worldwide showed that the Baltic Sea bacterial communities clustered primarily with metagenomes from the western English Channel (Thureborn et al. 2013). Unique to the Baltic Sea, however, was the community derived from the metagenome collected at the oxic-anoxic interface, being an outlier with few taxonomic similarities to any other community (Thureborn et al. 2013). It should, however, be emphasized that the reference metagenomes used in these analyses were all from marine environments (except two that were from terrestrial), and that only one geographic site in the Baltic Sea was included. The increased availability of metagenomes from the whole Baltic Sea salinity transect now warrants expanded comparative studies.

Other findings from comparing HTS analyses from the Baltic Sea to other marine environments relate to the diversity of the Baltic Sea microbial community. In one of the first HTS/16S rRNA based studies, a lower bacterial diversity was observed in the central eastern Baltic Sea (northern Baltic proper/Gulf of Finland) compared to that of some investigated fully marine oceanic habitats (Koskinen et al. 2011). However, later HTS-based analyses (16S rRNA and metagenomic) of the genetic diversity of bacteria along the whole longitudinal expansion of the Baltic Sea did not show any such reduced diversity at intermediate salinities (Herlemann et al. 2011; Dupont et al. 2014). In contrast, these studies rather demonstrate a surprisingly high microbial diversity throughout the Baltic Sea. This suggests that Baltic Sea microorganisms are less

impaired by genetic isolation than are macro-organisms (Johannesson and Andre 2006), presumably underpinned by a combination of short microbial doubling times (hours/days) and comparatively small and flexible genomes. A characteristic of microbes is also frequent horizontal gene transfer events between sympatric microbes, a phenomenon that was recently investigated in Baltic Sea picocyanobacteria (Larsson et al. 2011).

MICROBIAL COMMUNITY FUNCTION: FROM GENE FREQUENCIES TO GENOME ASSEMBLY

Metagenomic data may be used to predict metabolic potentials of microbes as they expose existing gene repertoires and related metabolic processes. Although other factors such as regulation of gene transcription and enzyme activities, and the availability of substrates, are of critical importance, the relative frequencies of specific genes may point to their functional importance in an environment. While the effect of salinity in shaping microbial communities is well known (see e.g., Lozupone and Knight 2007; Campbell and Kirchman 2013), the mechanism behind salinity being such a strong barrier to cross for bacteria is not. However, a recent metagenomic analysis, encompassing the Baltic Sea salinity gradient, revealed that salinity does not only influence the distribution of traits such as ion transporters (e.g., Na, K) and biosynthesis and transport of compatible solutes, but also bacterial metabolic core functions such as respiration, glycolysis, and cofactor biosynthesis (Dupont et al. 2014). Notably, analogous metabolic pathways, with approximately the same outcome but via different intermediate metabolites and genes, were found to have opposite abundance patterns along the salinity gradient. For instance, the glycolytic Entner Doudoroff (ED) pathway dominated at high salinity while the Embden–Meyerhof (EM) pathway at low salinity (Dupont et al. 2014). Low-salinity adapted bacteria overall use pathways with a higher ATP yield than bacteria in marine environments. These metabolic differences may explain the distinct divide known to exist between fresh and marine microbial communities. It further suggests that adaptation to a lower salinity may be based on a core gene set with higher energy yield. This discovery now calls for further exploration.

Other important findings from metagenomic functional analyses of the Baltic Sea microbes relate to the impact of eutrophication and pollution. Nutrients (nitrogen and phosphorus) are the second most important factor (besides salinity) in shaping the distribution of microbial taxa and their functional potential in Baltic Sea surface water communities (Dupont et al. 2014). More specific information on the influence of eutrophication and pollution was

obtained from a metagenomic study of the Landsort Deep microbial community (Thureborn et al. 2013). The functional gene repertoire showed a comparatively high abundance of microbial genes involved in attachment to and degradation of organic carbon and in heavy metal resistance (e.g., against cobalt, cadmium, and zinc). These findings are likely related to organic matter deposition and the high concentrations of metals in sediments at this site (Thureborn et al. 2013). Overall, the resulting gene diversity at this specific site appears to be shaped by anthropogenic pollution and eutrophication. However, it should be noted that the deeper waters at this site offer radically different conditions (hypoxia or anoxia) compared to the rest of the Baltic Sea, affecting the overall bacterial biodiversity dramatically (Dupont et al. 2014). The effect of dissolved organic carbon (DOC) on the Baltic Sea bacterial community was also recently investigated by combining mesocosms and metagenomic analyses (Dinasquet et al. 2013; Herlemann et al. 2014). While a weak effect of the DOC on structuring the community was the norm, some taxa were clearly influenced. This approach may be an efficient tool to evaluate details in successional changes as a response to nutrient regimes.

To further explore putative functions, (meta-)genomic assembly data in the form of contigs, i.e., longer genomic sequences obtained by aligning partly overlapping sequence reads, may be analyzed. For example, a recent analysis of light-harvesting genes in Baltic Sea identified a novel gene cluster in the picocyanobacterial population (Larsson et al. 2014). These cyanobacteria, with a cell size $<2 \mu\text{m}$, are major primary producers in oceans (Scanlan et al. 2009). This is also the case in the Baltic Sea where picocyanobacteria may constitute up to 80 % of the cyanobacterial population (Stal et al. 2003; Hajdu et al. 2007). Metagenomic analyses show that the Baltic Sea picocyanobacteria are dominated by strains belonging to the genera *Synechococcus* and *Cyanobium* (unpublished results) and that members of the dominant *Synechococcus* clade harbor a novel gene cluster encoding proteins for a unique set of light-harvesting antennae, i.e., pigment-associated phycobilisomes, not previously found in cyanobacteria (Larsson et al. 2014). The organization of the gene cluster suggests the involvement of multiple horizontal gene transfer events. The Baltic Sea picocyanobacteria may have evolved a set of phycobiliproteins with a potentially unique absorption spectrum, to specifically match light conditions offered by the Baltic Sea. These findings exemplify how “meta-omic” datasets can provide novel insights and generate hypotheses-driven research, particularly targeting processes in the large segment of still non-cultivable aquatic microbes, including those in the Baltic Sea.

With recent developments in bioinformatic analysis of metagenomic data, it is possible to assemble not only

contigs of a limited length but also assemble near complete genomes of abundant organisms (Iverson et al. 2012). For example, metagenomic time-series samples from the Baltic Sea were used to assemble a genome from an aquatic phylotype of the verrucomicrobia *Spartobacteria* (Herlemann et al. 2013), which is one of the dominant organisms in the Baltic Sea bacterial community during the summer (Herlemann et al. 2011). Analysis of the assembled *Spartobacteria* genome gave important information about the metabolic capacity of the bacterium, including the presence of 23 glycoside hydrolases, giving the bacterium the capacity to metabolize a number of different carbohydrates and suggesting a potentially important role in carbon cycling. Based on patterns of co-occurring abundances, it was further suggested that the carbon was mainly derived from cyanobacterial blooms. It is clear that this approach can take the metagenomic scope even further and provides an important step in increasing the number of microbial genomes available, one of the prerequisites for correct annotation of metagenomic sequences.

THE NITROGEN CYCLE: FROM ANOXIC ZONES TO SURFACE WATERS

Today, the Baltic Sea suffers from large and persistent anoxic bottom zones. This is partly a natural phenomenon caused by strong stratification, which prevents vertical mixing. Eutrophication of the Baltic Sea has increased the area of these zones (Carstensen et al. 2014), and they today constitute the largest anthropogenically induced hypoxic area in the world.

The nitrogen cycle consists of microbially mediated transformations of nitrogen, some of which are dependent on reducing conditions (e.g., denitrification and anamox). Research concerning hypoxic environments in the Baltic Sea has therefore often targeted these processes, more recently using HTS technologies. Microbial metagenomes from the Landsort Deep water column illustrate a stratification of the microbial functional capacities along the depth and oxygen profile (Thureborn et al. 2013; Dupont et al. 2014). While genes for the anamox reaction were absent, high frequency of genes involved in denitrification prevailed at the deepest anoxic sites (Thureborn et al. 2013). It was furthermore suggested that the denitrification at this depth was primarily carried out by chemolithotrophic (sulfur oxidizing) denitrifying epsilonbacteria. Later, also Dupont et al. (2014) observed a high prevalence of epsilonbacteria in these specific waters. Together the findings suggest an important role of these organisms in denitrification at the Landsort Deep.

The substrate for epsilonbacterial denitrification (nitrate) was in the Thureborn et al. (2013) suggested to originate

from aerobic ammonia oxidizing thaumarchaeota. High abundance of ammonia oxidizing thaumarchaeota was indeed previously observed in the suboxic zones (70–120 m depth) of the central Baltic Sea (Labrenz et al. 2010). More recently, metatranscriptome analyses substantiated these findings by showing that the transcript level of ammonia oxidation genes (*amoA*, *amoB*, and *amoC*) was high in the suboxic zones of both Landsort and Gotland Deep (Feike et al. 2012). The nitrifying potential of pelagic archaea was first demonstrated by an early metagenomic study (Venter et al. 2004). Since then ammonia-oxidizing archaea have been found to be widely distributed in the world's oceans and likely play a significant role in the global nitrogen cycle (Erguder et al. 2009).

Fixation of atmospheric dinitrogen is a microbial process that has received considerable research attention (Gruber 2005). This is particularly the case for the Baltic Sea, with its typical massive summer blooms of nitrogen-fixing cyanobacteria (Stal et al. 2003, Kahru and Elmgren 2014). In fact, the nitrogen fixation by the large filamentous cyanobacterial blooms represents the second largest source of “new” nitrogen (N) input into the Baltic Sea after riverine load (Larsson et al. 2001). The principal enzyme that catalyzes nitrogen fixation, nitrogenase, is encoded by highly conserved *nif* genes (*nifKDH* encoding the structural protein), and these genes are obvious targets in metagenomic and metatranscriptomic surveys. Metagenomic surveys for *nif* genes in the surface waters of the ocean have found surprisingly few sequence reads, despite high rates of nitrogen fixation repeatedly recorded by bloom-forming cyanobacteria (Johnston et al. 2005). The reasons for rare *nif* gene findings in metagenomes may be due to a comparatively restricted distribution of these genes among organisms in the massive microbial metagenomic datasets (Johnston et al. 2005). To address this difficulty, it has been suggested that to properly expose all potential nitrogen-fixers, a minimum set of six *nif* genes should be targeted, namely *nifHDK* and *nifENB*, (Dos Santos et al. 2012). Even when including these *nif* genes in metagenomic analysis of Baltic Sea surface waters, few *nif* sequences were retrieved (Thureborn et al. 2013). This was explained by pre-bloom sampling and to the use of a pre-filtration step (<3.0 μm), which may have excluded the dominant larger nitrogen-fixing filamentous cyanobacteria. A considerably higher number of *nif* gene sequences were later retrieved from a Baltic Sea metagenomic dataset sampled in July and including larger sized microbes (3.0–200 μm) (Dupont et al. 2014).

Despite the relatively few *nif* sequences retrieved from surface waters in the Baltic Sea, *nif* gene abundances increased with depth at Landsort Deep (Thureborn et al. 2013). This suggests the involvement of heterotrophic organisms, in this case, sulfate-reducing Deltaproteobacteria

(comprising 36 % of the *nif* genes at this site), in Baltic Sea nitrogen fixation. Combining *nifH* HTS analyses, gene expression measurements and nitrogen fixation rate determinations for Baltic Sea microbes (Farnelid et al. 2013) showed that heterotrophic nitrogen fixation may account for up to 6 % of the total annual nitrogen fixation. Heterotrophic nitrogen fixation rates have recently been documented in hypoxic waters of, for example, the eastern tropical South Pacific (Fernandez et al. 2011) and the Southern Californian Bight (Hamersley et al. 2011). Hence, heterotrophic nitrogen fixation may constitute an overlooked component of the nitrogen cycle not only in the Baltic Sea but also in other oceans. Additional spatial and temporal studies are now warranted to deepen our knowledge on nitrogen fixation and, in particular, on the variety of microbial *nif* gene operators in Baltic Sea waters, besides the well-known photoautotrophic cyanobacteria assumed to dominate.

METAGENOMICS IN BALTIC SEA MONITORING

HTS-based methods have huge capacity to provide detailed and all-encompassing information on microbial identity and potential function. In turn, this creates great potential for sequencing-based monitoring programs. Major advantages include the improved accessibility (i) to a widened coverage of uncultured organisms, (ii) to functional genes targeting specific metabolic processes of relevance for understanding ecosystem processes, (iii) to small bacteria, eukaryotic phytoplankton, and viruses, i.e., microbes lacking distinct morphologies, and finally, (iv) to a more consistent taxonomic identification. Efficient monitoring systems are characterized by continuous sampling and rapid handling, processing, and analyses of samples. Today, available HTS techniques enable simultaneous analyses of thousands of microbial samples with sufficient sequencing depths to reliably capture taxonomic diversity (Caporaso et al. 2011).

Initiatives to introduce genomic/metagenomic analyses in global monitoring programs and observatories are well underway, although not yet a standard (Bourlat et al. 2013; Davies et al. 2014). In an overview of the use of genomic tools in monitoring programs, Bourlat et al. (2013) identified thirteen indicators for qualitative descriptors from the “Marine Strategy Framework Directive” (MSFD, 2008/56/EC), for which genomic tools can be implemented (Fig. 1). The descriptor categories include biological diversity, non-indigenous species, food webs, human-induced eutrophication, and seafloor integrity, all of which are of great relevance in assessing the environmental status of the Baltic Sea. More specific examples of potential targets in a Baltic Sea metagenomics-based monitoring program

include genes and organisms of importance for eutrophication and nutrient cycles; with processes in focus encompassing, e.g., photosynthesis, nitrification/denitrification, nitrogen fixation, and phosphate uptake/metabolism. Other sets of target organisms/genes may be related to pollution (e.g. biodegradation of organic pollutants), or to public health concern and “early-warning systems,” including toxin producing microorganisms/toxin genes, as well as pathogens (e.g., *Vibrio*) and pathogenicity genes. Yet another example of microbe/gene sets is those involved in vitamin production in view of the ongoing thiamine (B1) deficiency in higher Baltic Sea organisms (see e.g., Balk et al. 2009). It should, however, be pointed out that while attempts have been made in defining sets of genes that can be used as indicators of environmental perturbations (see Yergeau et al. 2007; Bengtsson-Palme et al. 2014), the field still requires intensive research. Understanding of the relevant reporter genes is necessary for efficient use of metagenomics for monitoring purposes.

Phytoplankton is since long included in environmental monitoring programs covering the Baltic Sea water body, and methods for sampling and identification were standardized in 1991 through the establishment of the HELCOM phytoplankton expert group. This program, like most others, relies on a morphology-based identification of phytoplankton (light microscopy). However, these practices require considerable work efforts and cover a limited number of samples. Many phytoplankton also lack obvious morphological characteristics needed for identification. The advancement of HTS technologies therefore appears as a promising alternative (or complement) to morphology-based identification. A comparative analysis of a genetic (16S rRNA gene) and a morphology-based identification of phytoplankton revealed considerable discrepancies. For instance, Euglenophyta and Heterokonta were less frequently identified by the sequence-based approach while cyanobacteria were more frequently identified (Eiler et al. 2013). A similar comparison between metagenomic identifications (Dupont et al. 2014) and conventional monitoring data publicly available (SMHI) shows similar discrepancies (Fig. 2). Even though the datasets are not directly comparable (e.g., not based on the same samples), the difference in cyanobacterial, diatom, and green algae identification and abundances are worth noting. These discrepancies are likely indicative of inherent methodological issues for both methods. For the morphology-based identification, these include, for example, cell preservation and human biased microscopic identification, while for the sequence-based method, they relate to DNA/RNA nucleic acid extraction, primer biases, library preparation sequencing depth, and the assembly of short DNA sequences (Gomez-Alvarez et al. 2009; Niu et al. 2010; Schmieder and Edwards 2011). However, the

Metagenomics in monitoring

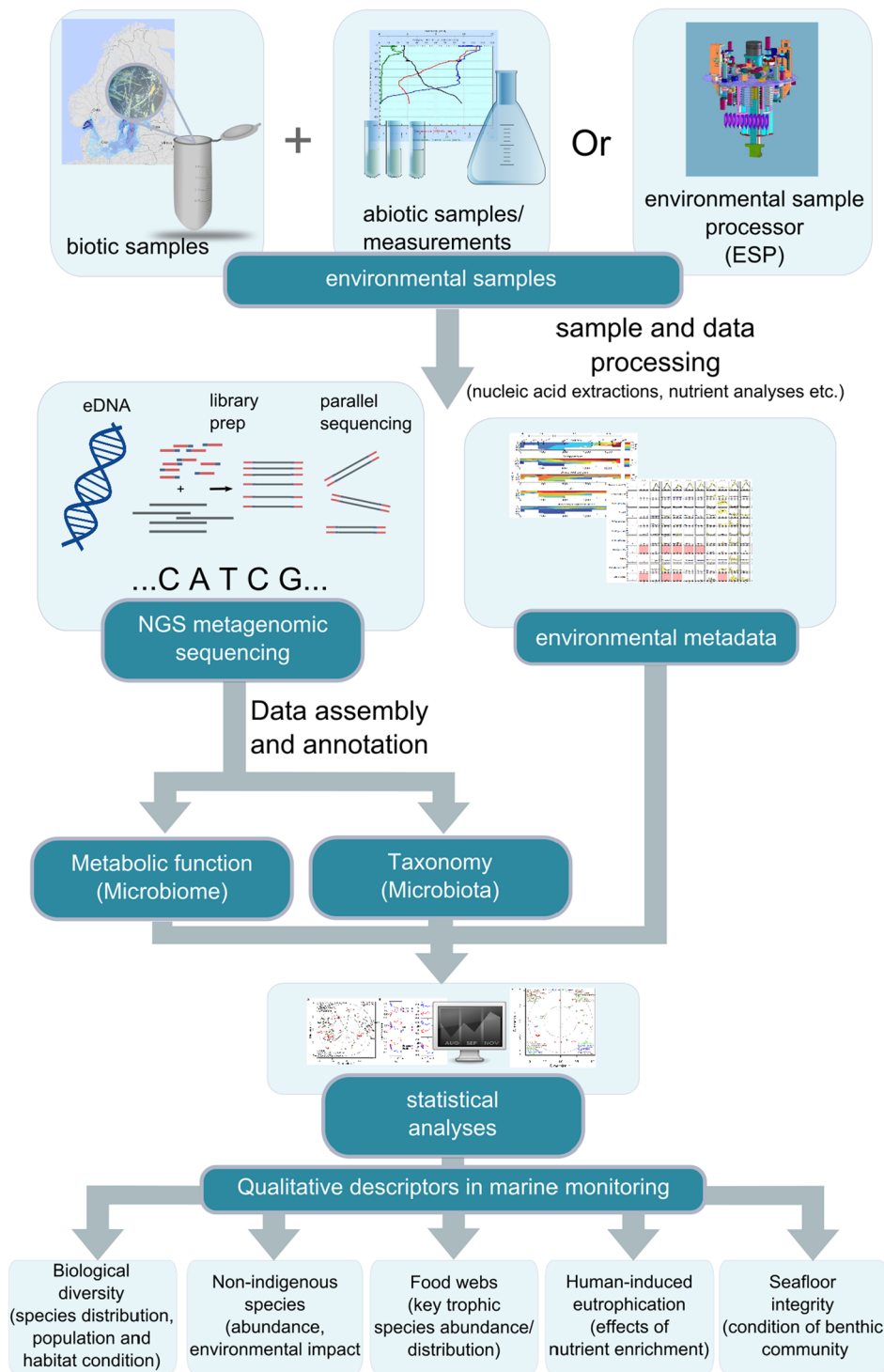


Fig. 1 Metagenomics in identifying and monitoring of microbes in the Baltic Sea. A schematic flowchart of the metagenomic approach used in the MiMeBS program and its potential integration in monitoring programs for the Baltic Sea. Criteria for which genomic methods can be used to assess environmental status were derived from the “Marine Strategy Framework Directive” (Bourlat et al. 2013)

perhaps largest challenge for a sequence-based monitoring program is the limited availability of sequenced reference strains. Metagenomic sequencing projects generate vast

amounts of data, but more than half of the reads may end up as “unclassified” due to lack of such reference (genome/genetic) material. One recently introduced way of

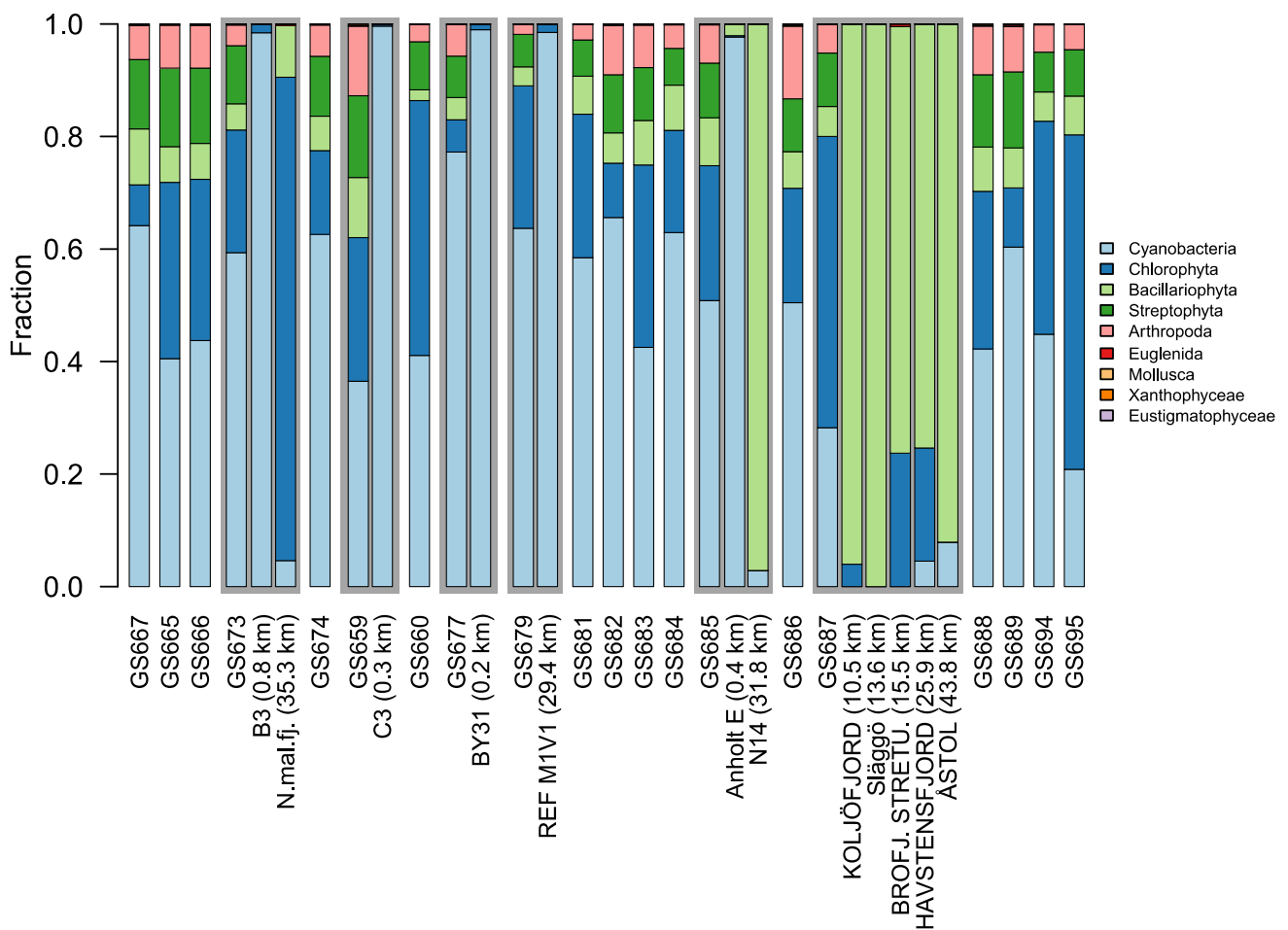


Fig. 2 Comparison of phytoplankton and metazoan classifications in environmental samples via genetic (metagenomic) and microscopy-based methods. Samples starting with “GS” represent metagenomic sequencing where classifications were made using similarity searches of protein-coding genes against reference databases. The remaining samples represent microscopy-based classifications available in the SMHI database of environmental parameters (www.smhi.se). SMHI-sites closest to and within 50 km of the “GS” sampling locations were identified and are shown within the same shaded area as their nearest metagenomic samples with distances shown in *parentheses*. Abbreviated SMHI-sites are as follows: *N.mal.fj* Nordmalingsfjärden, 1, *BY31* BY31 LANDSORTSDJ, *BROFJ. STRETU.* BROFJORDEN/STRETTUDDEN

resolving these issues is the above-discussed assembly or “binning” of genomes using metagenomic sequences (Herlemann et al. 2013; Alneberg et al. 2014; Nielsen et al. 2014). Of specific interest is also genome-sequencing projects focusing on maximizing phylogenetic coverage (Wu et al. 2009; Shih et al. 2013) and accessing the “rare microbial biosphere” (Dini-Andreote et al. 2012). For eukaryotes with larger genomes, DNA barcoding and metabarcoding are still likely more realistic options for monitoring purposes than both genome sequencing and metagenomics. In the barcoding approach, signature DNA sequences are collected from type-organisms that are either cultured or documented, e.g., by micrographs (Pawlowski et al. 2012). Barcoding initiatives furthermore ensure that sequence-based monitoring data can be harmonized with and constitute a direct continuation of, long-term morphology-based data already at hand.

Novel sequencing systems are currently being developed and evaluated for monitoring purposes in various marine ecosystems, and combined with innovative automated sampling devices such as real-time water monitoring buoys and environmental sample processors (Preston et al. 2009; Ottesen et al. 2011), these are promising tools for both expanding our knowledge and the protection of microbial life. These devices will be particularly important considering ongoing global warming, likely to negatively affect services delivered by natural aquatic ecosystems (Worm et al. 2006; White et al. 2012). Climate change predictions indicate that life in the Baltic Sea will be exposed to more drastic negative effects than those expected in global oceans (Meier et al. 2012), stressing the need for improved sustainable management practices for this body of waters. However, even in a global perspective, research efforts have only recently started to target consequences of such changes for

microbial life and hence on the global biogeochemical cycling of nutrients. Expanded genomics-based monitoring programs are therefore as urgent prerequisites for the Baltic Sea as for oceans and other water bodies.

CONCLUSION

The ability of an organism to survive through long-term and/or rapid changes in the environment is determined by its genetic repertoire and capacity to adapt physiologically. While metazoans are restricted by long generation times, genetic adaptations in bacteria can be fast and substantial. The steep physicochemical gradients and geographic isolation in the Baltic Sea pose challenging prerequisites for most organisms; however, fast growth and small microbial genomes enable drastic genetic modifications as a response to this variable environment. In addition to potentially profound effects of the microbial biodiversity on productivity and nutrient retention, the large pool of genetic diversity in Baltic Sea microbes discovered through HTS projects provides a valuable resource for resilience. In spite of the documented microbial diversity and functional potential in the Baltic Sea metagenomic datasets generated, many microbial-driven biochemical processes, ecosystem interactions and environmental adaptations remain insufficiently investigated. The studies reviewed here, based on microbial metagenomics, create a necessary base-line for monitoring the effects of a changing environment. Still, in order to develop efficient tools for monitoring programs, additional vertical, horizontal, and seasonal sampling and analyses of microbes in the Baltic Sea are required. Information obtained can then be used to create cost-effective screenings of, e.g., key-players in biogeochemical cycles, pathogenic organisms, and important components of Baltic food webs. This will in turn ensure a scientifically sound, knowledge-based management of the Baltic Sea, and its organismal resources in the future.

Acknowledgments The financial support from the Baltic Sea 2020 and Olle Engkvist, Byggmästare Foundations for the “Microbial Metagenomics of the Baltic Sea” program (MiMeBS), is acknowledged, as is the financial support from the Baltic Ecosystem Adaptive Management (BEAM) program at Stockholm University. The MiMeBS program is a collaborative effort between J. Craig Venter Institute’s Global Ocean Sampling expedition and Stockholm University/Science for Life Lab, Stockholm. We are grateful to Dr. R. Foster for valuable comments on the manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

REFERENCES

- Alneberg, J., B.S. Bjarnason, I. de Bruijn, M. Schirmer, J. Quick, U.Z. Ijaz, L. Lahti, N.J. Loman, et al. 2014. Binning metagenomic contigs by coverage and composition. *Nature Methods* 11: 1144–1146.
- Andersson, A.F., L. Riemann, and S. Bertilsson. 2010. Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *The ISME Journal* 4: 171–181.
- Balk, L., P.A. Hägerroth, G. Åkerman, M. Hanson, U. Tjärnlund, T. Hansson, G.T. Hallgrímsson, Y. Zebuhr, et al. 2009. Wild birds of declining European species are dying from a thiamine deficiency syndrome. *Proceedings of the National Academy of Sciences of the United States of America* 106: 12001–12006.
- Bengtsson-Palme, J., M. Alm Rosenblad, M. Molin, and A. Blomberg. 2014. Metagenomics reveals that detoxification systems are underrepresented in marine bacterial communities. *BMC Genomics* 15: 749. doi:10.1186/1471-2164-15-749.
- Bourlat, S.J., A. Borja, J. Gilbert, M.I. Taylor, N. Davies, S.B. Weisberg, J.F. Griffith, T. Lettieri, et al. 2013. Genomics in marine monitoring: New opportunities for assessing marine health status. *Marine Pollution Bulletin* 74: 19–31.
- Campbell, B.J., and D.L. Kirchman. 2013. Bacterial diversity, community structure and potential growth rates along an estuarine salinity gradient. *The ISME Journal* 7: 210–220.
- Caporaso, J.G., C.L. Lauber, W.A. Walters, D. Berg-Lyons, C.A. Lozupone, P.J. Turnbaugh, N. Fierer, and R. Knight. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America* 108: 4516–4522.
- Carstensen, J., D.J. Conley, E. Bonsdorff, B.G. Gustafsson, S. Hietanen, U. Janas, T. Jilbert, A. Maximov, et al. 2014. Hypoxia in the Baltic Sea: Biogeochemical cycles, benthic fauna, and management. *AMBIO* 43: 26–36.
- Davies, N., D. Field, L. Amaral-Zettler, K. Barker, M. Bica, S. Bourlat, J. Coddington, J. Deck, et al. 2014. Report of the 14(th) genomic standards consortium meeting, Oxford, UK, September 17–21, 2012. *Standards in Genomic Sciences*, 9: 1236–1250.
- DeLong, E.F., C.M. Preston, T. Mincer, V. Rich, S.J. Hallam, N.U. Frigaard, A. Martinez, M.B. Sullivan, et al. 2006. Community genomics among stratified microbial assemblages in the ocean’s interior. *Science* 311: 496–503.
- Dinasquet, J., T. Kragh, M.L. Schroter, M. Søndergaard, and L. Riemann. 2013. Functional and compositional succession of bacterioplankton in response to a gradient in bioavailable dissolved organic carbon. *Environmental Microbiology* 15: 2616–2628.
- Dini-Andreote, F., F.D. Andreote, W.L. Araujo, J.T. Trevors, and J.D. van Elsas. 2012. Bacterial genomes: Habitat specificity and uncharted organisms. *Microbial Ecology* 64: 1–7.
- Dos Santos, P., Z. Fang, S. Mason, J. Setubal, and R. Dixon. 2012. Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics* 13: 162.
- Dupont, C.L., J. Larsson, S. Yooseph, K. Ininbergs, J. Goll, J. Asplund-Samuelsson, J.P. McCrow, N. Celepli, et al. 2014. Functional tradeoffs underpin salinity-driven divergence in microbial community composition. *PLoS One* 9: e89549.
- Eiler, A., S. Drakare, S. Bertilsson, J. Pernthaler, S. Peura, C. Rofner, K. Simek, Y. Yang, et al. 2013. Unveiling distribution patterns of freshwater phytoplankton by a next generation sequencing based approach. *PLoS One* 8: e53516.
- Erguder, T.H., N. Boon, L. Wittebolle, M. Marzorati, and W. Verstraete. 2009. Environmental factors shaping the ecological

- niches of ammonia-oxidizing archaea. *FEMS Microbiology Reviews* 33: 855–869.
- Farnelid, H., M. Bentzon-Tilia, A.F. Andersson, S. Bertilsson, G. Jost, M. Labrenz, K. Jürgens, and L. Riemann. 2013. Active nitrogen-fixing heterotrophic bacteria at and below the chemocline of the central Baltic Sea. *The ISME Journal* 7: 1413–1423.
- Feike, J., K. Jürgens, J.T. Hollibaugh, S. Krüger, G. Jost, and M. Labrenz. 2012. Measuring unbiased metatranscriptomics in suboxic waters of the central Baltic Sea using a new in situ fixation system. *The ISME Journal* 6: 461–470.
- Fernandez, C., L. Farias, and O. Ulloa. 2011. Nitrogen fixation in denitrified marine waters. *PLoS One* 6: e20539.
- Fierer, N., M. Breitbart, J. Nulton, P. Salamon, C. Lozupone, R. Jones, M. Robeson, R.A. Edwards, et al. 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Applied and Environmental Microbiology* 73: 7059–7066.
- Fonselius, S., and J. Valderrama. 2003. One hundred years of hydrographic measurements in the Baltic Sea. *Journal of Sea Research* 49: 229–241.
- Frias-Lopez, J., Y. Shi, G.W. Tyson, M.L. Coleman, S.C. Schuster, S.W. Chisholm, and E.F. Delong. 2008. Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences of the United States of America* 105: 3805–3810. doi:10.1073/pnas.0708897105.
- Gast, V., and K. Gocke. 1988. Vertical distribution of number, biomass and size-class spectrum of bacteria in relation to oxic anoxic conditions in the central Baltic Sea. *Marine Ecology Progress Series* 45: 179–186.
- Gilbert, J.A., D. Field, Y. Huang, R. Edwards, W. Li, P. Gilna, and I. Joint. 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PLoS One* 3: e3042.
- Giovannoni, S.J., J.C. Thrash, and B. Temperton. 2014. Implications of streamlining theory for microbial ecology. *The ISME Journal* 8: 1553–1565.
- Gomez-Alvarez, V., T.K. Teal, and T.M. Schmidt. 2009. Systematic artifacts in metagenomes from complex microbial communities. *The ISME Journal* 3: 1314–1317.
- Gruber, N. 2005. Oceanography: A bigger nitrogen fix. *Nature* 436: 786–787.
- Hagström, Å., and U. Larsson. 1984. Diel and seasonal variation in growth rates of pelagic bacteria. In *Heterotrophic activity in the sea: NATO conference series*, ed. J. Hobbie, and P.I. Williams, 249–262. New York: Springer.
- Hajdu, S., H. Höglander, and U. Larsson. 2007. Phytoplankton vertical distributions and composition in Baltic Sea cyanobacterial blooms. *Harmful Algae* 6: 189–205.
- Hammersley, M.R., K.A. Turk, A. Leinweber, N. Gruber, J.P. Zehr, T. Gunderson, and D.G. Capone. 2011. Nitrogen fixation within the water column associated with two hypoxic basins in the Southern California Bight. *Aquatic Microbial Ecology* 63: 193–205.
- Heinänen, A.P. 1991. Bacterial numbers, biomass and productivity in the Baltic Sea—A cruise study. *Marine Ecology Progress Series* 70: 283–290.
- Herlemann, D.P., M. Labrenz, K. Jürgens, S. Bertilsson, J.J. Waniek, and A.F. Andersson. 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *The ISME Journal* 5: 1571–1579.
- Herlemann, D.P., D. Lundin, M. Labrenz, K. Jürgens, Z. Zheng, H. Aspeborg, and A.F. Andersson. 2013. Metagenomic de novo assembly of an aquatic representative of the verrucosomic class Spartobacteria. *AMBIO* 4: e00569–e00612.
- Herlemann, D.P., M. Manecki, C. Meeske, F. Pollehne, M. Labrenz, D. Schulz-Bull, T. Dittmar, and K. Jürgens. 2014. Uncoupling of bacterial and terrigenous dissolved organic matter dynamics in decomposition experiments. *PLoS One* 9: e93945.
- Huttenhower, C., D. Gevers, R. Knight, S. Abubucker, J.H. Badger, A.T. Chinwalla, H.H. Creasy, A.M. Earl, et al. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486: 207–214.
- Hårdeman, F., and S. Sjöling. 2007. Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. *FEMS Microbiology and Ecology* 59: 524–534.
- Iverson, V., R.M. Morris, C.D. Frazar, C.T. Berthiaume, R.L. Morales, and E.V. Armbrust. 2012. Untangling genomes from metagenomes: Revealing an uncultured class of marine Euryarchaeota. *Science* 335: 587–590. doi:10.1126/science.1212665.
- Johannesson, K., and C. Andre. 2006. Life on the margin: Genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Molecular Ecology* 15: 2013–2029.
- Johnston, A.W.B., Y. Li, and L. Ogilvie. 2005. Metagenomic marine nitrogen fixation—Feast or famine? *Trends in Microbiology* 13: 416–420.
- Kahru, M., and R. Elmgren. 2014. Multidecadal time series of satellite-detected accumulations of cyanobacteria in the Baltic Sea. *Biogeosciences* 11: 3619–3633.
- Karl, D.M. 2007. Microbial oceanography: Paradigms, processes and promise. *Nature Reviews Microbiology* 5: 759–769.
- Konstantinidis, K.T., J. Braff, D.M. Karl, and E.F. DeLong. 2009. Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific subtropical gyre. *Applied and Environmental Microbiology* 75: 5345–5355.
- Koskinen, K., J. Hultman, L. Paulin, P. Auvinen, and H. Kankaanpää. 2011. Spatially differing bacterial communities in water columns of the northern Baltic Sea. *FEMS Microbiology and Ecology* 75: 99–110.
- Labrenz, M., E. Sintes, F. Toetzke, A. Zumsteg, G.J. Herndl, M. Seidler, and K. Jürgens. 2010. Relevance of a crenarchaeotal subcluster related to *Candidatus Nitrosopumilus maritimus* to ammonia oxidation in the suboxic zone of the central Baltic Sea. *The ISME Journal* 4: 1496–1508.
- Larsson, J., N. Celepli, K. Ininbergs, C.L. Dupont, S. Yooseph, B. Bergman, and M. Ekman. 2014. Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *The ISME Journal* 8: 1892–1903.
- Larsson, J., J.A. Nylander, and B. Bergman. 2011. Genome fluctuations in cyanobacteria reflect evolutionary, developmental and adaptive traits. *BMC Evolutionary Biology* 11: 187.
- Larsson, U., S. Hajdu, J. Walve, and R. Elmgren. 2001. Baltic Sea nitrogen fixation estimated from the summer increase in upper mixed layer total nitrogen. *Limnology and Oceanography* 46: 811–820.
- Lauro, F.M., D. McDougald, T. Thomas, T.J. Williams, S. Egan, S. Rice, M.Z. DeMaere, L. Ting, et al. 2009. The genomic basis of trophic strategy in marine bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 106: 15527–15533.
- Lozupone, C.A., and R. Knight. 2007. Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences of the United States of America* 104: 11436–11440.
- Meier, M.H.E., H.C. Andersson, B. Arheimer, T. Blenckner, B. Chubarenko, C. Donnelly, K. Eilola, B.G. Gustafsson, et al. 2012. Comparing reconstructed past variations and future projections of the Baltic Sea ecosystem—First results from multi-model ensemble simulations. *Environmental Research Letters* 7: 034005.

- Morris, R.M., M.S. Rappe, S.A. Connon, K.L. Vergin, W.A. Siebold, C.A. Carlson, and S.J. Giovannoni. 2002. SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* 420: 806–810.
- Nielsen, H.B., M. Almeida, A.S. Juncker, S. Rasmussen, J. Li, S. Sunagawa, D.R. Plichta, L. Gautier, A.G. Pedersen, et al. 2014. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nature Biotechnology* 32: 822–828.
- Niu, B.F., L.M. Fu, S.L. Sun, and W.Z. Li. 2010. Artificial and natural duplicates in pyrosequencing reads of metagenomic data. *BMC Bioinformatics* 11: 187.
- Ottesen, E.A., R. Marin III, C.M. Preston, C.R. Young, J.P. Ryan, C.A. Scholin, and E.F. DeLong. 2011. Metatranscriptomic analysis of autonomously collected and preserved marine bacterioplankton. *The ISME Journal* 5: 1881–1895.
- Pawlowski, J., S. Audic, S. Adl, D. Bass, L. Belbahri, et al. 2012. CBOL protist working group: Barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biology* 10: e1001419. doi:10.1371/journal.pbio.1001419.
- Pinhassi, J., and Å. Hagström. 2000. Seasonal succession in marine bacterioplankton. *Aquatic Microbial Ecology* 21: 245–256.
- Pinhassi, J., U.L. Zweifel, and Å. Hagström. 1997. Dominant marine bacterioplankton species found among colony-forming bacteria. *Applied and Environmental Microbiology* 63: 3359–3366.
- Preston, C.M., R. Marin III, S.D. Jensen, J. Feldman, J.M. Birch, E.I. Massion, E.F. DeLong, M. Suzuki, et al. 2009. Near real-time, autonomous detection of marine bacterioplankton on a coastal mooring in Monterey Bay, California, using rRNA-targeted DNA probes. *Environmental Microbiology* 11: 1168–1180.
- Rheinheimer, G., K. Gocke, and H.G. Hoppe. 1989. Vertical distribution of microbiological and hydrographic-chemical parameters in different areas of the Baltic Sea. *Marine Ecology Progress Series* 52: 55–70.
- Riemann, L., C. Leitet, T. Pommier, K. Simu, K. Holmfeldt, U. Larsson, and Å. Hagström. 2008. The native bacterioplankton community in the central Baltic Sea is influenced by freshwater bacterial species. *Applied and Environmental Microbiology* 74: 503–515.
- Savage, C., P.R. Leavitt, and R. Elmgren. 2010. Effects of land use, urbanization, and climate variability on coastal eutrophication in the Baltic Sea. *Limnology and Oceanography* 55: 1033–1046.
- Scanlan, D.J., M. Ostrowski, S. Mazard, A. Dufresne, L. Garczarek, W.R. Hess, A.F. Post, M. Hagemann, et al. 2009. Ecological genomics of marine picocyanobacteria. *Microbiology and Molecular Biology Reviews* 73: 249–299.
- Schmieder, R., and R. Edwards. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27: 863–864.
- Shih, P.M., D. Wu, A. Latifi, S.D. Axen, D.P. Fewer, E. Talla, A. Calteau, F. Cai, et al. 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 110: 1053–1058.
- Stal, L.J., P. Albertano, B. Bergman, K. Bröckel, J.R. Gallon, P.K. Hayes, K. Sivonen, and A.E. Walsby. 2003. BASIC: Baltic Sea cyanobacteria. An investigation of the structure and dynamics of water blooms of cyanobacteria in the Baltic Sea—Responses to a changing environment. *Continental Shelf Research* 23: 1695–1714.
- Thureborn, P., D. Lundin, J. Plathan, A.M. Poole, B.M. Sjöberg, and S. Sjöling. 2013. A metagenomics transect into the deepest point of the Baltic Sea reveals clear stratification of microbial functional capacities. *PLoS One* 8: e74983.
- Väätänen, P. 1980. Effects of environmental factors on microbial populations in the brackish waters off the southern coast of Finland. *Applied and Environmental Microbiology* 40: 48–54.
- Venter, J.C., K. Remington, J.F. Heidelberg, A.L. Halpern, D. Rusch, J.A. Eisen, D.Y. Wu, I. Paulsen, et al. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304: 66–74.
- White, C., B.S. Halpern, and C.V. Kappel. 2012. Ecosystem service tradeoff analysis reveals the value of marine spatial planning for multiple ocean uses. *Proceedings of the National Academy of Sciences of the United States of America* 109: 4696–4701.
- Worm, B., E.B. Barbier, N. Beaumont, J.E. Duffy, C. Folke, B.S. Halpern, J.B.C. Jackson, H.K. Lotze, et al. 2006. Impacts of biodiversity loss on ocean ecosystem services. *Science* 314: 787–790.
- Wu, D., P. Hugenholtz, K. Mavromatis, R. Pukall, E. Dalin, N.N. Ivanova, V. Kunin, L. Goodwin, et al. 2009. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* 462: 1056–1060.
- Yergeau, E., S. Kang, Z. He, J. Zhou, and G.A. Kowalchuk. 2007. Functional microarray analysis of nitrogen and carbon cycling genes across an Antarctic latitudinal transect. *The ISME Journal* 1: 163–179. doi:10.1038/ismej.2007.24.
- Yooseph, S., G. Sutton, D.B. Rusch, A.L. Halpern, S.J. Williamson, K. Remington, J.A. Eisen, K.B. Heidelberg, et al. 2007. The Sorcerer II Global Ocean Sampling expedition: Expanding the universe of protein families. *PLoS Biology* 5: 432–466.
- Yooseph, S., K.H. Nealson, D.B. Rusch, J.P. McCrow, C.L. Dupont, M. Kim, J. Johnson, R. Montgomery, et al. 2010. Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* 468: 60–66.
- Zinger, L., L.A. Amaral-Zettler, J.A. Fuhrman, M.C. Horner-Devine, S.M. Huse, D.B.M. Welch, J.B.H. Martiny, M. Sogin, et al. 2011. Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *PLoS One* 6: e24570.

AUTHOR BIOGRAPHIES

Karolina Ininbergs (✉) is a researcher and co-head of the MiMeBS program at Stockholm University/SciLifeLab, investigating microbial diversity and function in the Baltic Sea. Karolina's primary research focus is marine microbiology, with specific interest in physiology and genetic diversity of cyanobacteria.
Address: Science for Life Laboratory, Department of Ecology, Environment and Plant Sciences, Stockholm University, Box 1031, 171 21 Solna, Sweden.
 e-mail: karolina.ininbergs@su.se

Birgitta Bergman is full Professor in Plant Physiology at Stockholm University. She is a member of The Royal Swedish Academy of Sciences (since 1996). Bergman's research focus is on interactive and evolutionary aspects of symbiotic and marine free-living cyanobacteria/microbes. Professor Bergman is the founder of the MiMeBS program.
Address: Science for Life Laboratory, Department of Ecology, Environment and Plant Sciences, Stockholm University, Box 1031, 171 21 Solna, Sweden.

John Larsson headed the metagenomic analyses within the MiMeBS program. John is currently a post doc at Linnaeus University/SciLifeLab. His main expertise is primarily bioinformatic analyses including phylogenetics, statistical analyses and comparative genomics, genome assembly, and binning.

Address: Science for Life Laboratory, Department of Ecology, Environment and Plant Sciences, Stockholm University, Box 1031, 171 21 Solna, Sweden.

Address: Science for Life Laboratory, Linnaeus University Kalmar, 171 21 Solna, Sweden.

Martin Ekman is a researcher and co-head of the MiMeBS program at Stockholm University/SciLifeLab. He has a background in molecular biology and physiology of cyanobacteria, with specific focus on symbiotic systems.

Address: Science for Life Laboratory, Department of Ecology, Environment and Plant Sciences, Stockholm University, Box 1031, 171 21 Solna, Sweden.