

# Accumulation of DNA in an anoxic sediment – rDNA and rRNA presence of members of the microbial community from Landsort Deep, the Baltic Sea

Petter Thureborn<sup>1</sup>, Yue O. O. Hu<sup>2</sup>, Andrea Franzetti<sup>3</sup>, Sara Sjöling<sup>1</sup> and Daniel Lundin<sup>4\*</sup>

5 <sup>1</sup> School of Natural Sciences, Technology and Environmental Studies, Södertörn University, Huddinge, Sweden

<sup>2</sup> School of Biotechnology, Division of Gene Technology, Science for Life Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden

<sup>3</sup> Dept. of Earth and Environmental Sciences, University of Milano - Bicocca, Milano, Italy

10 <sup>4</sup> Centre for Ecology and Evolution in Microbial Model Systems, Linnaeus University, Kalmar, Sweden

\* Correspondence: [daniel.lundin@lnu.se](mailto:daniel.lundin@lnu.se)

Running title: Landsort Deep graveyard

# Abstract

Numerous investigations of bacterial communities using sequence analysis of environmental DNA have revealed extensive diversity of microbial taxa in an array of different environmental habitats. Community analysis based solely on DNA, however, does not reveal whether the detected community members are actively contributing to community functioning, or whether they are dormant or remnants of dead cells. This dilemma is of particular concern when analyzing microbial community structure of sites with a high degree of deposited matter, such as marine sediments. For example, the Baltic Sea's deepest point, the Landsort Deep, consists of anoxic sediments with a large deposition of allochthonous organic matter from the highly stratified 460 m water column above. Our previous metagenomics results indicated the presence of potential obligately aerobic and phototrophic microorganisms in the Landsort Deep sediment. To further elucidate which taxa may contribute to ecosystem function at this site, we here present three different datasets – rDNA amplicons, rDNA reads from a shotgun metagenome and expressed rRNA from a shotgun metatranscriptome. By comparing the three datasets and the ratios between rRNA and rDNA we seek to estimate the protein synthesis potential of the community members in order to provide an indication of what taxa may have cellular activity and metabolic potential. The variation in protein synthesis potential was large, both within and between taxa, in the sediment community. Many typically anaerobic taxa, e.g. from Deltaproteobacteria and Euryarchaeota, showed a high protein synthesis potential, while typical aerobes like Flavobacteria showed a low protein synthesis potential. More surprisingly, some common Baltic Sea surface water bacteria also displayed a high protein synthesis potential, suggesting they have an active role in the anoxic sediment ecosystem at 460 m depth. Both filamentous and unicellular Cyanobacteria exhibited very high protein synthesis potential, which implies a more complex role of these bacteria in carbon cycling in the Baltic Sea than previously suggested. Moreover, Mycobacteria, that were abundant in Landsort Deep sediment metagenome compared with other marine sediment metagenomes, showed protein synthesis potentials consistent with a functional role in the sediment community. Our results provide a new window of insight into the complexities of the microbial community of Landsort Deep with implications for the understanding of other anoxic accumulation sediments.

# Introduction

Microbial community analyses based on total community DNA methods have revealed important information about the microbial diversity in a wide range of habitats globally. However, since the DNA pool includes nucleotides both from living and dead cells, taxa

identified in total DNA are not necessarily active nor alive (Blazewicz et al. 2013; Corinaldesi et al. 2011; Torti, Lever, and Jørgensen 2015). A habitat that is particularly sensitive for misinterpretations using DNA-based techniques is bottom sediment of eutrophied or deep seas. Besides hosting local microbial populations, sediments receive cells and cellular material from the water column (Turley 2000; Zinger et al. 2011). Furthermore, anoxic sediments are hotspots of preserved extracellular DNA (Corinaldesi et al. 2011; Torti, Lever, and Jørgensen 2015). Consequently, analysis of DNA may misleadingly include species that are not part of the active community at the site but a result of accumulated dead cells or extracellular DNA. The sediments at Landsort Deep – the deepest point of the Baltic Sea – receive and accumulate organic material from the water column, in particular deposited as a consequence of riverine input and the eutrophication process (Conley et al. 2009). Furthermore, the sediments are anoxic, likely extending the half time of DNA from sedimenting cells. In a previous study we could show, by comparative metagenomics, that the Landsort Deep sediment holds a diverse bacterial and archaeal community (Thureborn et al. 2013). However, the presence of phototrophic Cyanobacteria and typically aerobic bacteria such as flavobacteria and *Mycobacterium* spp., raised the question: which portion of the taxa were part of the active community and which were remnants of dead cells sinking through the water column?

In this study, we sought to shed some light on this question by investigating to what extent individual populations of the anoxic Landsort Deep sediment community had a potential to actively contribute to ecosystem functioning. In the absence of an indisputable genetic marker for activity, we used expressed rRNA as a proxy for protein synthesis potential. We compared the profiles of operational taxonomic units (OTUs) in 16S rDNA amplicons (rDNA), shotgun total DNA (totalDNA) and shotgun total RNA (totalRNA). Protein synthesis potential for each OTU was estimated from the ratio between its relative abundance in totalRNA and rDNA. Moreover, we qualitatively confirmed the findings from the amplicon-dependent analysis using *ab initio* identified rRNA sequences in the totalDNA and totalRNA datasets. We report large variation in protein synthesis potential among the Landsort Deep sediment community members. Several high-level taxa and OTUs displayed high totalRNA/rDNA ratios, providing new insight into what members of the Landsort Deep sediment community have a potential to directly contribute to ecosystem functioning. Low protein synthesis potential was detected for some taxa, consistent with an expected inflow of dead cells from the water column to the sediment.

## Results

### *Sequencing output*

The rDNA dataset produced 23,888 merged and quality-controlled amplicon read pairs that were clustered into 447 OTUs at a 99% similarity threshold. All OTUs could be matched to a taxon in the SILVA database (Quast et al. 2013) (data not shown). After deletion of chloroplast-annotated OTUs, 22,350 rDNA reads from 424 OTUs remained. Reads from the totalRNA and totalDNA datasets were mapped to the seed sequences of the rDNA OTUs, to quantify OTUs in these datasets. Around one million totalRNA reads matched the rDNA OTUs, with all 424 OTUs represented. Only two thousand totalDNA reads matched the OTU seed sequences, matching less than half of the OTUs (**Table 1**). The rDNA sample had the by far highest Shannon diversity at 4.91 (**Table 1**).

### *Relative abundance of taxa in rDNA, totalRNA and totalDNA*

Comparison of the relative abundance of high-level taxa – phyla and proteobacterial classes – in the rDNA, totalDNA and totalRNA datasets revealed both differences and similarities between the datasets (**Figure 1**). Most strikingly, the relative abundance of Cyanobacteria in totalRNA was ten-fold higher compared to rDNA. Conversely, Bacteroidetes and Firmicutes were overrepresented in rDNA compared to totalRNA. The remaining large taxa – Deltaproteobacteria and Actinobacteria – had similar relative abundance in the three datasets (**Figure 1**). The totalDNA dataset displayed some differences from the rDNA. In particular there was a higher frequency of Actinobacteria and lower frequencies of less abundant taxa, for instance Betaproteobacteria. The latter is an expected consequence of the five-fold smaller size of the totalDNA dataset and the incomplete coverage of OTUs (**Table 1**), and the totalDNA was consequently excluded from the amplicon-based analysis and only used in an amplicon-independent analysis (see subsection below, “Amplicon-independent relative abundances of taxa in totalRNA and totalDNA”).

### *Per phylum distribution of protein synthesis potential*

We calculated ratios between the relative abundances of each OTU (n=424) in totalRNA and rDNA (totalRNA/rDNA ratios). The ratios had a wide distribution: minimum 0.0, maximum 207, mean 3.3 and median 0.47. Breaking down the totalRNA/rDNA ratios in terms of phyla and proteobacterial classes, large variations of totalRNA/rDNA ratios among taxa were observed (**Figure 2**). Furthermore, we applied a sign test ( $p < 0.05$ ) to identify high-level taxa – phyla and proteobacterial classes – with ratio distributions significantly higher or lower than the community median (0.47) (**Figure 2**). Actinobacteria, Cyanobacteria and Deltaproteobacteria had totalRNA/rDNA ratio distributions among their OTUs that were

significantly higher than the community median. Taxonomically unclassified OTUs (“Unclassified”), bacteria not classified at phylum level (“Bacteria - other”), Bacteroidetes and Betaproteobacteria had significantly lower ratio distributions than the community median (**Figure 2**).

Most phyla and proteobacterial classes display wide distributions of totalRNA/rDNA ratios (**Figure 2**). To investigate how the distribution in protein synthesis potential relate to abundance of OTUs in the rDNA dataset within phyla and proteobacterial classes, we plotted relative abundances in the rDNA dataset to percent intervals of totalRNA/rDNA ratios. Most taxa displayed the expected trend of a rather strong decline of rDNA abundance with increasing totalRNA to rDNA ratios (**Figure 3**). Potential exceptions were the Delta- and Gammaproteobacteria. Moreover, the two largest orders of Deltaproteobacteria – Desulfarculales and Desulfobacterales – had high abundances mostly at higher totalRNA/rDNA ratios than the community median (**Figure 3**). Cyanobacteria stood out as different from all other phyla by being represented almost exclusively at high totalRNA/rDNA ratios at relatively high abundances (**Figure 3**). This was true both for the small unicellular *Synechococcus*-related OTUs in subsection I and the filamentous *Anabaena*-related OTUs in subsection IV.

The majority of the most abundant families (>1% in either the rDNA or the totalDNA dataset) had ratios higher than the community median (**Figure 4**). The clearest exceptions was Flavobacteriaceae (Bacteroidetes) and, although rarer, Christensenellaceae (Firmicutes). Mycobacteriaceae were among the high ratio and abundant families together with several anaerobic families from the Chloroflexi and Deltaproteobacteria. Three of the Firmicutes families, all from the Clostridiales order, were found close to the community median. The cyanobacterial families, both the unicellular in Subsection I and the multicellular in Subsection IV had high median ratios, with the former at high rDNA abundance and the latter at low (**Figure 4**).

#### *Amplicon-independent relative abundances of taxa in totalRNA and totalDNA*

In light of potential biases with amplicon-based analyses, we undertook an amplicon-independent classification of the totalDNA and totalRNA datasets. After identifying potential 16S reads, the sequences were compared with NCBI’s NT database and imported into MEGAN (Huson et al. 2007). 8,833 reads from the totalDNA dataset and 18,636,493 in the totalRNA dataset were annotated to Bacteria or Archaea. In the absence of a standardized classification of sequences similar to the OTU in the amplicon-based analysis, we did not

attempt to calculate precise ratios between the two datasets. We thus only present relative abundance data at the phylum/proteobacterial class level plus data for a few orders that were sufficiently well covered in the totalDNA dataset.

Most strikingly, archaeal taxa were much more abundant in the amplicon-independent analysis, due to the poor coverage of Archaea with the 16S primers used. Archaea were similarly abundant in the totalDNA and totalRNA datasets (**Figure 5A**). For bacterial taxa, at the level of phyla and proteobacterial classes, patterns were similar to the amplicon-based analysis except for Chloroflexi and Betaproteobacteria which diverged from their pattern in the amplicon-based analysis. Betaproteobacteria was more evenly abundant in totalDNA and totalRNA, while Chloroflexi, which had a median ratio very close to the community, had much higher relative abundance in totalDNA than in totalRNA (**Figures 3 and 5A**). High abundances and relatively equal frequencies in the totalDNA and totalRNA datasets were observed for Actinobacteria, Deltaproteobacteria and Euryarchaeota. Cyanobacteria had very high relative abundance in the totalRNA dataset, but, similarly to the amplicon-based analysis, very low abundance in the totalDNA dataset. The ratio between the two was approximately 7:1. Low ratios were found in Lentisphaerae (1:8), Bacteroidetes (1:7), Chloroflexi (1:5) and Verrucomicrobia (1:4) (**Figure 5A**). Only a few orders could be detected in the totalDNA amplicon-independent analysis, allowing only a glimpse of information regarding the distribution of ratios of totalRNA/totalDNA in orders (**Figure 5B**). Similarly to the amplicon-based analysis, the majority of phyla and proteobacterial classes contained orders both with high ratios of totalRNA/totalDNA and with low ratios. An exception was Bacteroidetes, which only contained orders with low ratios.

## Discussion

Anoxic sediments of deep seas are recognized as ecosystems harboring diverse microbial communities (Thureborn et al. 2013; Zinger et al. 2011). However, due to anoxic conditions and sedimentation of cells (Corinaldesi et al. 2011; Torti, Lever, and Jørgensen 2015; Turley 2000), diversity estimates from DNA recovered from these ecosystems are likely a mixture of both microbial sediment and water column populations. To provide insight into which microbial community members of the anoxic Landsort Deep sediment have a potential to contribute to ecosystem functioning, we compared frequencies in expressed rRNA (totalRNA dataset) and genomic rDNA (rDNA and totalDNA datasets) as an estimate of protein synthesis potential. We observed large groups of taxa – both by counting amplicon OTUs (**Figure 2**) and in the amplicon-independent analysis (**Figure 5**) – in the DNA datasets that were not matched by similarly high abundances in the RNA dataset. We could thus confirm

that analysis of bacterial communities of the anoxic sediment at Landsort Deep based only on genomic DNA markers will be influenced by what is likely DNA of dormant, dying and dead cells accumulating onto the sediment. Communities will therefore appear more diverse than they really are, as exemplified by the Shannon diversity of the rDNA sample being considerably higher (4.91) than the other two (4.58 and 4.40 respectively, **Table 1**). This conclusion likely extends not only to other anoxic and low temperature sediments (Edlund et al. 2008) but to accumulation environments in general, although the problem will likely be smaller at oxic conditions due to the shorter halftime of DNA under such conditions.

Furthermore, the potential problems arising from accumulation of DNA will be reinforced by the frequently high abundances of taxa with low RNA to DNA ratios. Most high-level taxa displayed significantly declining relative abundances with increasing RNA to DNA ratios in a linear log-log relationship (**Table 2**). In a standard amplicon-based community analysis, some of these taxa would be considered community dominants.

Bacteroidetes had a distribution of totalRNA/rDNA ratios that was significantly below the community median at 0.47 and Betaproteobacteria as well as the “other Bacteria”, i.e. Bacteria from smaller phyla, had a distribution that was highly skewed to the lower quartiles, indicating that sequence reads from these taxa are mostly an effect of “marine snow”, dead cells sedimenting through the water column (**Figure 2**). The pattern from the amplicon-based analysis corroborates the amplicon-independent analysis for Bacteroidetes (**Figure 5**). Bacteroidetes consisted to a large proportion of the Flavobacteriales order, common heterotrophs in Baltic Sea surface water (Herlemann et al. 2011; Thureborn et al. 2013), consistent with the accumulation of dead cells onto the sediment.

We identified several taxa, e.g. Deltaproteobacteria, Euryarchaeota, Thaumarchaeota and Spirochaetes with high protein synthesis potential, although not significantly higher than the community median, (**Figure 2 and 5**), indicating that populations from these taxa play active roles in the anoxic Landsort Deep sediment community. These taxa are known to have a capacity for anaerobic metabolism (Hoover et al. 2003; Hug et al. 2013; Liu and Whitman 2008; Muyzer and Stams 2008). Unexpectedly, actinobacterial OTUs also had a distribution of totalRNA/rDNA ratios skewed towards the higher quartiles of the totalRNA to rDNA ratios (**Figures 2 and 3**). Moreover, they were overall slightly more represented in the totalRNA dataset than the totalDNA dataset (**Figure 5**). Especially OTUs from the Mycobacteriaceae family of the Corynebacteriales order and the PeM15 had high protein synthesis potential at relatively high rDNA abundances (**Figure 3**). Representatives of the PeM15 group have been identified in various habitats including surface water of the Baltic



Sea (Lindh et al. 2015) and Chesapeake Bay estuary (Kan et al. 2008) as well as anaerobic habitats, such as sediments of Lake Kinneret (Bar-Or et al. 2015) and the midgut of humus-feeding larva (Egert et al. 2003), which suggests that PeM15 members can thrive in both aerobic and anaerobic environments. We previously showed that Mycobacteria was abundant and a characteristic feature of the Landsort Deep sediment metagenome in comparison to other habitats globally (Thureborn et al. 2013). In addition, growth of strains of Mycobacteria have been shown to be stimulated by fulvic and humic acids (Kirschner, Parker, and Falkinham 1999), which are both common constituents of Baltic Sea sediments. Our results suggest that species of PeM15 and Mycobacteria indeed play active roles in the Landsort Deep sediment, possibly in anaerobic degradation of organic compounds.

The most striking observation in our data was the very high protein synthesis potential of another characteristic Baltic Sea surface water bacteria, Cyanobacteria (**Figures 2 and 3**). The two major taxa observed – unicellular *Synechococcus*-like (Subsection I) and filamentous *Anabaena*-like (Subsection IV) – displayed the same pattern, while a few OTUs with unknown taxonomy below phylum had lower protein synthesis potential. Filamentous Cyanobacteria, e.g. *Anabaena* spp., form recurrent blooms in surface water of the Baltic Sea during July and August and can thus be expected to constitute a substantial part of the “marine snow” of sinking cells through the water column (Sivonen et al. 2007). Unicellular Cyanobacteria, e.g. *Synechococcus*, are also commonly found in surface water (Andersson, Riemann, and Bertilsson 2009) but do not form the conspicuous blooms that *Anabaena* spp. do. However, the comparatively high totalRNA/rDNA ratios and hence protein synthesis potentials of OTUs within Cyanobacteria identified in this study were unexpected, assuming that the presence of cyanobacterial nucleic acids in the sediment was merely a result of marine snow (**Figure 2 and 3**). Akinetes (dormancy cells) in filamentous Cyanobacteria increase both the number of ribosomes and genome copies, compared to vegetative cells (Sukenik et al. 2012). This could potentially explain high protein synthesis potential as a result of dormancy. Possibly it signals a physiological preparedness to access sediment phosphorous and subsequent recruitment to the water column, in case the cells had sedimented at shallow depths. Similar behavior has been observed for the colony forming cyanobacterium *Gloeotrichia echinulata* in shallow lakes as a response to phosphorus limitation in the surface water (see Cottingham et al. 2015, and references therein). Furthermore, benthic akinetes have been suggested to have an important role in overwintering and bloom dynamics of Baltic Sea *Anabaena* spp. (Olli, Kangro, and Kabel 2005; Suikkanen et al. 2010). However, the potential of vertical migration through buoyancy



250 regulation in oceanic Cyanobacteria has been estimated to be possible up to depths of approximately 70 m (Villareal and Carpenter 2003). Vertical migration of *Anabaena* spp. from Landsort Deep sediment at 460 m thus appear less likely. Regarding *Synechococcus*, cellular contents of RNA and DNA have been shown to be positively correlated to growth rate, in both freshwater and marine *Synechococcus* spp., and the RNA/DNA ratio was

255 positively correlated to growth rate in marine *Synechococcus* sp. strain WH 8103 (Lepp and Schmidt 1998). By using diurnal batch cultures Lepp and Schmidt also showed that nucleic acids (i.e. DNA, RNA, 16S rRNA) increased during periods of darkness and decreased during periods of light in *Synechococcus* sp. strain PCC 6301. This was suggested to be a reflection of a period of de novo synthesis without concurrent cell division (Lepp and

260 Schmidt 1998). The high protein synthesis potential of *Synechococcus*-like OTUs we observed, could hence be a reflection of growing cells or of cells arrested in a "dark" mode with no cell division in the dark Landsort Deep sediment. Furthermore, the comparatively high ratios of ribosomal RNA to DNA in Cyanobacteria may partly be explained by the physiology of Cyanobacteria, with large amounts of protein bound in photosystems in

265 thylakoid membranes. These proteins may need frequent replacement due to photo-induced damage, and large numbers of membrane-bound ribosomes would thus be required to replenish photosystems. Given that the protein synthesis potential of Cyanobacteria were an order of magnitude higher even than taxa that are expected to be active community members, such as Deltaproteobacteria and Euryarchaeota, it seems unlikely that these cells are dead.

270 Whether Cyanobacteria in the Landsort Deep sediment are active, growing or dormant remains an open question, but either scenario would challenge our view of sedimenting cyanobacterial blooms as solely a nutrient input for heterotrophic growth in the sediment.

Comparing the two types of analyses presented here – amplicon-based and amplicon-independent, respectively – there are clear strengths and drawbacks with both. The primers

275 that were used were designed for Bacteria (Hugerth, Wefer, et al. 2014) and Archaea were hence all but absent from the amplicon-based analysis. Less expectedly, Spirochaeta were also clearly underrepresented in the rDNA dataset and thus went virtually undetected in the amplicon-based analysis (**Figures 2 and 5**). Shotgun sequencing, as in the amplicon-independent analysis, can be expected to result in a more even coverage of taxa, although

280 biases due to differences in performance in DNA isolation and sequencing library preparation can not be excluded. While sequencing bias may have influenced our results in the amplicon-independent analysis – e.g. the low abundance of Betaproteobacteria in the totalDNA dataset – it is by its nature difficult to study. A clearer influence on the amplicon-independent

analysis, was the fact that shotgun sequencing is more difficult to standardize since it contains  
 285 reads from different parts of the genome or even a gene, , the 16S rRNA gene in this study,  
 resulting in differences in how much taxonomical information can be garnered from each  
 read. Clustering reads into OTUs as in the amplicon-based analysis, will hence likely be  
 misleading. By comparing and applying the two approaches, we address these limitations by  
 navigating into a middle ground. While several taxa remain difficult to categorize as to  
 290 whether they are active and directly contribute to the microbial ecosystem in the Landsort  
 Deep, we identify other taxa that most likely are present in the sediment as a result of  
 accumulation of dead cells and cell debris. We also identify taxa that are most likely  
 dominating the metabolic processes in the sediment, in particular the Desulfobacterales and  
 Desulfarculales orders of the Deltaproteobacteria (**Figure 3**) and several orders from the  
 295 Euryarchaeota. Somewhat surprisingly, the Mycobacteriaceae family and PeM15 of the  
 Actinobacteria also appeared according to their protein synthesis potential to play active roles  
 in the community.

Most surprising to us however, was the order of magnitude higher ratios of expressed rRNA  
 compared with both rDNA datasets exhibited by the Cyanobacteria. Although the  
 300 interpretation is not straightforward and cyanobacterial metabolic activity and physiology  
 require further investigation, results raise the question whether Cyanobacteria contribute  
 more to the ecosystem in the dark, anoxic sediment of the Landsort Deep than reduced carbon  
 compounds and nutrients.

## Material and Methods

### *Sampling*

305 Triplicate sediment cores for nucleic acid extraction were retrieved at 466 m depth, using a  
 Gemini sampler, on the 21<sup>st</sup> of April 2010 at Landsort Deep (lat 583591 N, long 01814.26 E)  
 in the Baltic Sea, Sweden (Thureborn et al. 2016).

### *Nucleic acid extraction, cDNA synthesis and PCR amplification*

310 Nucleic acids were extracted from the 0-10 cm top-layer of the triplicate sediment cores  
 (Thureborn et al. 2016). Total RNA extraction, using the FastRNA Pro Soil-Direct Kit (MP  
 Biomedicals, Solon, USA), was immediately initiated on-board. 500 mg sediment aliquots  
 were immersed in RNAPro Soil Lysis Solution and processed in the FastPrep instrument  
 followed by transportation on dry ice to the lab and storage at -80°C. The extraction  
 315 procedure continued once back in the lab following the manufacturer's instructions. Total  
 RNA extracts were treated with Turbo DNA-free kit (Life Technologies, New York, USA)

and DNaseI (Thermo Fisher Scientific, Waltham, USA) according to the manufacturer's protocols followed by quantification and RNA integrity check using a Qubit fluorometer (Life Technologies) and Bioanalyzer (Agilent Technologies, Santa Clara, USA), respectively.

Prior to synthesis of cDNA the triplicate total RNA extracts were pooled and amplified using the MessageAmp II Bacteria kit (Life Technologies) following the manufacturer's instructions, except substituting the kit oligo (dT) with T7-BpmI-(dT)<sup>16</sup> VN (Frias-Lopez et al. 2008). The antisense RNA product was synthesized to cDNA using the SuperScript II Reverse Transcriptase (Life Technologies) with random hexamer priming for first-strand synthesis and Superscript Double Stranded cDNA synthesis kit (Life Technologies) for second-strand synthesis following the manufacturer's instructions. The synthesized cDNA was purified using the Qiaquick PCR Purification Kit (Qiagen, Hilden, Germany) and subsequently treated with 2-3 units *BpmI* (New England Biolabs, Ipswich, USA) per µg cDNA for 2-3 hour at 37°C to remove poly-A tails. The *BpmI* treated cDNA was purified using Qiaquick PCR Purification Kit before being sent to sequencing. Total DNA was extracted as described previously (Thureborn et al. 2013) and subsequently pooled before PCR amplification and sequencing. The V3-V4 region from 16S rDNA was amplified with primers 341F (CCTACGGGNGGCWGCAG) and 805R (GACTACHVGGGTATCTAATCC) (Herlemann, ISME, 2011, Hugerth et al. 2014), using KAPA HiFi HotStart ReadyMix (2X) (KAPA Biosystem, Kit Code KK2602). A two-step PCR procedure was applied for generating and barcoding amplicons (Hugerth, Muller, et al. 2014) with 25 and 8 cycles separately. For the amplification step, each DNA sample was individually PCR-amplified by initial denaturation at 98°C for 1 min followed by 25 cycles of 10s at 98°C, 20s at 54°C, 6s at 72°C, and a final extension at 72°C for 2 mins. For the second PCR-barcoding step, the annealing temperature was 62°C lasting for 30 s. Between the first and second PCR, and prior to pooling library, amplicons were purified with 8.8% PEG 6000 (polyethylene glycol 6000) (Merck Millipore co., Cat. No. 528877-1KG) precipitation buffer and CA beads (carboxylic acid-coated superparamagnetic beads) (Dynabeads® MyOne™ Carboxylic Acid, Cat. No. 65012) (Lundin et al. 2010). Gel electrophoresis (1% agarose in TAE buffer 1x) and Qubit 2.0 Fluorometer (Invitrogen, Qubit-IT dsDNA kit) were used for checking amplicon size and for DNA quantification.

### Sequencing

Total RNA and DNA was shotgun sequenced using Illumina HiSeq 2000 instrumentation and amplicon sequencing was performed with Illumina MiSeq (Illumina, Inc, USA). All sequencing processes were performed at the Science for Life Laboratory (SciLifeLab),

Stockholm. All sequence data are deposited at the European Nucleotide Archive (ENA) under the study accession number PRJEB13658 (16S rDNA amplicons) and PRJEB6616 (totalRNA and totalDNA), respectively.

### *Bioinformatics*

16S rDNA amplicon sequence pairs were merged, quality filtered and clustered into operational taxonomic units (OTUs) at 1.0% dissimilarity using the UPARSE pipeline (Edgar, 2013). Total RNA and total DNA sequence reads were mapped, forward and reverse reads from pairs separately, against OTU centroid sequences using BWA (Li and Durbin 2009) with default settings. See Table 1 for sequence count summaries. The taxonomic affiliation of OTUs was determined using the LCA algorithm of the SINA aligner (Pruesse, Peplies, and Glöckner 2012) against the SILVA database release 111 (Quast et al. 2013). The abundances of OTUs in the totalRNA and totalDNA libraries were calculated by mapping reads against the seed sequence of each OTU cluster.

Additionally, 16S sequence reads in totalRNA and totalDNA datasets were taxonomically classified using BLAST against the SILVA database and subsequent MEGAN analysis using default settings. Shannon diversity was calculated using the Vegan R package (Oksanen, J. et al. 2011), subsampling the samples to the size of the smallest (totalDNA, 4,411 observations) and second smallest (rDNA, 22,350 observations) datasets 50 times and calculating the mean.

Ratios of totalRNA to rDNA significantly higher or lower than the community median were calculated in R using the sign test (binom.test function in R) after counting the number of OTUs in each major taxon higher or lower than the median. Correction for multiple comparisons was applied using the p.adjust method with the 'fdr' method.

### **Acknowledgements**

We thank Anders Andersson for help with 16S PCRs and Illumina MiSeq sequencing. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under project b2011178. The work was funded by the Foundation for Baltic and East European Studies 1169/42/2007:17 ([www.ostersjostiftelsen.se](http://www.ostersjostiftelsen.se)) to SS, Britt-Marie Sjöberg and Anthony Poole, the EU Metaexplore project (KBBE-222625) to SS and Stockholm Läns Landsting to SS and PT. YOOH was supported by a scholarship from the China Scholarship Council (CSC#201206950024).

### **References**

- Andersson, Anders F., Lasse Riemann, and Stefan Bertilsson. 2009. "Pyrosequencing Reveals Contrasting Seasonal Dynamics of Taxa within Baltic Sea Bacterioplankton Communities." *The ISME Journal* 4 (2): 171–81. doi:10.1038/ismej.2009.108.
- Bar-Or, I., E. Ben-Dov, A. Kushmaro, W. Eckert, and O. Sivan. 2015. "Methane-Related Changes in Prokaryotes along Geochemical Profiles in Sediments of Lake Kinneret (Israel)." *Biogeosciences* 12 (10): 2847–60. doi:10.5194/bg-12-2847-2015.
- Blazewicz, Steven J., Romain L. Barnard, Rebecca A. Daly, and Mary K. Firestone. 2013. "Evaluating rRNA as an Indicator of Microbial Activity in Environmental Communities: Limitations and Uses." *The ISME Journal* 7 (11): 2061–68. doi:10.1038/ismej.2013.102.
- Conley, Daniel J., Svante Björck, Erik Bonsdorff, Jacob Carstensen, Georgia Destouni, Bo G Gustafsson, Susanna Hietanen, et al. 2009. "Hypoxia-Related Processes in the Baltic Sea." *Environmental Science & Technology* 43 (10): 3412–20.
- Corinaldesi, C., M. Barucca, G. M. Luna, and A. Dell'anno. 2011. "Preservation, Origin and Genetic Imprint of Extracellular DNA in Permanently Anoxic Deep-Sea Sediments." *Molecular Ecology* 20 (3): 642–54. doi:10.1111/j.1365-294X.2010.04958.x.
- Cottingham, Kathryn L., Holly A. Ewing, Meredith L. Greer, Cayelan C. Carey, and Kathleen C. Weathers. 2015. "Cyanobacteria as Biological Drivers of Lake Nitrogen and Phosphorus Cycling." *Ecosphere* 6 (1): 1–19. doi:10.1890/ES14-00174.1.
- Edlund, Anna, Fredrik Hårdeman, Janet K Jansson, and Sara Sjöling. 2008. "Active Bacterial Community Structure along Vertical Redox Gradients in Baltic Sea Sediment." *Environmental Microbiology* 10 (8): 2051–63. doi:10.1111/j.1462-2920.2008.01624.x.
- Egert, Markus, Bianca Wagner, Thorsten Lemke, Andreas Brune, and Michael W. Friedrich. 2003. "Microbial Community Structure in Midgut and Hindgut of the Humus-Feeding Larva of Pachnoda Ehippiata (Coleoptera: Scarabaeidae)." *Applied and Environmental Microbiology* 69 (11): 6659–68. doi:10.1128/AEM.69.11.6659-6668.2003.
- Frias-Lopez, Jorge, Yanmei Shi, Gene W. Tyson, Maureen L. Coleman, Stephan C. Schuster, Sallie W. Chisholm, and Edward F. DeLong. 2008. "Microbial Community Gene Expression in Ocean Surface Waters." *Proceedings of the National Academy of Sciences* 105 (10): 3805–3810.
- Herlemann, Daniel PR, Matthias Labrenz, Klaus Jürgens, Stefan Bertilsson, Joanna J. Waniek, and Anders F. Andersson. 2011. "Transitions in Bacterial Communities along the 2000 Km Salinity Gradient of the Baltic Sea." *The ISME Journal* 5 (10): 1571–79. doi:10.1038/ismej.2011.41.
- Hoover, Richard B., Elena V. Pikuta, Asim K. Bej, Damien Marsic, William B. Whitman, Jane Tang, and Paul Krader. 2003. "Spirochaeta Americana Sp. Nov., a New Haloalkaliphilic, Obligately Anaerobic Spirochaete Isolated from Soda Mono Lake in California." *International Journal of Systematic and Evolutionary Microbiology* 53 (3): 815–21. doi:10.1099/ijs.0.02535-0.
- Hug, Laura A., Cindy J. Castelle, Kelly C. Wrighton, Brian C. Thomas, Itai Sharon, Kyle R. Frischkorn, Kenneth H. Williams, Susannah G. Tringe, and Jillian F. Banfield. 2013. "Community Genomic Analyses Constrain the Distribution of Metabolic Traits across the Chloroflexi Phylum and Indicate Roles in Sediment Carbon Cycling." *Microbiome* 1: 22. doi:10.1186/2049-2618-1-22.
- Hugerth, Luisa W., Emilie E. L. Muller, Yue O. O. Hu, Laura A. M. Lebrun, Hugo Roume, Daniel Lundin, Paul Wilmes, and Anders F. Andersson. 2014. "Systematic Design of 18S rRNA Gene Primers for Determining Eukaryotic Diversity in Microbial Consortia." *PLoS ONE* 9 (4): e95567. doi:10.1371/journal.pone.0095567.
- Hugerth, Luisa W., Hugo A. Wefer, Sverker Lundin, Hedvig E. Jakobsson, Mathilda Lindberg, Sandra Rodin, Lars Engstrand, and Anders F. Andersson. 2014. "DegePrime, a Program for Degenerate Primer Design for Broad-Taxonomic-Range



- PCR in Microbial Ecology Studies.” *Applied and Environmental Microbiology* 80 (16): 5116–23. doi:10.1128/AEM.01403-14.
- Huson, Daniel H., Alexander F. Auch, Ji Qi, and Stephan C. Schuster. 2007. “MEGAN Analysis of Metagenomic Data.” *Genome Research* 17 (3): 377–86. doi:10.1101/gr.5969107.
- Kan, Jinjun, Sarah E. Evans, Feng Chen, and Marcelino T. Suzuki. 2008. “Novel Estuarine Bacterioplankton in rRNA Operon Libraries from the Chesapeake Bay.” *Aquatic Microbial Ecology* 51 (1): 55–66.
- Kirschner, Richard A., Bruce C. Parker, and Joseph O. Falkinham. 1999. “Humic and Fulvic Acids Stimulate the Growth of Mycobacterium Avium.” *FEMS Microbiology Ecology* 30 (4): 327–32. doi:10.1111/j.1574-6941.1999.tb00660.x.
- Lepp, P. W., and T. M. Schmidt. 1998. “Nucleic Acid Content of Synechococcus Spp. during Growth in Continuous Light and Light/Dark Cycles.” *Archives of Microbiology* 170 (3): 201–7. doi:10.1007/s002030050634.
- Li, Heng, and Richard Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform.” *Bioinformatics* 25 (14): 1754–60. doi:10.1093/bioinformatics/btp324.
- Lindh, Markus V., Johanna Sjöstedt, Anders F. Andersson, Federico Baltar, Luisa W. Hugerth, Daniel Lundin, Saraladevi Muthusamy, Catherine Legrand, and Jarone Pinhassi. 2015. “Disentangling Seasonal Bacterioplankton Population Dynamics by High-Frequency Sampling.” *Environmental Microbiology* 17 (7): 2459–76. doi:10.1111/1462-2920.12720.
- Liu, Yuchen, and William B. Whitman. 2008. “Metabolic, Phylogenetic, and Ecological Diversity of the Methanogenic Archaea.” *Annals of the New York Academy of Sciences* 1125 (1): 171–89. doi:10.1196/annals.1419.019.
- Lundin, Sverker, Henrik Stranneheim, Erik Pettersson, Daniel Klevebring, and Joakim Lundeberg. 2010. “Increased Throughput by Parallelization of Library Preparation for Massive Sequencing.” *PLOS ONE* 5 (4): e10029. doi:10.1371/journal.pone.0010029.
- Muyzer, Gerard, and Alfons J. M. Stams. 2008. “The Ecology and Biotechnology of Sulphate-Reducing Bacteria.” *Nature Reviews Microbiology* 6 (6): 441–54. doi:10.1038/nrmicro1892.
- Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., O’Hara, R. B., Simpson, G. L., Solymos, P., Henry, M., Stevens, H., and Wagner, H. 2011. “Vegan: Community Ecology Package.” <http://CRAN.R-project.org/package=vegan>.
- Olli, K., K. Kangro, and M. Kabel. 2005. “Akinete Production of Anabaena Lemmermannii and A. Cylindrica (Cyanophyceae) in Natural Populations of N- and P-Limited Coastal Mesocosms.” *Journal of Phycology* 41 (6): 1094–98. doi:10.1111/j.1529-8817.2005.00153.x.
- Pruesse, Elmar, Jörg Peplies, and Frank Oliver Glöckner. 2012. “SINA: Accurate High-Throughput Multiple Sequence Alignment of Ribosomal RNA Genes.” *Bioinformatics* 28 (14): 1823–29. doi:10.1093/bioinformatics/bts252.
- Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. O. Glockner. 2013. “The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools.” *Nucleic Acids Research* 41 (D1): D590–96. doi:10.1093/nar/gks1219.
- Sivonen, Kaarina, Katrianna Halinen, Leila M. Sihvonen, Kerttu Koskenniemi, Hanna Sinkko, Kaisa Rantasärkkä, Pia H. Moisander, and Christina Lyra. 2007. “Bacterial Diversity and Function in the Baltic Sea with an Emphasis on Cyanobacteria.” *AMBIO: A Journal of the Human Environment* 36 (2): 180–85. doi:10.1579/0044-7447(2007)36[180:BDAFIT]2.0.CO;2.
- Suikkanen, Sanna, Hermann Kaartokallio, Seija Hällfors, Maija Huttunen, and Maria Laamanen. 2010. “Life Cycle Strategies of Bloom-Forming, Filamentous



- Cyanobacteria in the Baltic Sea.” *Deep Sea Research Part II: Topical Studies in Oceanography*, Phytoplankton Life-Cycles and Their Impacts on the Ecology of Harmful Algal Bloom, 57 (3–4): 199–209. doi:10.1016/j.dsr2.2009.09.014.
- Sukenik, Assaf, Ruth N. Kaplan-Levy, Jessica Mark Welch, and Anton F. Post. 2012. “Massive Multiplication of Genome and Ribosomes in Dormant Cells (Akinetes) of *Aphanizomenon Ovalisporum* (Cyanobacteria).” *The ISME Journal* 6 (3): 670–79. doi:10.1038/ismej.2011.128.
- Thureborn, Petter, Andrea Franzetti, Daniel Lundin, and Sara Sjöling. 2016. “Reconstructing Ecosystem Functions of the Active Microbial Community of the Baltic Sea Oxygen Depleted Sediments.” *PeerJ* 4 (January): e1593. doi:10.7717/peerj.1593.
- Thureborn, Petter, Daniel Lundin, Josefin Plathan, Anthony M. Poole, Britt-Marie Sjöberg, and Sara Sjöling. 2013. “A Metagenomics Transect into the Deepest Point of the Baltic Sea Reveals Clear Stratification of Microbial Functional Capacities.” *PLoS ONE* 8 (9): e74983. doi:10.1371/journal.pone.0074983.
- Torti, Andrea, Mark Alexander Lever, and Bo Barker Jørgensen. 2015. “Origin, Dynamics, and Implications of Extracellular DNA Pools in Marine Sediments.” *Marine Genomics* 24, Part 3 (December): 185–96. doi:10.1016/j.margen.2015.08.007.
- Turley, Carol. 2000. “Bacteria in the Cold Deep-Sea Benthic Boundary Layer and Sediment—water Interface of the NE Atlantic.” *FEMS Microbiology Ecology* 33 (2): 89–99. doi:10.1111/j.1574-6941.2000.tb00731.x.
- Villareal, T. A., and E. J. Carpenter. 2003. “Buoyancy Regulation and the Potential for Vertical Migration in the Oceanic Cyanobacterium *Trichodesmium*.” *Microbial Ecology* 45 (1): 1–10. doi:10.1007/s00248-002-1012-5.
- Zinger, Lucie, Linda A. Amaral-Zettler, Jed A. Fuhrman, M. Claire Horner-Devine, Susan M. Huse, David B. Mark Welch, Jennifer B. H. Martiny, Mitchell Sogin, Antje Boetius, and Alban Ramette. 2011. “Global Patterns of Bacterial Beta-Diversity in Seafloor and Seawater Ecosystems.” *PLoS ONE* 6 (9): e24570. doi:10.1371/journal.pone.0024570.

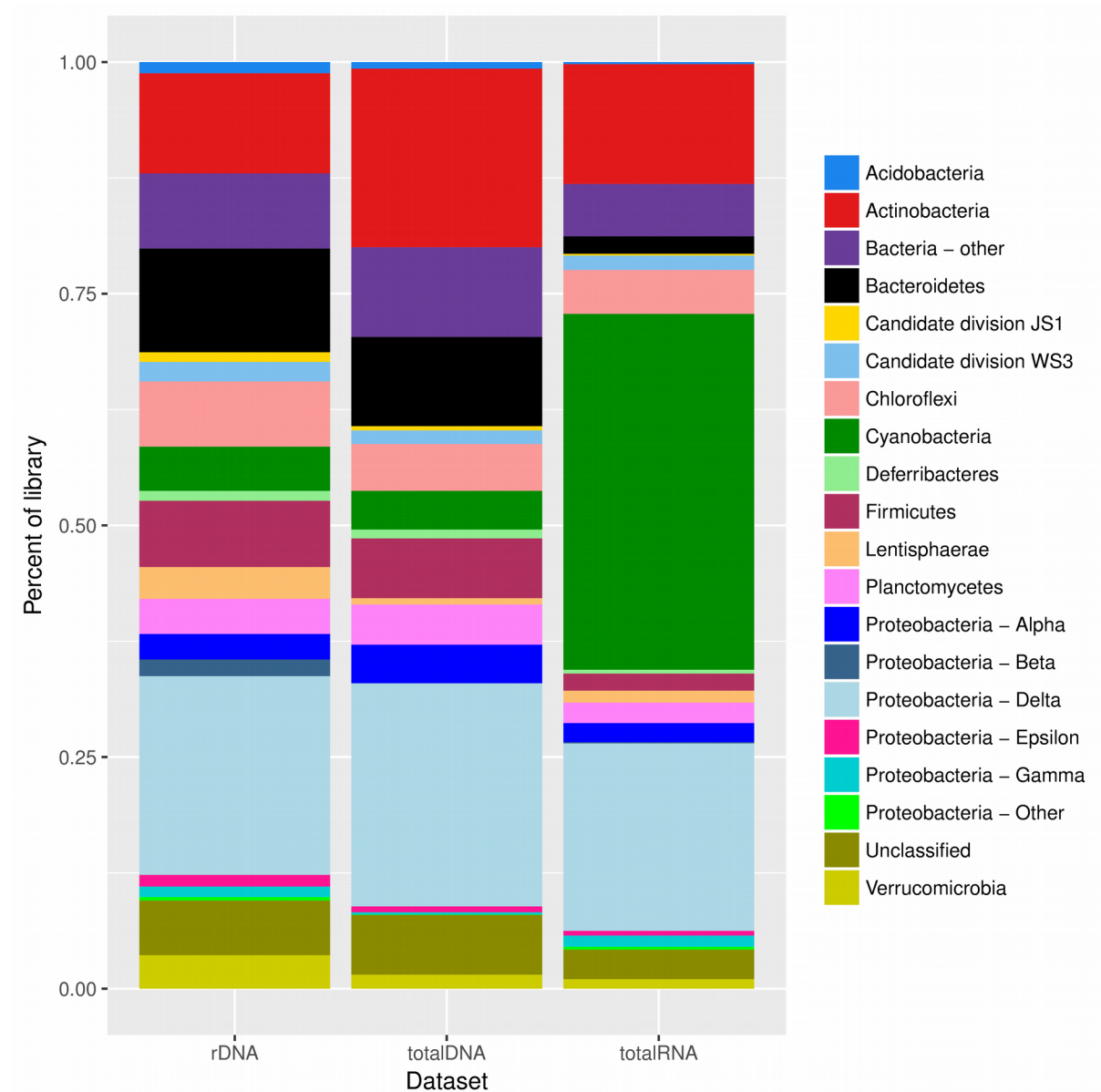
# Tables

385 **Table 1. Sequencing output and Shannon diversity.** Number of sequence reads per  
sequence dataset that match OTU seed sequences, excluding chloroplast OTUs. Shannon  
diversity calculated as the mean of 50 subsampling iterations to 4,411 observations per  
sample (size of the totalDNA sample). In parentheses, values after subsampling to 22,350  
(size of the rDNA sample).

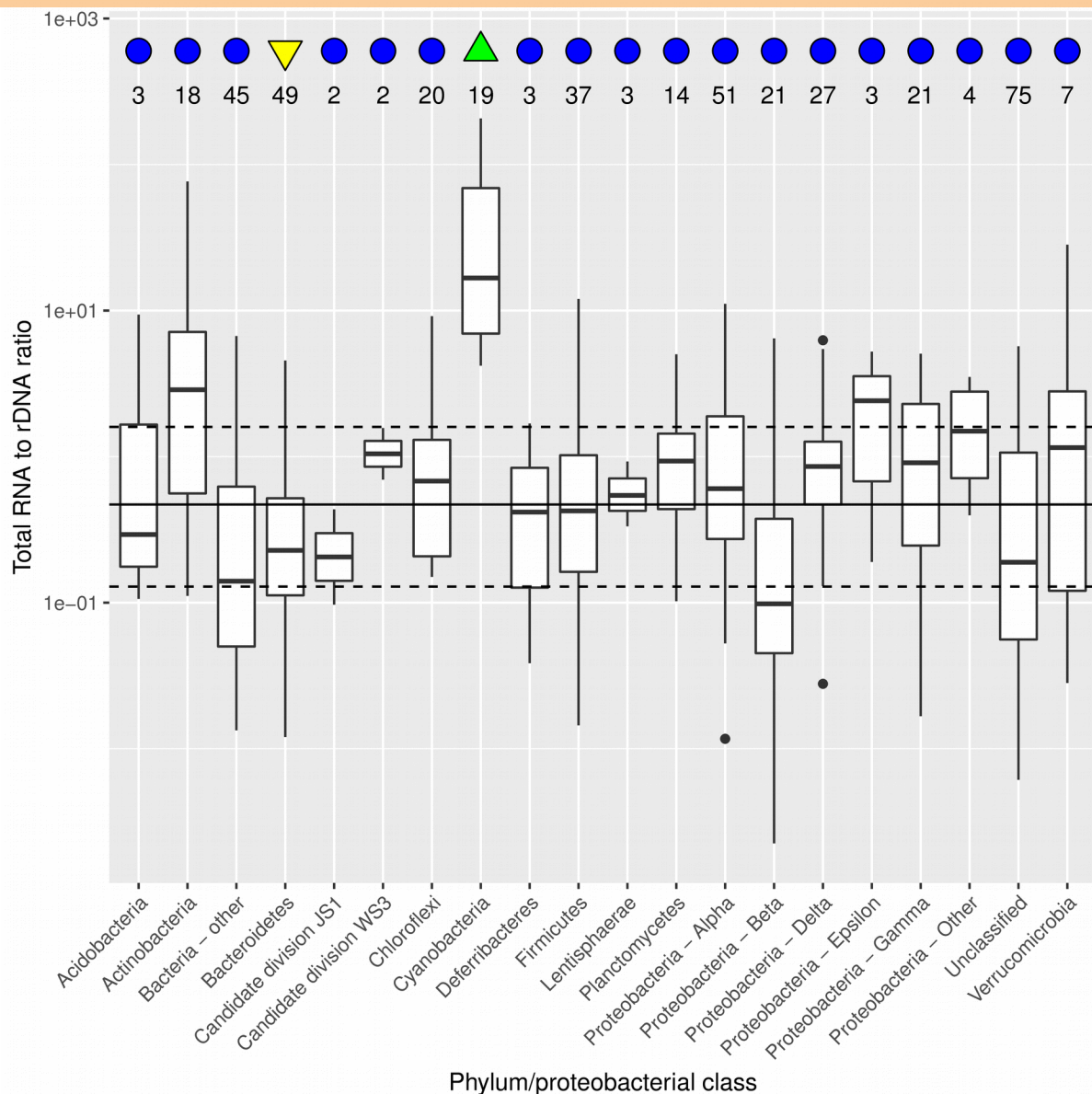
	rDNA	totalDNA	totalRNA
Count	22,350	4,411	2,231,099
N. OTUs	424	205	424
Shannon	4.91 (4.95)	4.58	4.40 (4.44)

390

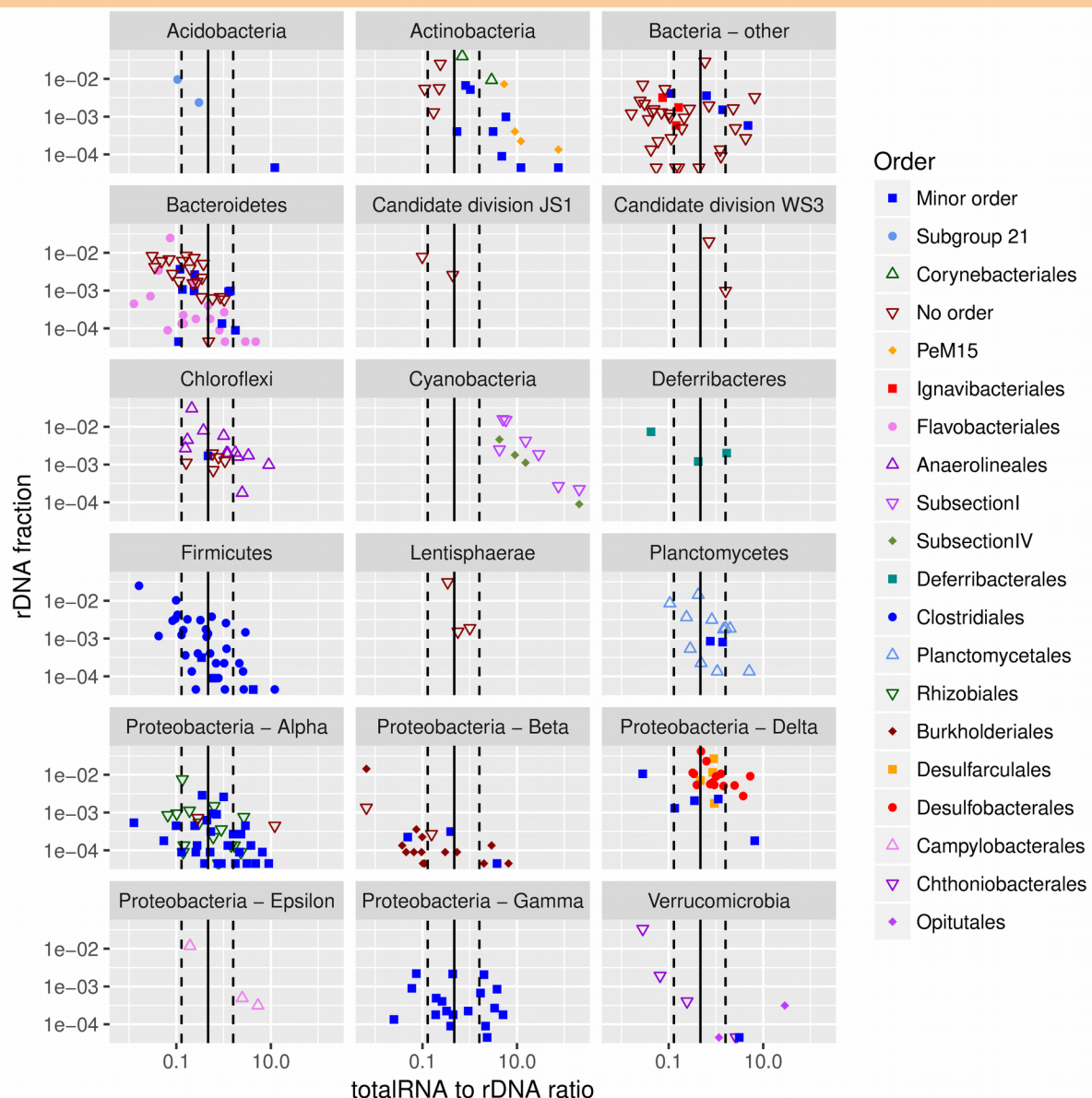
# Figures



**Figure 1. Distribution of major taxonomic groups among amplicon OTUs.** Distribution of phyla and proteobacterial classes in the three datasets, counted in rDNA amplicon OTUs.

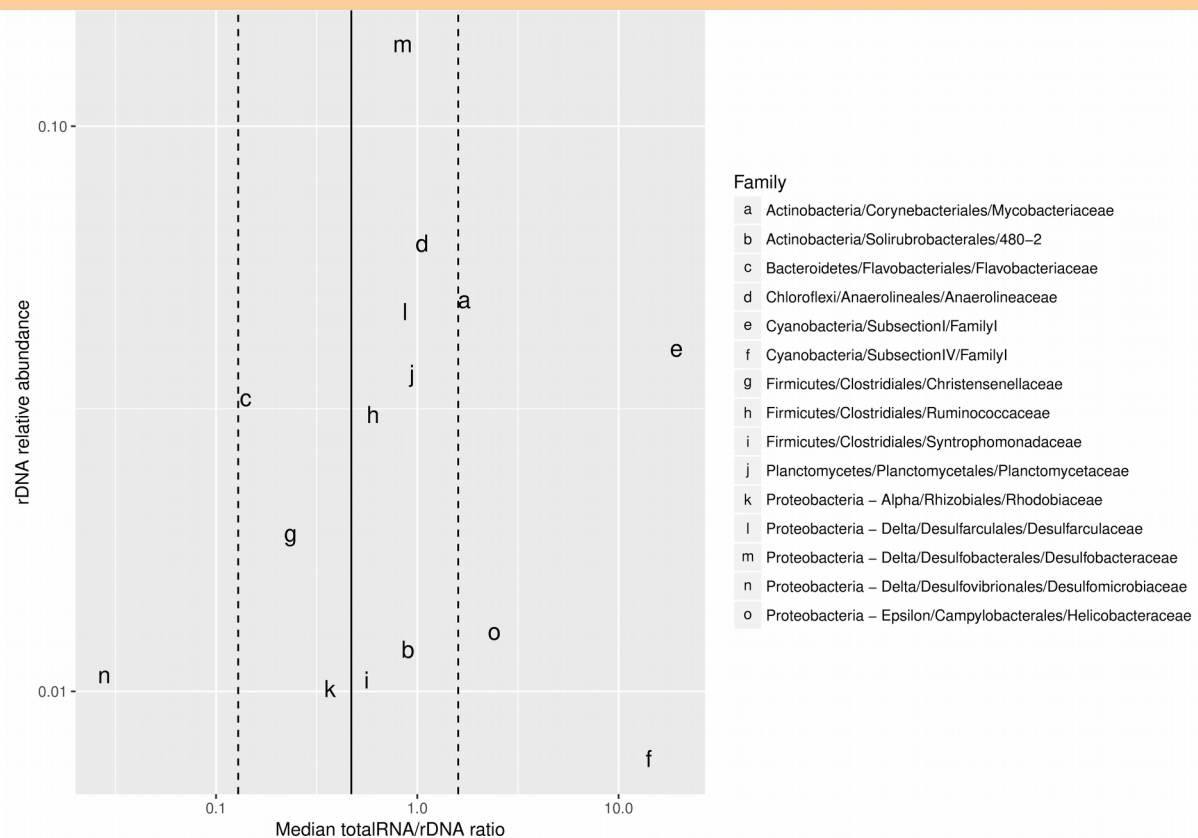


395 **Figure 2. Distribution of ratios between totalRNA and rDNA in OTUs among the major bacterial phyla and proteobacterial classes.** Taxa with a distribution of ratios significantly greater or lower than the community median (0.47) are indicated with green and yellow triangles, respectively. Taxa with a ratio not significantly higher nor lower than the community median (sign test,  $p < 0.05$ ) are indicated by a blue circle (sign test; adjusted for multiple comparisons). The community median (0.47) is indicated by a solid black line and interquartile limits are indicated by dashed lines. Only taxa with a frequency of at least 1 % in  
400 one of the datasets are shown. Note the log scale on the y-axis.



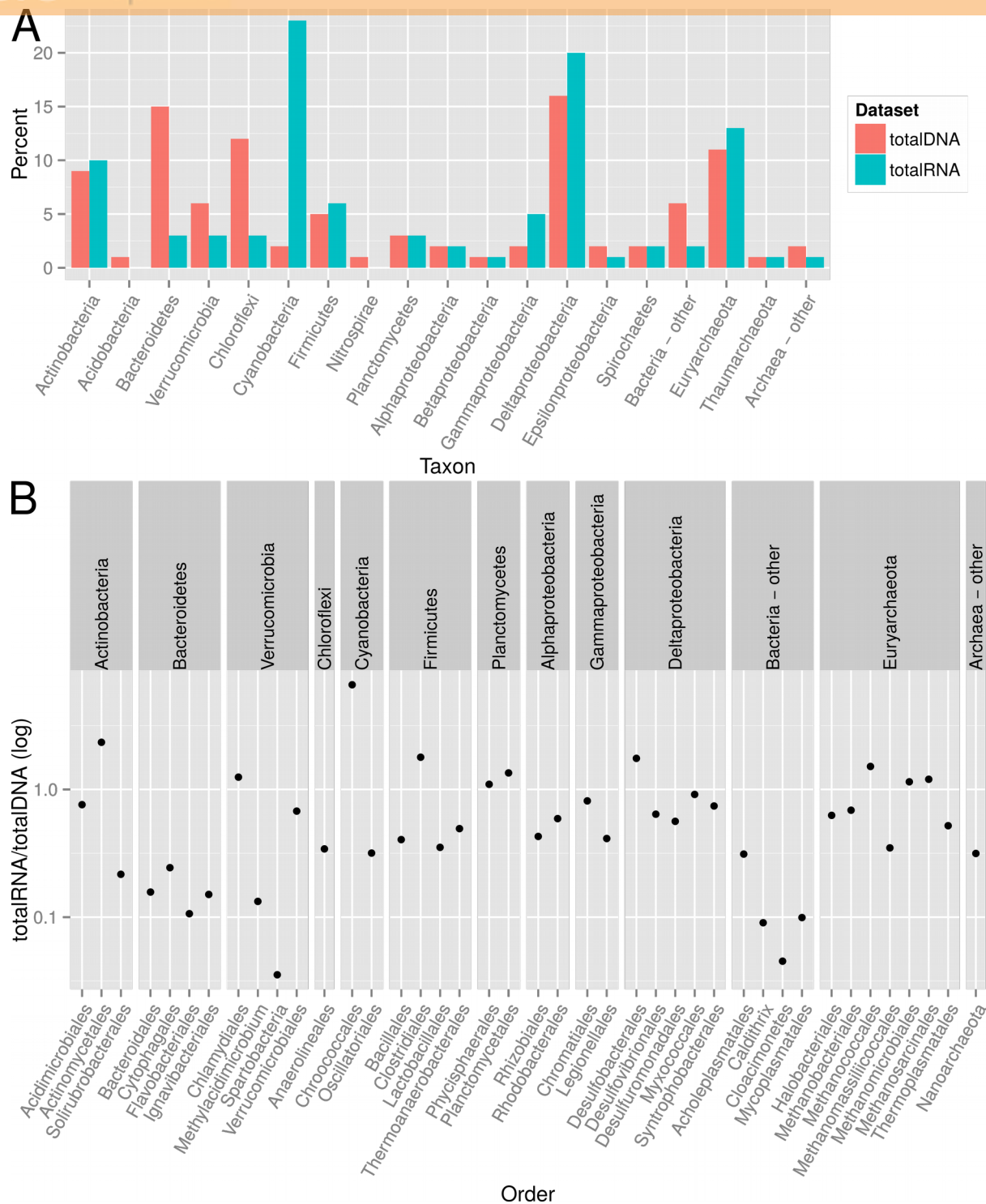
**Figure 3. rDNA abundances as a function of totalRNA/rDNA ratios for the most abundant phyla and proteobacterial classes.** For clarity, ratios were grouped into

percentiles for which relative abundances were calculated by summing all OTUs' relative abundances. Each symbol is one ratio percentile, plotted on the x-axes at the top totalRNA/rDNA ratio within the percentile. In each taxon, the top two orders are indicated with colour and shape, excluding orders with a relative abundance lower than 0.5 %; all other orders are grouped in "Minor orders". Community-wide median (solid line) and interquartile ranges (dashed lines) are indicated. Note that both the x-axes and y-axes are logarithmic.



**Figure 4. TotalRNA/rDNA ratio medians and rDNA abundances per family.** Families larger than 1 % in either the totalRNA or the rDNA datasets are presented with their median totalRNA/rDNA ratios and rDNA relative abundances. Note that both scales are logarithmic.





**Figure 5. Ab initio identified 16S sequences.** A) Proportion of phyla and proteobacterial classes in amplicon-independent totalRNA and totalDNA analysis. B) TotalRNA/totalDNA ratios for orders appearing in the totalDNA dataset.