Meta-analysis cum machine learning approaches address the structure and biogeochemical potential of marine copepods associated bacteriobiome

Balamurugan Sadaiappan[1, +], PrasannaKumar Chinnamani[1, +], Uthara V Nambiar[1], Mahendran Subramanian[2], Manguesh U Gauns[1, *]

[1]Plankton Ecology Lab, Biological Oceanography Division, CSIR-National Institute of Oceanography, Dona Paula, Panaji- 403004, Goa, India
[2]Department of Bioengineering and Department of Computing, Imperial College London, South Kensington- SW72AZ, London, United Kingdom.

+Equal contribution

*** corresponding author Manguesh U Gaun's email id: gmangesh@nio.org**

Abstract

Copepods are dominant members of the zooplankton and the most abundant forms of life. Studying the bacterial diversity associated with copepods will helps in understanding the impact of global climate change on these organisms. It is important to address the core microbiome of copepods which has a key role in their host health and ocean biogeochemical cycle. Early studies have identified few bacterial phyla and orders as core microbiome. So to predict the important Operational taxonomic units (OTUs), we used meta-analysis, and machine learning (RandomForest Classifier) approaches. Also, we explore the biogeochemical potential of copepods associated bacteriobiome (CAB). Overall, 50 important s-OTUs were predicted by machine learning; among them, 38 s-OTUs were specific to *Calanus* spp. and 17 s-OTUs were specific to *Acartia* spp. Six bacterial genera were identified as important core sub-OTUs in copepods for the first time, i.e. *Micrococcus luteus*, *Krokinobacter eikastus*, *Vibrio shilonii*, *Acinetobacter johnsonii Burkholderia* and *Sphingobium.* From the PICRUST2 analysis, the potential genes responsible for methanogenesis (aerobic and anaerobic), methanotrophy and iron fertilization were high in the CAB of *Pleuromamma* spp.. The potential nitrogen-fixing genes were relatively high in the CAB of *Pleuromamma* spp.. Whereas the potential genes for denitrification were relatively high in the CAB of *Temora* spp., and the potential Dissimilatory Nitrate Reduction (DRNA) genes were relatively high in Acartia spp.. All the CAB of the copepod genera investigated in the present study has potential genes for cobalamin synthesis.

Keywords: Copepod associated bacteriobiome, machine learning, *Acartia* spp*., Temora* spp*., Pleuromamma* spp*., Centropages* spp*., Calanus* spp*.,* methanogenesis, cyanocobalamine.

1. Introduction

Copepods (Subphylum Crustacea; Class Maxillopoda; Subclass Copepoda) are an abundant and diverse group of zooplankton in the ocean (Datta et al., 2018; Shoemaker and Moisander, 2017). They play a key role in the energy transfer within the pelagic food web (Steinberg et al., 2000). They are also well-known for their wide-ranging and flexible feeding approaches (Mianrun Chen et al., 2018). Copepods, usually not more than a millimetre in length, supports a wide range of bacterial associations, due to the release of organic and inorganic nutrients during feeding and excretion (Shoemaker and Moisander, 2017; Datta et al., 2018). Exchange of bacterial community between the copepods and water-column is a well-established fact (De Corte et al., 2014). Moreover, the bacterial association with copepods differ within the body parts of a copepod, also during the vertical migration and the life stages (Datta et al., 2018; Moller et al., 2007; Tang et al., 2010). Understanding the relationship between copepods and its bacterial community could predict the impacts of future oceanic conditions on copepods.

Next-generation DNA sequencers such as Illumina platforms are known for massive data generation for understanding CAB. Through sequencing the V3-V4 hypervariable regions of 16S rDNA genes, it was observed that the percentage of Gammaproteobacteria was more copious in starved *Centropages* sp. And *Acartia* sp. than their full gut counterparts (Moisander et al., 2015). Likewise, Gammaproteobacteria was observed to be abundant in *Pleuromamma* sp. (Cregeen, 2016). Also, eight bacterial orders such as Lactobacillales, Bacillales, Actinomycetales, Rhizobiales, Vibrionales, Pseudomonadales and Flavobacteriales were found as core members in *Pleuromamma* spp. (Shoemaker and Moisander, 2017). The phylum Proteobacteria was identified as core OTUs along with Actinobacteria and Bacteroidetes in C*alanus finmarchicus* (Datta et al., 2018). Datta et al. (2018) found the distinct bacterial communities between the diapause phase and actively feeding *Calanus finmarchicus*. The bacterial family, Flavobacteriaceae, was meagre in copepods during diapause and abundant in actively feeding counterparts. Datta et al. (2018) reported that *Marinimicrobium* (*Alteromonadaceae*) was relatively abundant in deep-dwelling copepods than its shallow counterparts and concluded that the copepods have inter-individual microbiome variations and the factors driving these variations are still unknown.

Moreover, the gut of *Calanus* species has low pH and different oxygen gradient from the anal opening to the metasome region. It may selectively have certain groups of bacteria which could be specialized in iron dissolution, anaerobic methanogenesis (Tang et al., 2011) and dinitrogen (N2) –fixation (Proctor, 1997). If we assume one copepod per litre of seawater, the relative contribution of CAB to the total bacteria in seawater would be less than 2–3 orders, but the contribution of CAB to the marine biogeochemical cycles will be significant (Shoemaker and Moisander, 2017). Already various studies have shown that CAB has a potential role in biogeochemical processes, such as nitrogen-fixation, (Proctor, 1997; Scavotto et al., 2015), denitrification (De Corte et al., 2018), carbon, sulfur (Dong et al., 2013) and iron mineralization processes (Tang et al., 2011). It is important to address the core microbiota of copepods which has a key role in their host health and ocean biogeochemical cycle.

The masking effect of the abundant bacterial community associated with copepod diet, copepod life stage, and environmental conditions was considered the main hindrance in defining core bacterial operational taxonomic units (OUTs; equivalent to species) specific to copepod genera (example; Wage et al., 2019, Moisander et al., 2015; De Corte et al., 2018), which we aimed to overcome by using meta-analysis cum machine learning approaches.

The meta-analysis, a set of methods used to organize and combine "the results of several reports to create a single, and more precise results" (Ferrer, 1998). It is a powerful approach (Rocca et al., 2018; Wirbel et al., 2019) to understand the relationship between the copepods and its associated bacterial community. We analyzed 16S rDNA gene sequences (V3-V4 & V4-V5 regions; ~16.5 million reads) of CAB belonging to 5 different copepod genera using Quantitative Insights Into Microbial Ecology (QIIME2) software package (Bolyen et al., 2019). We hypothesized that if copepod genera have specific OTUs then different copepod has a differential CAB, and the biogeochemical potential of the CAB will differ. We used Random Forest classifier, a machine learning approach and Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2) (Douglas et al., 2020) analysis to test this hypothesis.

2. Methodology

2.1 Data collection

We systematically reviewed the studies related to copepod associated microbiome. The relevant published research articles were searched and retrieved from PubMed, Google scholar, and SCOPUS using keywords such as copepods gut microbiome, copepod associated microbiome, copepods gut flora, copepod microbiome and zooplankton associated microbiome on the Jan 30th, 2020. Apart from the article search, we also searched in public databases (for published and unpublished ion torrent, pyro and Illumina sequence data) such as the NCBI-SRA, ENA and figshare using the above-mentioned keywords.

Herein, the terminology 'bacteriobiome' means the total bacterial composition inhabiting in a specific biological niche (example; copepods), including their genomic content and metabolic products (Marchesi &Ravel, 2015). It is a well-known fact that host-associated microbial communities remain essential for maintaining any ecosystems, and any variation in these communities can be unfavorable, i.e. the human microbiome plays an import role in development, immunity, and even behavior of their hosts (Gilbert et al., 2018).

Overall of 11 study data were retrieved for meta-analysis (Table S1) containing 549 next-generation sequence libraries. We separately pre-processed every individual file within the study and prepared the quality control (QC) report (Table 1).

2.2. Pre-processing

The sequence quality was checked with FastQC tool (Joseph Brown et al., 2017) and the minimum base per quality for future analysis was fixed as PHRED >25. Based on the QC high rates of erroneous sequences form Illumina, 454 and ion torrent files (Table 1) were removed from the further meta-analysis. The two major reasons for the exclusion are 1) erroneous sequences (of PHRED <25) and 2) Short reads (<200 bps) screened by DADA2 (Callahan et al., 2016) while picking sub-Operational Taxonomic Units (s-OUT). Overall, Illumina sequences contained better quality than the Ion-torrent and Pyrosequence (Table 1). Finally, we did meta-analysis with 453 files of copepods associated microbiome to test the proposed hypothesis.

2.3. Meta-analysis

2.3.1. Sequence screening and preparations for meta-analysis

We used Quantitative Insights Into Microbial Ecology (QIIME2) version 2019.10 (Bolyen et al., 2019), for the meta-analysis. QIIME2 pipeline provides a start-to-finish workflow, beginning with demultiplexing sequence reads and finishing with taxonomic and phylogenetic profiles. The sequences from the individual study were imported to QIIME2 using CasavaOneEight format, and the quality of the sequences was checked by the default settings in QIIME2. Based on the sequence quality, the sequence was trimmed, denoised, aligned and checked for chimera using DADA2 (single and paired-ends sequence were trimmed based on the length of primer used) (Callahan et al., 2016). The feature table and representative sequence of each file were merged using QIIME2 feature merge table and merge representative sequences.

2.3.2. Taxonomic classification

The merged files were aligned to phylogeny against the Greengene reference sequence sepp-refs-gg-13-8 using q2-fragment-insertion (Janssen et al., 2018). Incorrect taxonomic and phylogenetic assignments due to differences in 16S rDNA hypervariable regions and merging the variable lengths during analysis were solved with q2-fragment insertion technique (SATe-enabled phylogenetic placement in QIIME2 plugin) (Janssen et al., 2018). The core diversity was calculated before (to calculate the impact on diversity) and after removing mitochondria (mtDNA) and chloroplast (clDNA) sequence from the dataset. The mtDNA and clDNA filtered dataset was further used for calculating diversity, taxonomy, important (core) s-OTUs and the difference in composition estimation using QIIME2 and the diversity graph was plotted using R phyloseq (McMurdie & Holmes, 2013). We used Unweighted, Weighted Unifrac and Jaccard distance matrix to compute the beta diversity, and the outcomes were envisaged using Principal Coordinates Analysis (PCoA) in QIIME2. A Permutational Multivariate Analysis Of Variance (PERMANOVA) (Anderson, 2017) thru the Unweighted, Weighted unifrac along with Jaccord distance-based beta-diversity was calculated within QIIME2.

We also, implement the Analysis Of the Composition of Microbiome (ANCOM) (Mandal et al., 2015) in QIIME2 plugin to identify the significantly different s-OTUs between the copepod genera. ANCOM uses F-statistics and W-statistics to determine the difference, where W represents the vigor of the ANCOM test for the tested number of species and F represents the measure of the effect size difference for a particular species between the groups (Copepods). To Predict the important bacteria associated with the copepods, we used sophisticated supervised machine learning classifier; RandomForest Classifier (Breiman, 2001) in build-in QIIME2.

The mtDNA and clDNA filtered table and representative sequence were also used as an input for predicting CAB potential metabolic function using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2) (Douglas et al., 2020). The output abundance KEGG data were analyzed in Statistical Analysis of Taxonomic and Functional Profiles (STAMP) which includes Principle Component Analysis (PCA) (Parks et al., 2014) to find the significant difference in potential functions of CAB between the copepods genera using Kruskal–Wallis H-test (Kruskal & Wallis, 1952) with Tukey–Kramer parameter (Tukey–Kramer, 2013).

2.4. Copepod phylogeny

Cytochrome Oxidase Subunit 1 (COI) gene (mined from Genbank) of 5 copepod genera (of the present study) constituting 42 COI sequences (28th, Dec 2019) were aligned, and five consensus sequences, representing from each copepod genera were synthesized using Bio-edit (Hall, 1999). The phylogenetic Neighbor-joining tree was constructed using MEGA ver. 10 (Tamura et al., 2007).

3. Result and discussion

New bioinformatics tools have been created to cope up with data generated by the next-generation sequencers (Siegwald et al., 2019). To overcome the bias in the tools we used standard, well-recognized pipelines such as FastQC and QIIME2 demultiplexing statistics for reading the quality of sequence, DADA2 algorithm for clustering, aligning and filtering of chimaeric sequences, (Callahan et al., 2016). About, 12% (n=62), i.e. 35 Roche, 6 ion torrent and 21 Illumina generated sequence files) of the files failed during the QC were removed from the further analysis. Finally, 453 raw files belonging to 5 different copepod genera were subjected to downstream sequence analysis.

3.1. DNA sequence data analysis

We analyzed 16.5 million V3-V4 regions, (except 13 files of V4-V5 archaea specific primer files of Wage et al., (2019), Table 1) of bacterial-16S rDNA gene sequences that belongs to 5 copepod genera, i.e. *Acartia* spp., *Calanus* spp., *Centropages* spp., *Pleuromamma* spp., *Temora* spp. After quality filtering through DADA2 package, an average of 0.1 to 7.8% of sequences was removed (Table 1), and a total of 1, 39, 87, 186 sequences were used for downstream analysis. The present study represents one of the biggest CAB related DNA sequence data analyzed to date.

3.2. CAB diversity (Alpha & Beta)

We found the bacterial diversity Shannon ('H') index for the 5 copepod genera and *Calanus* spp. showed the maximum (5.36±1.29), followed by *Centropages* spp. ('H'=5.029±0.60). Furthermore, the least was observed in *Temora* spp. 2.78±1.30 (Figure 1). However, H indices were 2-3 order higher in the ambient seawater than in copepods guts (Shoemaker and Moisanders, 2017).

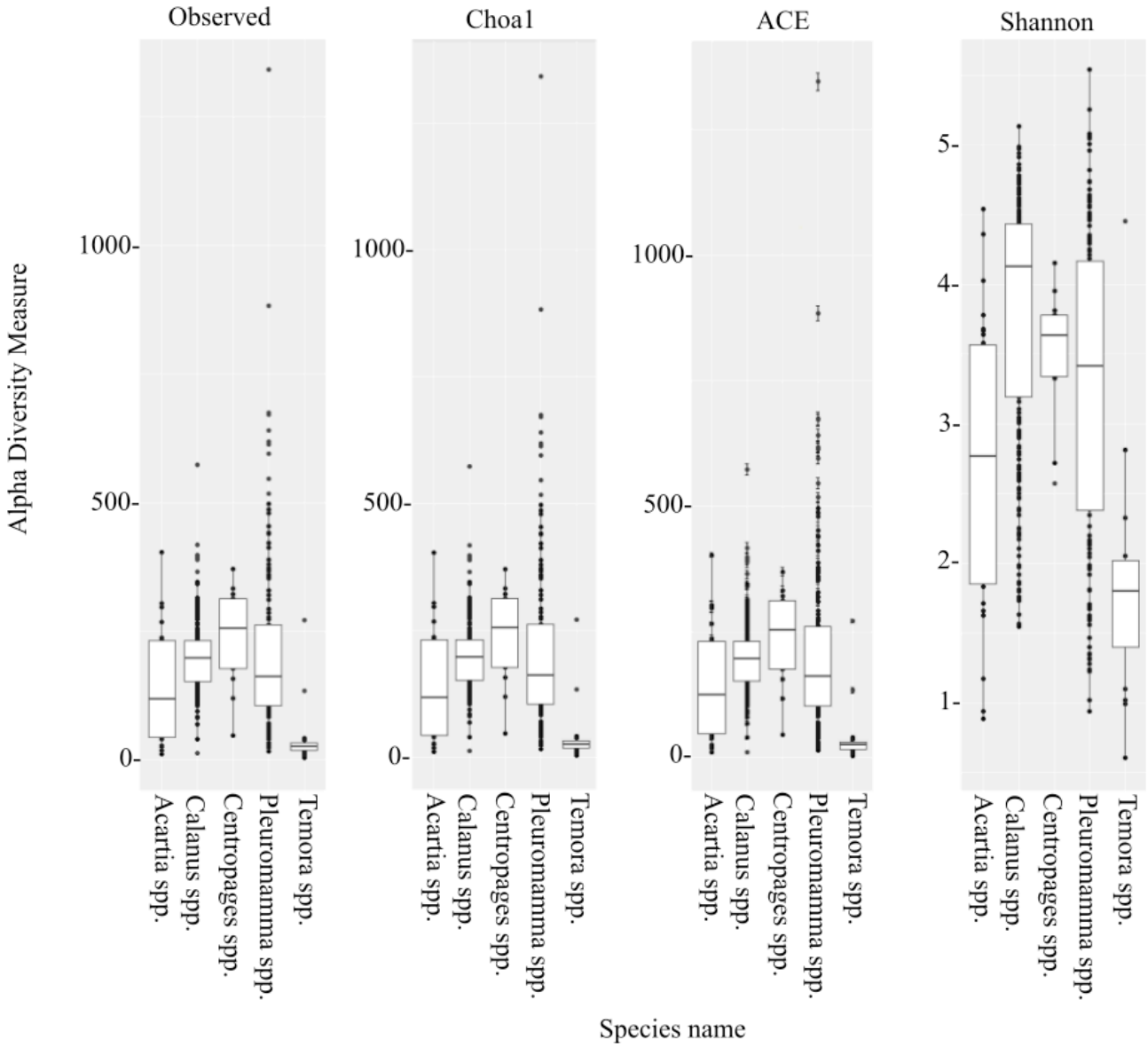

Figure.1: Alpha diversity index (Observed PD, Choa1 and Shannon) correspond to CAB in 5 different copepod genera.

The Kruskal-Wallis analysis revealed that the H index of *Acartia* spp. CAB was significantly different from the *Calanus* spp., *Centropages* spp. and *Pleuromamma* spp. with p-value in range from 0.0000002 to 0.0019 (Figure S1a). The differences may be due to their feeding habit, as *Acartia* spp. are primarily omnivores, feeds on phytoplankton and occasionally on ciliates, and rotifers (Saiz et al., 2007). Whereas, some genus in *Calanus* spp. like C. *finmarchicus* is known as filter feeders, and during energy shortfall and reproduction, they feed on ciliates and other heterotrophic protists (Ohman & Runge, 1994; Nejstgaard et al., 2001). The H index of *Temora* spp. was significantly different from *Centropages* spp. (p=0.0003) and *Pleuromamma* spp. (p=0.00006). One should note, *Temora* spp. frequently switches its feeding behavior between omnivore and herbivore based on food availability and season (Dam and Lopes, 2003).

The Kruskal-Wallis analysis with evenness index of CAB showed that all the copepods genera have significantly different evenness (p-value: 0.0003 to 0.03) except *Centropages* spp. and *Pleuromamma* spp. ($p>0.97$) (Figure S1b). Note that, different genera of the copepods carry an uneven number of CAB species, as different copepod genera have different body volume (Datta et al., 2018). We also observed maximum faith phylogenetic genetic diversity (Faith_PD) index (52.00±35.66) in *Pleuromamma* spp. Both the *Calanus* spp. and *Centropages* spp. showed very less Faith_PD (19.9±6.3 and 13.3±3.02, respectively) (Figure S2). The gradient of the micro-environment (pH and O2 gradients) within the *Pleuromamma* spp. and its range of distribution in the water column may be reasoned for the observed maximum CAB phylogenetic diversity. All known 11 species of *Pleuromamma* (Goswami et al., 1994; Beaugrand et al., 2002) are well-known vertical migrators and have an important role in nutrient and carbon export from the shallow to innate mesopelagic waters (Steinberg et al., 2000).

The variation in faith_PD of CAB was assessed by Kruskal-Wallis test, which revealed that different copepod genera have highly significant and phylogenetically distinct bacteriobiome (Figure S2). Datta et al. (2018) identified 14% of OTUs (n=34) as core OTUs in 90% of individual *Calanus* spp. analyzed. Hence, defining copepod genera specific core OTUs would be an important task in understanding the phylogenetic distinctness of CAB.

3.4. Beta-Diversity

We hypothesize that if bacteriobiome were copepod type-specific, does phylogenetically closer copepod genera harbor phylogenetically close bacterial species diversity? To test this hypothesis, a consensus phylogram of 5 copepod genera was constructed and compared with the Unweighted, Weighted UniFrac and Jaccard distance matrix of CAB using PCoA plot. Phylogenetic relationships among the order Calanoida remains problematic mainly due to the wide range of morphological characteristics, widespread and overlapping geographical ranges and a sizeable magnitude of cryptic species complexity (Blanco-Bercial et al., 2014). We extracted 19 different *Acartia* spp., 9 different *Calanus* spp., 5 different *Centropages* spp., 6 different *Pleuromamma* spp., and 3 different *Temora* spp., sequences (Figure S3) for phylogram construction. The consensus phylogram revealed that *Calanus* spp. were phylogenetically closer to *Pleuromamma* spp. and form two distinct clusters. Whereas, rest of the genera were clustered into one cluster.

In the present study, beta-diversity (P-value 0.001) patterns and PERMANOVA analyses support the hypothesis that the CAB composition differed between and within copepod genera. As we closely investigate, Unweighted Unifrac distance matrix showed the CAB of *Pleuromamma* spp. and *Calanus* spp. separated into two different clusters (Figure 2a, b), whereas, the CAB of *Calanus* spp. was clustered into a single large cluster in a weighted distance matrix (Figure 2b). But in Jaccard distance matrix PCoA revealed *Calanus* spp. had three phylogenetic distinct CAB clusters (Figure 2c). Unweighted unifrac PCoA reveals that, *Pleuromamma* spp. and *Calanus* spp. has phylogenetically distinct CAB (Figure 2a and S3) with a variation of 6.318% in axis 1. This difference of CAB may be attributed to the difference in vertical migration and feeding behavior between the two genera. *Pleuromamma* spp. are known as omnivorous feeders (including phytoplankton, microzooplankton and

detritus) (Teuber et al., 2014; Cregene, 2016), and migrate vertically up to 1000m (Goswami et al., 1992; Beaugrand et al., 2002). Whereas, *Calanus* sp. are mostly herbivores feeders, but feeds on ciliates and other heterotrophic protists during lack of food availability and egg production (Nejstgaard et al., 2001) and adult *Calanus* spp. could migrate up to 600m (Irigoien, 1999). *Calanus carinatus* are known to tolerate low oxygen concentrations (<1ml l-1), and *Pleuromamma robusta* withstands hypoxic conditions (<0.8 ml l-1) in the Atlantic OMZ (Auel and Verheye, 2007).

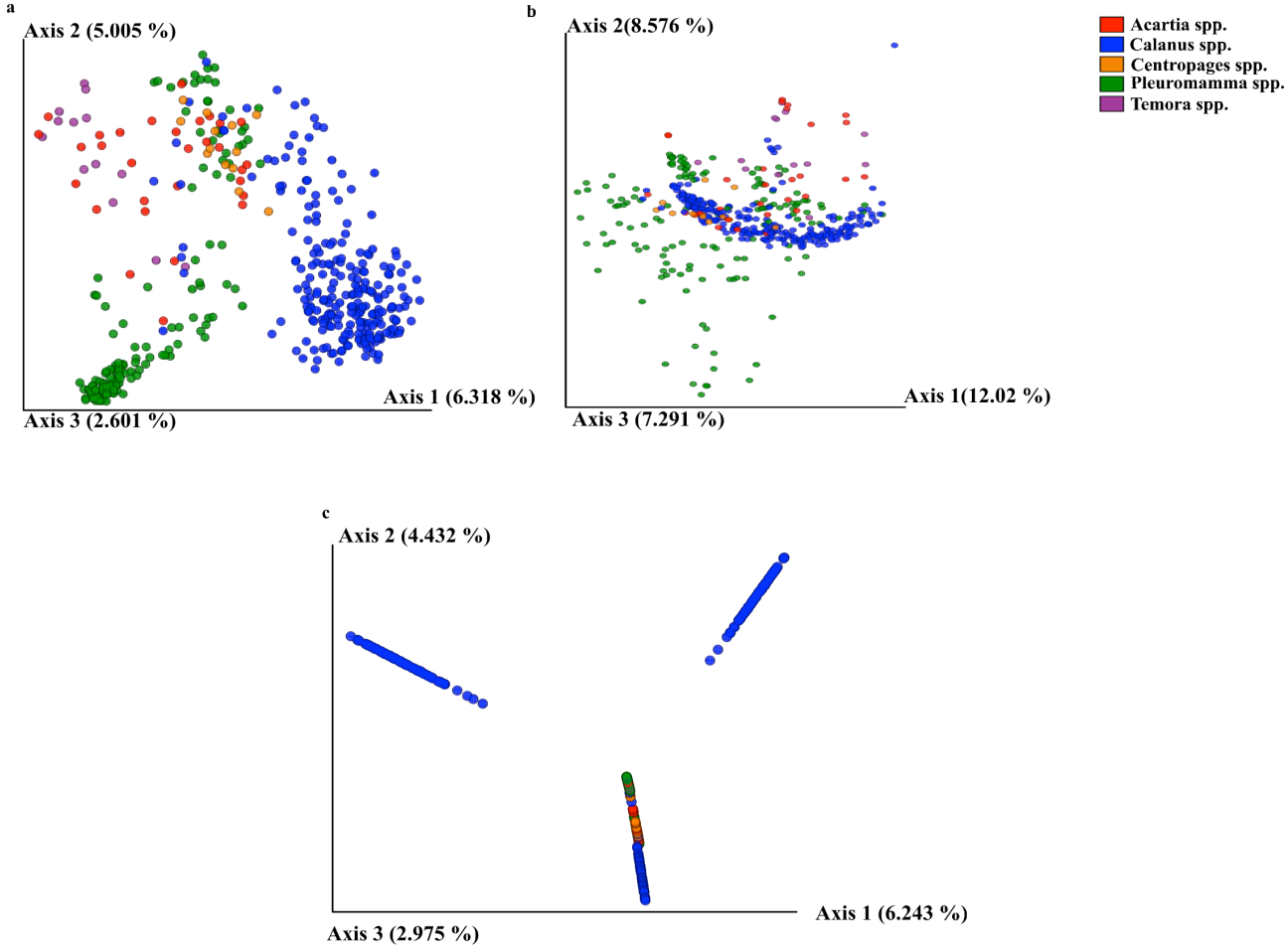

Figure.2 a) Unweighted Unifra distance matrix showed *Calanus* spp. and *Pleuromamma* spp. harbors phylogenetically distinct CAB and the CAB of other copepods genera were scattered on the plot b) Weighted unifrac distance matrics plot shows the *Pleurommama* spp. harbors phylogenetically distinct and diverse bacterial assemblages within the genera (green dots distributed in the plot) whereas *Calanus* spp. harbors phylogenetically conserved (relatively) groups of bacteria (blue dots; the middle portion of the plot) c) Jaccord distance-based beta-diversity reveals *Calanus* spp. and *Pleuromamma* spp. harbors distinct bacterial population. Nevertheless, they do share common bacterial groups with *Acartia* spp., *Centropages* spp., and *Temora* spp.

3.5. Differential abundance of CAB revealed through ANCOM

ANCOM results showed that a total of 23 bacterial phyla, viz., Cyanobacteria, Spirochaetes, Crenarchaeota, Firmicutes, GN02, Bacteroidetes, Proteobacteria, Planctomycetes, Actinobacteria, Acidobacteria, Euryarchaeota, Verrucomicrobia, WPS-2, Parvarchaeota, Thermi, TM6, Elusimicrobia, Fusobacteria, Chlorobi, Gemmatimonadetes, SBR1093, Chlamydiae and OD1 were significantly different between the copepod genera with W and F statistics ranged between 40 to 30 and 53 to 2.7, respectively (Supplementary File S1). The 23-bacterial phylum consists of 39 classes, 78 Order, 146 Family and 242 genera which were significantly different between the copepods (Supplementary File S2). We choose the top two percentile different genera (with W value of 809 and 808 and representative genera F-statistical value are given in supplementary File S2) to explain the percentile compositional difference of bacteriobiome between the copepod genera.

Bacterial taxa's like *Pseudomonas, Anaerospora,* Methylobacteriaceae*,* HTCC2207*,* Flavobacteriaceae*, Acinetobacter,* Bacteriovoracaceae and *Ochrobactrum* (F statistical value are given in supplementary File S2) were found high percentile in *Calanus* spp. (Figure 3). Prevalence of *Pseudomonas* and members of Methylobacteriaceae was also observed in *Pleuromamma* spp. (Cregene, 2016). Whereas Flavobacteriaceae was observed in low numbers in empty copepod guts, and its abundances increase with active feeding *Calanus finmarchicus* (Datta et al., 2018) and show the characteristic feature of surface dwellers. Also, *Sedinimicola* sp. (Flavobacteriaceae) was observed to be dominant in *Acartia* spp., *Temora* spp. and *Centropages* spp. (Moisander et al., 2015). Members of Bacteriovoracaceae known as a predatory bacterial group that regulate the populations of other bacteria in estuarine environments (Davidov & Jurkevitch, 2004).

In the present study, ANCOM showed that bacterial genera like *Paulinella, RS62, Candidatus portiera, Planktotalea, Segetibacter, Octadecabacter* and order Bacteroidales were found in high percentile in *Acartia* spp. (Figure 3). The copepod type and type of food ingested were known to influence the cultivable bacterial load in *Acartia* spp. (Tang 2005). In the case of *Centropages* spp. the bacterial genus like *Alteromonas, Pseudoalteromonas, Fluviicola, Oleispira, Ralstonia* and order Colwelliaceae and Cryomorphaceae percentile was found to be in high. Members of Oceanospirillales like *Pseudoalteromonas sp*. and Aletromonadaceae (Colwellia sp.) were known to be dominantly abundant in *Centropages* spp. (Moisander et al., 2015). Furthermore, the dominance of *Alteromonas* was observed in *Pleuromamma* spp. (Cregene, 2016). Moisander et al., (2015) reported that *Marinomonas* sp. (Gammaproteobacteria) was predominantly observed in *Centropages* spp. but it was not observed in our analysis. *Temora* spp. showed to have high percentile of *Comamonas, Planctomyces, Flavobacterium, Synechococcus, Chryseobacterium* and *Nitrosopumilus.* Only four genera like *Bradyrhizobium, Marinobacter, Photobacterium* and *Variovorax* were significantly high in *Pleuromamma* spp. (Figure 3).

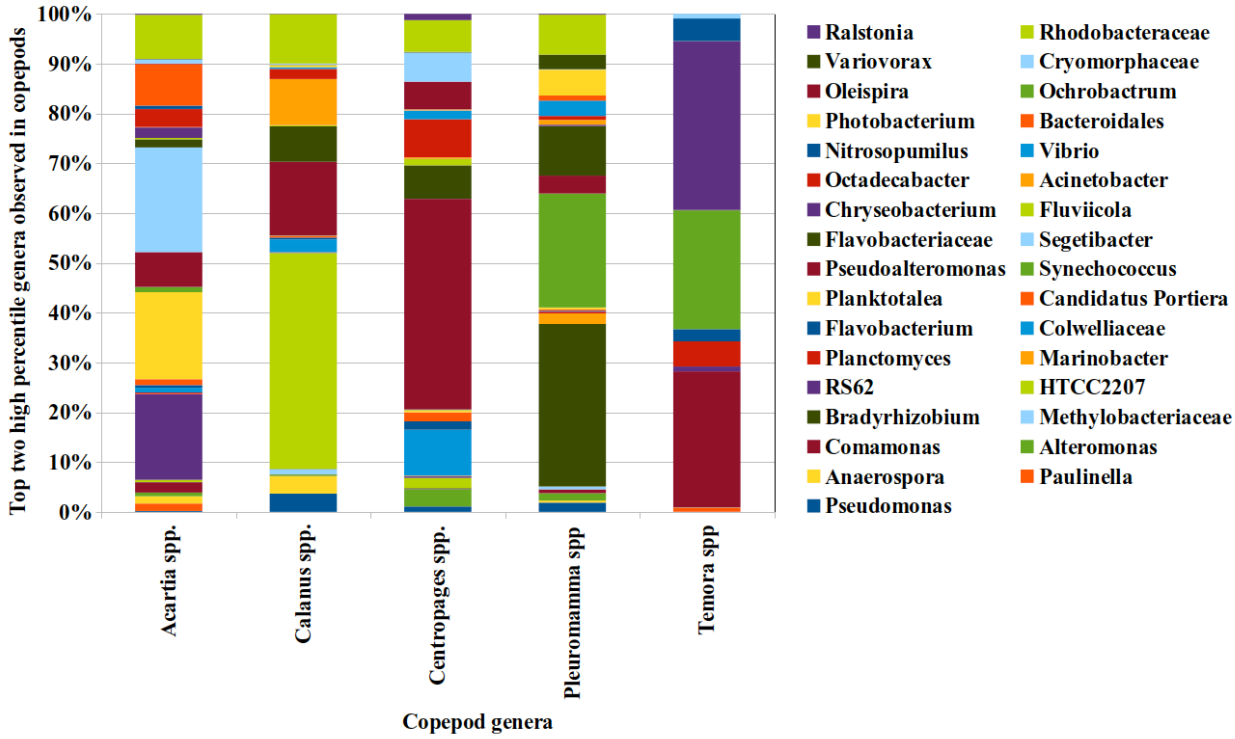

Figure 3: Top two genera with high percentile abundance observed in the 5 copepod genera using ANCOM.

3.6. Machine learning (RandomForest classifier) to predict important s-OTUs

The masking effect of the abundant bacterial community associated with copepod diet and ambient water column should not hinder the detection of core-OUTs, as evidenced from previous studies (Moisander et al., 2015: DeCorte et al., 2018; Wage et al., 2019; Datta et al., 2019). QIIME2 core_abundance algorithms used in the present study did not predict single bacterial s-OTUs (Data not presented). Hence, we use the machine learning Random Forest Classifier approaches to detect important core sub-OTUs specific to copepod genera.

Overall, the accuracy of the model was 0.956 and with the accuracy ratio of 1.69, indicating high reliability of the RandomForest classifier result. The accuracy of predicting important bacterial s-OTUs in copepod genera (Figure 4a) were in the range of 1 to 0.16 (Figure 4b). The graphical representation of machine learning model Receiver Operating Characteristic (ROC) curve (Figure 4c) was in ranging of 0.98 to 1, and it showed the high positive prediction rate and low rates of the false prediction. The prediction accuracy was found high in *Calanus* spp. and *Pleuromamma* spp. (AUC=1.00) (Figure 4c).

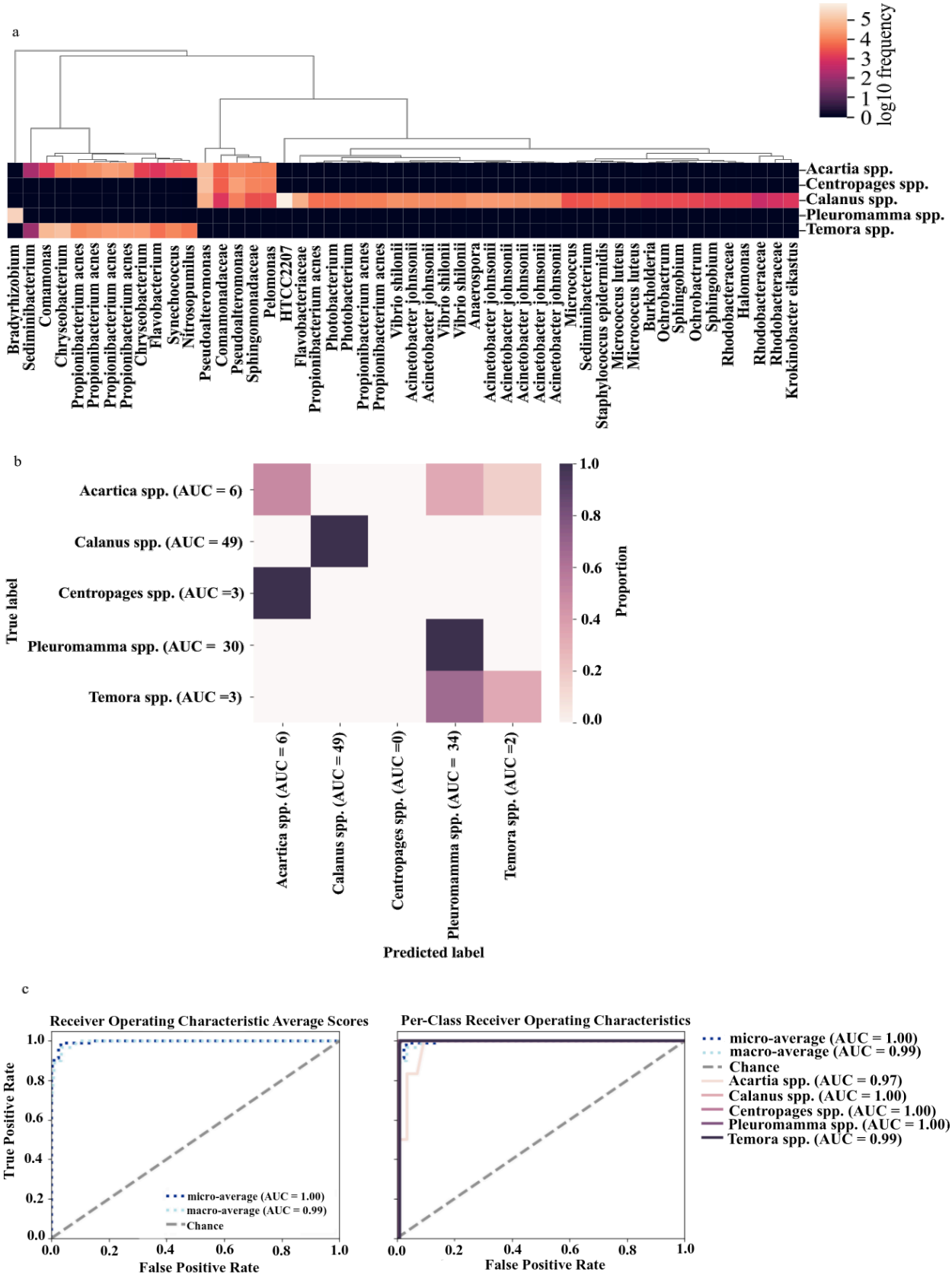

Figure.4: a) RandomForest classifier heatmap representing important microbial s-OTUs in five copepods genera, the colour scale indicates log10 frequency ("0" black to "5" pale). b) The overall prediction accuracy of RandomForest method represented in the confusion

matrix. c) Receiver Operating Characteristic (ROC) curves represent the classification accuracy of a machine-learning model. d) The area under the curve (AUC) indicates the better performance of RandomForest classifier.

Machine learning approach predicted 26 bacterial and one archaeal taxon in 5 copepod genera as important s-OTUs with differential hierarchical resolutions ranging from family to sub-OTUs (equivalent to subspecies or strains) level. It was evident that copepod genera had specific bacteria, not only at the species level but also in sub-species or strain level. A similar observation has made in phyla Nematoda and Annelida, i.e. their symbiotic sulfur-oxidizing bacteria (*Candidatus Thiosymbion*), showed coupled evolution along with their host (Zimmermann et al., 2016). Only *Calanus* spp. and *Pleuromamma* spp. found to have specific important s-OTUs, i.e. all s-OTUs of *Photobacterium, Micrococcus luteus*, three s-OTUs of *Vibrio shilonii* and all s-OTUs of *Acinetobacter johnsonii* were specific to *Calanus* spp. and one s-OTUs of *Bradyrhizobium* was predicted in *Pleuromamma* spp.. The unclassified genera of *Bradyrhizobiaceae* were significantly higher in *Centropages* sp. with full gut (Moisander et al., 2015). The *Bradyrhizobium* was known to have nifH gene, and this genus can be ruled out from core s-OTUs because they usually occur in seawater (Jayakumar & Ward, 2020). Specific important s-OTUs for other 3 genera of copepods was not evident. The *Synechococcus* (a free-living Cyanobacteria) genera abundance was influenced by the diet and even found after 24 hours in starved copepod gut. So, this OTUs can be ruled from the important s-OTUs (Moisander et al., 2015). Even though HTCC2207 (Gammaproteobacteria) was the most frequent predicted s-OTUs, their association as core OTUs could be ruled out. Because of their known proteorhodopsin gene and being free water living bacteria (Stingl et al., 2007), and hence the probability of detecting this bacteria in the copepod gut was highly due to food ingestion.

Among the 27 taxa detected by machine learning approach, 10 taxa's relative percentile was low in ANCOM analysis, which may be due to the masking effect of other abundant dominant taxa's. So, the machine learning approach adopted here was successful in picking rare but important s-OTUs. The 10-important s-OTUs belonged to *Micrococcus luteus, Sediminibacterium, Krokinobacter eikastus, Pelomonas, Vibrio shilonii, Acinetobacter johnsonii, Burkholderia, Sphingobium, Halomonas* and *Nitrosopumilus.*

Among that 10 s-OTUS, 5 OTUs were previously reported as important s-OTUs by earlier studies. Example; the present study observed *Sediminibacterium* as important s-OTUs in *Temora* spp. and *Acartia* spp. rather than *Pleuromamma* spp. However, even with low abundance of *Sediminibacterium* was regularly present in *Pleuromamma* spp. (Cargeen, 2016). *Halomonas* and *Pelomonas* were ruled out from core OTUs in *Calanus* spp. because it was also found in non-calanoid copepods (Datta et al., 2018). However, in the present analysis, the Proteobacterial genus *Pelomonas* was found to be an important s-OTUs in *Acartia* spp., *Calanus* spp., and *Centropages* spp.. Earlier, studies showed that the genus *Photobacterium* (Phylum: Proteobacteria) was abundant in *Pleuromamma* spp. (Cargeen, 2016), *Centropages* spp. (Moisander et al., 2015), *Calanus* spp., and non-calanoid species (Datta et al., 2018). Nevertheless, machine learning predicts the 2 s-OTUs of *Photobacterium* as an important s-OTUs only in *Calanus* spp. Even though the bacterial primers used rarely capture archaeal sequences, machine learning algorithm used here detected archaeal

sequences (*Nitrosopumilus*) as important s-OTUs in *Acratia* spp. and *Temora* spp. and this genus *Nitrosopumilus* was also reported to contribute 89 and 99 percentage on the overall community composition in *Acartia* spp. and *Temora* spp. (Wage et al., 2019). The *Pseudoalteromonas* was reported as a constant and stable OTU in *Acartia* sp*., Calanus* sp. and *Centropages* spp. (Wage et al., 2019), and the present RandomForest classification predict the same as important core s-OTUs in the same *Acartia* spp*., Calanus* spp. and *Centropages* spp.

Based on the present analysis the 6 s-OTUs viz., 1) *Microccocus luteus*, 2) *Krokinobacter eikastus*, 3) *Vibrio shilonii*, 4) *Acinetobacter johnsonii* and 5) *Burkholderia* and 6) *Sphingobium* were detected for the first time as important s-OTUs in copepods.

3.7. Principle component analysis reveals that copepod genera do host functionally distinct bacterial diversity.

The functional PCA plot clearly showed that the phylogenetic relationships among the CAB were grouped into four clusters (Figure 5). *Calanus* spp. was separated from the rest of the copepods genera with Principle Component (PC) value of 28.4% in axis 1 and 9.2% in axis 3, whereas, *Pleuromamma* spp. showed a variation of 28.4% in axis PC1 and 16.7% in PC2. *Centropages* spp. did not have unique CAB functional diversity, whereas, *Acartia* spp. and *Temora* spp. shared the common functional CABs.

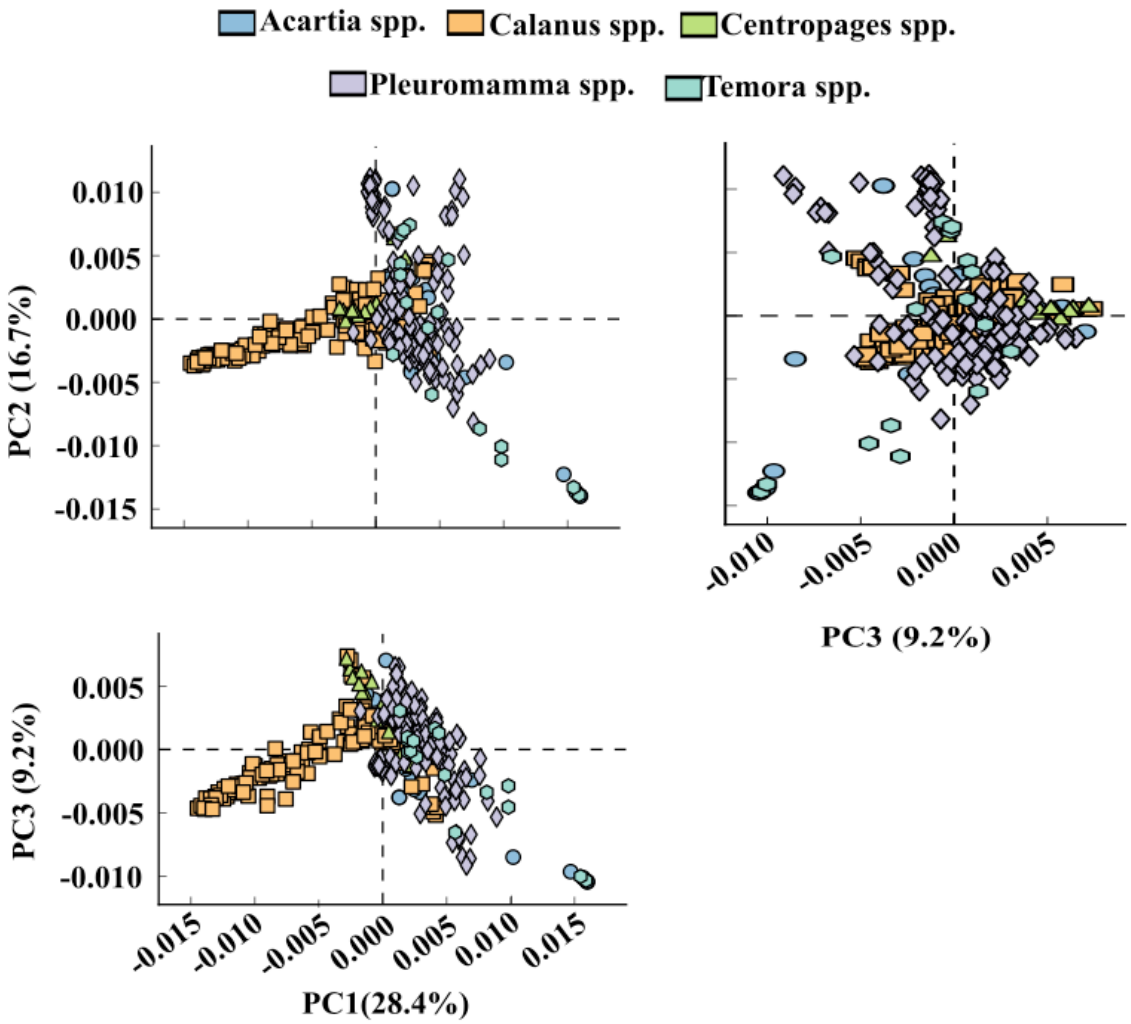

Figure 5: Overall functional diversity pattern observed among the copepod associated bacteria PCA.

3.8. Biogeochemical potentials of CAB

Bacterial communities exploit copepods as microhabitat by colonizing copepods' internal and external surfaces and mediate marine biogeochemical processes (De Corte et al., 2018). CAB also metabolize the complex organic compounds such as, chitin, taurine and other complex molecules in and around the copepod which could be a hot spot for the biogeochemical process (De Corte et al., 2018).

3.8.1. Potential methanogenesis by CAB: *Evidence of interlinking methanogenesis, DMSP degradation and phosphate utilization*

We observed methyl phosphonate, acetate, carbon dioxide, methylamine, and methanol, i.e. five major compounds that act as a substrate for methanogenesis (Yao et al.,2016; Evans et al., 2019). In the present analysis, we found that CAB has a complete set of aerobic methanogenesis genes (PhnL, M, J, H, G and mpnS) (Yao et al., 2016) which converts methylphosphonate (MPn) to methane (CH_4). Among the copepods, the CAB of *Pleuromamma* spp. and *Calanus* spp. had a relatively high proportion of MPn genes (Figure

S4), and the relative proportion significantly differ between the copepod genera (p values between 1.03e-08 to 1.78e-08), except for the gene mpnS (p=0.726) (Figure S4). Some copepods like *Acartia* sp. and *Temora* sp. were reported to have associate bacteria that involves in CH_4 production from MPn (Wage et al., 2019). CAB of *Pleuromamma* spp. could be a key player in potential MPn methanogenesis (Figure S4). Also, based on the present analysis *Pleuromamma* spp. CAB found to have a high relative proportion of genes (mtbC, mtbA, mttB) involve in the oxidation of Trimethylamine (TMA) to methyl-CoM (Figure S4) and mcrA gene (Figure S4). De Corte et al., (2018) suggested that different copepods species have different CAB, and only some copepods have specific CAB for methanogenesis and other biogeochemical cycles.

Early, *T. longicornis* fed with a high content of TMA/DMA phytoplankton's produce maximum amount of CH_4 and suggested the production was due to the micro-niches inside the copepods (Angelis &Lee, 1994). Instead of analyzing fecal pellets (Tang, 2001) and anaerobic incubation experiments (Ploug et al., 1997), further research should consider CAB mediated aerobic methanogenesis as one of the factors to solve the "Ocean methane paradox".

CAB of *Acartia* spp. and *Centropages* spp. contained high proportion of dmdA (Demethylation of DMSP) genes (p = <1e-15), whereas, *Temora* spp. holds the least (Figure. S5). But, the final step in CH_4 production by mtsA and mtsB genes were found abundant in *Pleuromamma* spp. (Figure S4). The taxons detected in the present study, like *Roseobacter* clade, SAR11 and Gammaproteobacteria are known to have dmdA genes (Howard et al.,2011, Varaljay et al., 2010). A previous study hypothesized the bacteria other than CAB as responsible for the methane build-up in the sub-thermocline layers of the central Baltic Sea (Stawiarski et al., 2019). However, in the present analysis showed that the CAB had potential dmdA gene which involves in CH_4 Production. Also, the methanogenic archeae like *Methanogenium organophilum, Methanolobus vulcani like sequences* and *Methanogenium organophilum* and *Methanobacterium bryantii* like sequences were noted in *Acartia clausi* and *Temora longicornis* fecal pellets (Ditchfield et al., 2012). Also, ^{14}C labeled experiment observed high methane production in *Temora longicornis* (Stawiarski et al., 2019). But, in the present study, we observed that *Pleuromamma* spp. could be a potential candidate to carry out archaeal methanogenesis with a high proportion of mcrA gene (Figure. S4).

3.8.2. Methanotrophic potential of CAB

In the present investigation, we found that the relative abundances of methanol dehydrogenases; mxaF and maxI genes were relatively high in *Pleuromamma* spp. with respect to other copepods (Figure S4). Even though, there is a lack of evidence for complete CH_4 utilization, CAB of *Pleuromamma* spp. have a high number of potential methanotrophic followed by CAB of *Calanus* spp.

3.8.3. Assimilatory sulfate reduction (ASR)

Based on our analysis, in all the copepod genera ASR pathway genes were predominant than the dissimilatory sulphate reduction (DSR) pathway genes. CAB of *Temora* spp. had a

higher number of sulfite reductase ferredoxin component (Figure S5a). Whereas, CAB of *Centropages* spp. has flavoprotein sulfite reductase gene in high proportions (Figure S5b). The relatively high abundance of genera like *Synechococcus* and Deltaproteobacterial family *Desulfovibrionaceae* (Supplementary File S3) in the CAB of *Temora* spp. may be responsible for the ASR pathway, as these genera are known to have ferredoxin-sulfite reductase activity.

3.9. Nitrogen fixation

We investigated the N_2-fixing potentiality of CAB by screening the abundances of nifH, nifD and nifK genes. *Pleuromamma* spp. had a higher proportion of nifH gene whereas *Temora* spp. had the least (Figure. S6). The abundance of nifH gene was found higher in full gut and starved *Acartia* spp. contributed by *Vibrio parahaemolyticus, V. cincinnatiensis* and unicellular cyanobacterium UCYN-A (Scavotto et al., 2015) and most bacteria with nifH gene are not genuine CAB (Scavotto et al. 2015). Also, the high abundance of *Bradyrhizobium* in *Pleuromamma* spp. (supplementary file) maybe the reason for the high percentile of nifH gene, which is present in the *Bradyrhizobium* genome. *Vibrio* attached to the exoskeleton, and gut lining of copepods (Rawlings et al., 2007) degrades chitin (Hirono et al., 1998; Meibom et al., 2004) and use this chitin as carbon and energy source for nitrogenase activity, which could give advantage for *Vibrio* spp. over non-cyanobacterial diazotrophs in nitrogen fixation (Moisander et al., 2012).

The abundance of nifH gene in the CAB of *Pleuromamma* spp. may be due to the presence of genera like *Synechococcus, Bradyrhizobium, Prochlorococcus, Microcystis***,** *Trichodesmium* and *Chroococcidiopsis***.** The previous study had also shown that *Pleuromamma*-gut has stable symbiotic cyanobacteria and Deltaproteobacteria (Cargeen, 2016).

3.9.1. Denitrification

3.9.1.1. Nitrate reductions; napA & napB

Gene involving in all the 3 steps of denitrification (nitrate reductions (napA and napB), nitrite reduction (nirK and nirS) and nitric oxide reduction (norB, C, D, Q)) were observed in all 5 copepod genera, whereas the relative proportions varied between them. The CAB of *Temora* spp. found to have a high proportion of potential denitrification genes, especially napA and napB genes, followed by *Pleuromamma* spp., *Acartia* spp., *Calanus* spp., and *Centropages* spp. (Figure S6). Moisander et al., (2018) reported the abundance of napA genes (similar to *Vibrio harveyi* and *V. campbellii*) in mixed copepods containing *Pleuromamma* sp., *Undinula vulgaris* and *Sapphirina* sp. The narG genes among the North Atlantic copepods were contributed by *Hahella ganghwensis* and *Alteromonas macleodii*.

3.9.1.2. Nitrite reduction; nirK and nirS

Among the nitrite reductase gene, we found the proportion of nirK gene to dominate nirS gene, in all the copepod genera (Figure S6). Furthermore, the proportion of nirK gene was high in *Acartia* spp. and *Temora* spp. Whereas, the proportion of nirS was high *Calanus* spp. and *Pleuromamma* spp..“Does feeding habit of copepods influence the denitrification process?” needs further investigation. Bacteria genera like *Pseudoalteromonas* and

Actinobacterium found in dead (sinking carcass) and live *Calanus finmarchicus* were reported to have nirS genes and known as a hotspot for denitrification (Glud et al., 2015).

3.9.1.3. Nitric oxide reductase; nor (B, C, D, Q)

The nor genes' presence was high in *Temora* spp., next to *Acratia* spp., while *Calanus* spp. and *Pleuromamma* spp. has an equal proportion of this gene. Whereas, in *Centropages* spp. we observed the least number of nor sequences and this nor genes are responsible for microaerobic bacterial growth (Mesa et al., 2002).

3.10. Anaerobic nitric oxide reduction

The norV (anaerobic nitric oxide reductase) and norW (flavorubredoxin reductase) genes sequences were high in CAB of *Pleuromamma* spp. compared to (of descending orders) *Centropages* spp., *Calanus* spp., *Acartia* spp. and least detected in *Temora* spp. (Figure S6). Interestingly, all the genes responsible for the anaerobic and microaerophilic biogeochemical process were found maximum in CAB of *Pleuromamma* spp.. which may play an important role in ocean anoxic biogeochemistry, and the membres of *Pleuromamma* genera are known to migrate hypoxic waters (Escribano et al., 2009; Teuber et al., 2013) contains a high abundance of norV and norW genes, the physiology (oxygen conditions) of the copepod gut condition may also favour the abundance of these genes.

3.10.1. Dissimilatory nitrate reduction into ammonia (DNRA)

In the previous analysis, the DNRA genes (narG, narI and narH) were observed in mixed copepod communities (De Corte et al., 2018). Whereas, the present analysis showed the high abundance of DNRA gene in *Acartia* spp. and *Temora* spp. followed by *Pleuromamma* spp. (Figure S6). The *Calanus* spp. and *Centropages* spp. had similar least relative proportions of DNRA genes.

3.11. Carbon processes

Phosphoenolpyruvate Carboxylase (PEPC) gene in CAB was related to its food intake (especially phytoplanktons). The PEPC gene was found to be equally distributed among the 5 copepods (Figure. S7a). The chitinase producing bacteria's like *Aeromonas, Erwinia, Chromobacterium, Flavobacterium, Arthrobacter, Serratia, Bacillus, Enterobacter,* and *Vibrio* are known to carbon mineralization like degradation and utilization of chitin (Donderski et al., 2000). The presence of chitinase gene in CAB is not surprising as their diet includes marine diatoms, which are known to have cell walls containing chitin (Teuber et al., 2014; Cregene, 2016). The CAB of *Centropages* spp. harbor high proportion of chitinase gene as compared to other copepods (Figure S7b) this may occur due to the feeding of ciliates or dinoflagellates by *Centropages* spp. (Calbet et al., 2007). The overall, outline of CAB mediated biogeochemical pathway is represented in Figure 6.

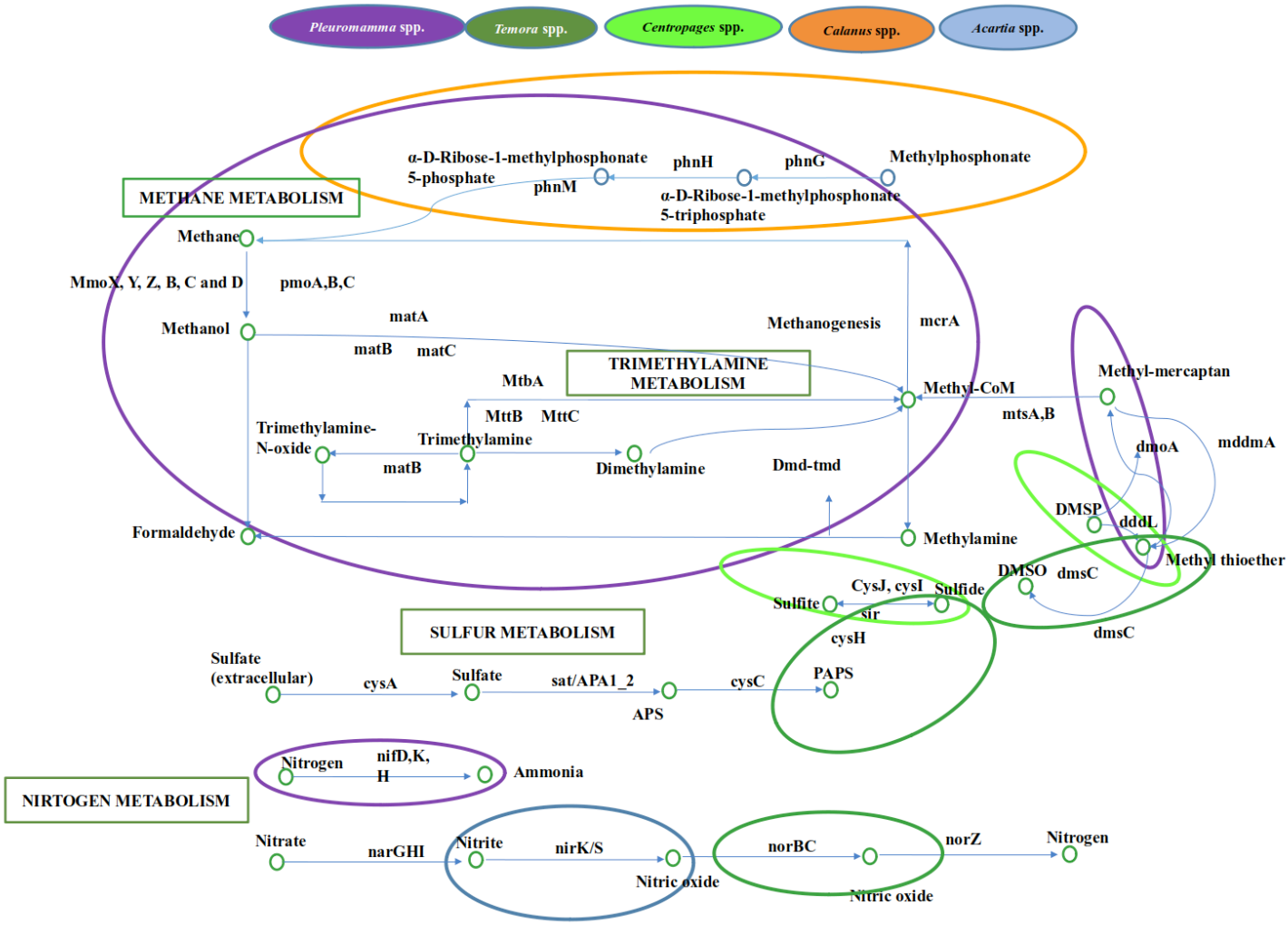

Figure.6. Overall representation of the biogeochemical potential of CAB. The circle and the color represent the copepod genera for that particular biogeochemical processes. Refer the text for the abbreviation of listed gene names.

3.12. Role of CAB in Iron fertilization

The key role of zooplankton, bacteria and viruses in supporting iron supply to the ocean biota is the emerging feature (Boyd et al., 2015). Several studies have documented regional and seasonal variation in the regeneration of iron in fueling phytoplankton carbon fixation (Boyd et al., 2015).

The meta-analysis revealed that the abundances of potential ferrous iron transport gene A (FeoA) and ferrous iron transport protein B (FeoB) were similarly distributed among the five copepod genera. While, *Temora* spp. was found to hold the largest proportion of potential ferrous iron transport proteins coding CAB, while the least proportion was observed in *Centropages* spp. (Figure S8a). Concerning ferric reduction, *Pleuromamma* spp. carries the largest proportion of ferric iron reductase gene (fhhF) gene (Figure S8b). The presence of the largest proportion of ferric iron reductase gene fhuF in *Pleuromamma* spp. needs detailed. Fe(II) may enhance the reduction of an intermediate (for example, NO_2−) which in turn enhance denitrification and DNRA processes (Michiels et al., 2017).

The acidic condition of zooplankton's digestive tract promotes iron recycling and solubilization by numerous microbial pathways (Tang et al., 2011; Schmidt et al., 2016). Thus increases the bioavailability of iron in the surrounding and promotes iron fertilization (Schmidt et al., 2016). The zooplankton-associated bacterial community (Bacteroidetes, Alphaproteobacteria and Gammaproteobacteria) are known to carry many genes involved in iron utilization, such as ferric reductase gene that encodes for an oxidoreductase to inter-convert ferric (Fe3+) and to ferrous (Fe2+) ion in *Calanus* sp. and *Paraeuchaeate* spp. (De Corte et al., 2018).

However, the differential iron contributions of different copepod genera were unknown until now. We hypothesis the different copepod genera have different bacteriobiome, that contribute to the ocean iron cycle differently and CAB community variation are due to multiple factors. The Ferric iron (Fe3+) mechanism was found to be dominant in an oxygenated environment, whereas ferrous iron (Fe2+) dominates the anaerobic conditions or at low pH (Lau et al., 2015). For organisms that must combat oxygen limitation for their survival (*Pleuromamma* spp.), pathways for the uptake of ferrous iron are essential. Several bacterial ferrous iron transport systems have been described; however, only the Feo system appears to be widely distributed and exclusively dedicated to the transport of iron. With this regard, we found CAB of *Pleuromamma* spp. to be a most significant contributor for iron fertilization. It has been shown that lower levels of nitrogen fixation in the South Atlantic are due to reduced iron availability (Moore et al., 2009). The meta-analysis demonstrated here showed *Pleuromamma* spp. could be a significant contributor to both nitrogen fixation and iron bioavailability.

3.13. CAB as a source of cyanocobalamine synthesizing prokaryotes

Organisms within all domains of life require the cofactor cobalamin (vitamin B12), which is produced only by a subset of bacteria and archaea (Doxey et al., 2015). We found that CAB could be one of the potential sources of cyanocobalamine production in the sea. Among the five genera analyzed, following were the descending order of genera based on their relative proportion of potential cobalamin synthesizing gene; *Temora* spp., *Acartia* spp., *Calanus* spp., *Pleuromamma* spp., and *Centropages* spp. (Figure. S9).

Previous studies reported that the cobalamin in ocean surface water is due to de nova synthesis by Thaumarchaeota and selective heterotrophic bacteria like *Sulfitobacter* sp. SA11 and *Ruegeria pomeroyi* DSS-3, *Methylophaga* and *Marinobacter* (Doxey et al., 2015). But, in the present study, CAB of *Temora* spp. had high proportions of cobalamine synthesis gene and (Thaumarchaeota) genus *Nitrosopumilus*. About 94% of Alphaproteobacteria, Gammaproteobacteria and Thaumarchaeota genomes have the cobalamin synthesizing and activation gene (Doxey et al., 2015).

The limitation of the present study could be related to the fact that all CAB sequences were from the Atlantic Ocean. Copepods genera from other different ocean may contain different CAB diversity (Datta et al., 2018). In this regard, further studies on CAB diversity from different ocean realms would throw the actual potential of CAB in the global biogeochemical cycle. Also, since oxygen minimum zone is globally increasing (see Stramma et al., 2011) and few copepod species such as *Pleuromamma robusta, Calanoides carinatus* and *Rhincalanus nasutus* were known to navigate to OMZ (Auel & Verheye,

2007), exploring the CAB diversity in OMZ of the Arabian Sea and the Pacific Ocean could expand our understating of mechanisms behind OMZ-copepod survival and varying their biogeochemical processes in deep migrating copepods.

Conclusion

We predicted 27 bacterial taxa (+1 archea) in 5 copepod genera using Machine learning approach as important s-OTUs. Among the predicted bacterial genera *Micrococcus luteus, Krokinobacter eikastus, Vibrio shilonii, Acinetobacter johnsonii, Burkholderia,* and *Sphingobium* were reported as important s-OTUs in copepods for the first time as per our knowledge. It is evident that the specific bacterial s-OTUs do exists for copepod genera, not only at the species level but also in sub-species or strain level.

A meta-analysis revealed that CAB was capable of mediating methanogenesis (with evidence of interlinking the methane production, DMSP degradation and phosphate utilization) and methane oxidation. We also found that CAB had more potential assimilatory sulphur reducing microbial community than the dissimilatory sulfate reduction. Likewise, CAB found to have potential gene involving in nitrogen fixation, denitrification, anammox, dissimilatory nitrate reduction into ammonia. We also found CAB is also carrying potential genes that perform carbon fixation, carbon mineralization, iron fertilization and vitamin B12 synthesis. Future studies should also consider the CAB as one of the factors in marine biogeochemical and climate modeling.

Acknowledgement

The authors thank the Director, CSIR-NIO, for encouraging this work. BS, PC, UVN and MG received the financial assistance from the Council of Scientific & Industrial Research, Government of India, under projects MLP1802 is gratefully acknowledged. We also thank the High-performance computing facility "Pravah" to carry out the bioinformatics work. This is NIO's contribution No___. MS is funded by the EPSRC and Imperial College London (EP/N509486/1: 1,979,819). We thank our funders.

References

1) Anderson, M. J. (2017). Permutational Multivariate Analysis of Variance (PERMANOVA). In Wiley StatsRef: Statistics Reference Online (pp. 1–15). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118445112.stat07841.
2) Auel, H., & Verheye, H. M. (2007). Hypoxia tolerance in the copepod Calanoides carinatus and the effect of an intermediate oxygen minimum layer on copepod vertical distribution in the northern Benguela Current upwelling system and the Angola–Benguela Front. Journal of Experimental Marine Biology and Ecology, 352(1), 234–243. https://doi.org/10.1016/j.jembe.2007.07.020.
3) Beaugrand, G., Ibañez, F., Lindley, J., Philip, C., & Reid, P. (2002). Diversity of calanoid copepods in the North Atlantic and adjacent seas: species associations and biogeography. Marine Ecology Progress Series, 232, 179–195. https://doi.org/10.3354/meps232179.
4) Blanco-Bercial, L., Cornils, A., Copley, N., & Bucklin, A. (2014). DNA barcoding of marine copepods: assessment of analytical approaches to species identification. PLoS

currents,6,ecurrents.tol.cdf8b74881f87e3b01d56b43791626d2.https://doi.org/10.1371/currents.tol.cdf8b74881f87e3b01d56b43791626d2.

5) Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology, 37(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9.
6) Boyd, P. W., Strzepek, R. F., Ellwood, M. J., Hutchins, D. A., Nodder, S. D., Twining, B. S., & Wilhelm, S. W. (2015). Why are biotic iron pools uniform across high- and low-iron pelagic ecosystems? Global Biogeochemical Cycles, 29(7), 1028–1043. https://doi.org/10.1002/2014gb005014.
7) Breiman, L. (2001). Machine Learning, 45(1), 5–32. https://doi.org/10.1023/a:1010933404324.
8) Brown, J., Pirrung, M., & McCue, L. A. (2017). FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. Bioinformatics, 33(19), 3137–3139. https://doi.org/10.1093/bioinformatics/btx373.
9) Calbet, A., Carlotti, F., & Gaudy, R. (2007). The feeding ecology of the copepod Centropages typicus (Kröyer). Progress in Oceanography, 72(2–3), 137–150. https://doi.org/10.1016/j.pocean.2007.01.003.

10) Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods, 13(7), 581–583. https://doi.org/10.1038/nmeth.3869.
11) Cregeen, S.J.J. (2016). Microbiota of dominant Atlantic copepods: *Pleuromamma* sp. as a host to a betaproteobacterial symbiont. Ph.D., Thesis, University of Southampton, pp-1-183.
12) Dam, H. G., & Lopes, R. M. (2003). Omnivory in the calanoid copepod *Temora longicornis*: feeding, egg production and egg hatching rates. Journal of Experimental Marine Biology and Ecology, 292(2), 119–137. https://doi.org/10.1016/s0022-0981(03)00162-x.
13) Datta, M. S., Almada, A. A., Baumgartner, M. F., Mincer, T. J., Tarrant, A. M., & Polz, M. F. (2018). Inter-individual variability in copepod microbiomes reveals bacterial networks linked to host physiology. The ISME Journal, 12(9), 2103–2113. https://doi.org/10.1038/s41396-018-0182-1.
14) Davidov, Y., & Jurkevitch, E. (2004). Diversity and evolution of *Bdellovibrio*-and-like organisms (BALOs), reclassification of *Bacteriovorax starrii* as *Peredibacter starrii* gen. nov., comb. nov., and description of the Bacteriovorax–Peredibacter clade as Bacteriovoracaceae fam. nov. International Journal of Systematic and Evolutionary Microbiology, 54(5), 1439–1452. https://doi.org/10.1099/ijs.0.02978-0.
15) de Angelis, M. A., & Lee, C. (1994). Methane production during zooplankton grazing on marine phytoplankton. Limnology and Oceanography, 39(6), 1298–1308. https://doi.org/10.4319/lo.1994.39.6.1298.
16) De Corte, D., Lekunberri, I., Sintes, E., Garcia, J., Gonzales, S., & Herndl, G. (2014). Linkage between copepods and bacteria in the North Atlantic Ocean. Aquatic Microbial Ecology, 72(3), 215–225. https://doi.org/10.3354/ame01696.0.
17) De Corte, D., Srivastava, A., Koski, M., Garcia, J. A. L., Takaki, Y., Yokokawa, T., Nunoura, T., Elisabeth, N. H., Sintes, E., & Herndl, G. J. (2017). Metagenomic insights into zooplankton-associated bacterial communities. Environmental Microbiology, 20(2), 492–505. https://doi.org/10.1111/1462-2920.13944.

18) Ditchfield, A., Wilson, S., Hart, M., Purdy, K., Green, D., & Hatton, A. (2012). Identification of putative methylotrophic and hydrogenotrophic methanogens within sedimenting material and copepod faecal pellets. Aquatic Microbial Ecology, 67(2), 151–160. https://doi.org/10.3354/ame01585.
19) Donderski, W., & Trzebiatowska, M. (2000). Influence of physical and chemical factors on the activity of chitinases produced by planktonic bacteria isolated from Jeziorak Lake. Polish Journal of Environmental Studies, 9(2), 77–82.
20) Dong, Y., Yang, G.-P., & Tang, K. W. (2013). Dietary effects on abundance and carbon utilization ability of DMSP-consuming bacteria associated with the copepod *Acartia tonsa* Dana. Marine Biology Research, 9(8), 809–814. https://doi.org/10.1080/17451000.2013.765587.
21) Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., Huttenhower, C., & Langille, M. G. I. (2020). PICRUSt2 for prediction of metagenome functions. Nature Biotechnology, 38(6), 685–688. https://doi.org/10.1038/s41587-020-0548-6.
22) Doxey, A. C., Kurtz, D. A., Lynch, M. D., Sauder, L. A., & Neufeld, J. D. (2015). Aquatic metagenomes implicate Thaumarchaeota in global cobalamin production. The ISME journal, 9(2), 461–471. https://doi.org/10.1038/ismej.2014.142.
23) Escribano, R., Hidalgo, P., & Krautz, C. (2009). Zooplankton associated with the oxygen minimum zone system in the northern upwelling region of Chile during March 2000. Deep Sea Research Part II: Topical Studies in Oceanography, 56(16), 1083–1094. https://doi.org/10.1016/j.dsr2.2008.09.009.
24) Evans, P. N., Boyd, J. A., Leu, A. O., Woodcroft, B. J., Parks, D. H., Hugenholtz, P., & Tyson, G. W. (2019). An evolving view of methane metabolism in the Archaea. Nature Reviews Microbiology, 17(4), 219–232. https://doi.org/10.1038/s41579-018-0136-7.
25) Ferrer R. L. (1998). Graphical methods for detecting bias in meta-analysis. Family medicine, 30(8), 579–583.
26) Gilbert, J. A., Blaser, M. J., Caporaso, J. G., Jansson, J. K., Lynch, S. V., & Knight, R. (2018). Current understanding of the human microbiome. Nature Medicine, 24(4), 392–400. https://doi.org/10.1038/nm.4517.
27) Glud, R. N., Grossart, H.-P., Larsen, M., Tang, K. W., Arendt, K. E., Rysgaard, S., Thamdrup, B., & Gissel Nielsen, T. (2015). Copepod carcasses as microbial hot spots for pelagic denitrification. Limnology and Oceanography, 60(6), 2026–2036. https://doi.org/10.1002/lno.10149.
28) Goswami, S.C., (1994). Distribution of *Pleuromamma* spp. (Copepoda-Calanoida) in the northern Arabian Sea. Indian Journal Marine Science, 23, 178–179.
29) Hall TA (1999). "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT". Nucl. Acids. Symp. Ser. 41: 95-98.
30) Heidelberg, J. F., Heidelberg, K. B., & Colwell, R. R. (2002). Bacteria of the gamma-subclass Proteobacteria associated with zooplankton in Chesapeake Bay. Applied and environmental microbiology, 68(11), 5498–5507. https://doi.org/10.1128/aem.68.11.5498-5507.2002.
31) Howard, E. C., Sun, S., Biers, E. J., & Moran, M. A. (2008). Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. *Environmental microbiology*, *10*(9), 2397–2410. https://doi.org/10.1111/j.1462-2920.2008.01665.x.
32) Irigoien, X. (2000). Vertical distribution and population structure of *Calanus finmarchicus* at station India (59°N, 19°W) during the passage of the great salinity anomaly, 1971–1975. Deep Sea Research Part I: Oceanographic Research Papers, 47(1), 1–26. https://doi.org/10.1016/s0967-0637(99)00045-x.

33) Janssen, S., McDonald, D., Gonzalez, A., Navas-Molina, J. A., Jiang, L., Xu, Z. Z., Winker, K., Kado, D. M., Orwoll, E., Manary, M., Mirarab, S., & Knight, R. (2018). Phylogenetic Placement of Exact Amplicon Sequences Improves Associations with Clinical Information. mSystems, 3(3). https://doi.org/10.1128/msystems.00021-18.
34) Jayakumar, A., & Ward, B. B. (2020). Diversity and distribution of Nitrogen Fixation Genes in the Oxygen Minimum Zones of the World Oceans. Copernicus GmbH. https://doi.org/10.5194/bg-2019-445.
35) Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association, 47(260), 583–621. https://doi.org/10.1080/01621459.1952.10483441.
36) Lau, C. K. Y., Krewulak, K. D., & Vogel, H. J. (2015). Bacterial ferrous iron transport: the Feo system. FEMS Microbiology Reviews, 40(2), 273–298. https://doi.org/10.1093/femsre/fuv049.
37) Lau, C. K. Y., Krewulak, K. D., & Vogel, H. J. (2015). Bacterial ferrous iron transport: the Feo system. FEMS Microbiology Reviews, 40(2), 273–298. https://doi.org/10.1093/femsre/fuv049.
38) Mandal, S., Van Treuren, W., White, R. A., Eggesbø, M., Knight, R., & Peddada, S. D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. Microbial Ecology in Health & Disease, 26(0). https://doi.org/10.3402/mehd.v26.27663.
39) Marchesi, J. R., & Ravel, J. (2015). The vocabulary of microbiome research: a proposal. Microbiome, 3(1). https://doi.org/10.1186/s40168-015-0094-5.
40) Mark Moore, C., Mills, M. M., Achterberg, E. P., Geider, R. J., LaRoche, J., Lucas, M. I., McDonagh, E. L., Pan, X., Poulton, A. J., Rijkenberg, M. J. A., Suggett, D. J., Ussher, S. J., & Woodward, E. M. S. (2009). Large-scale distribution of Atlantic nitrogen fixation controlled by iron availability. Nature Geoscience, 2(12), 867–871. https://doi.org/10.1038/ngeo667.
41) Mesa, S., Velasco, L., Manzanera, M. E., Delgado, M. J., & Bedmar, E. J. (2002). Characterization of the norCBQD genes, encoding nitric oxide reductase, in the nitrogen fixing bacterium *Bradyrhizobium japonicum* b bThe GenBank accession number for the *B. japonicum* norCBQD genes reported in this paper is AJ132911. Microbiology, 148(11), 3553–3560. https://doi.org/10.1099/00221287-148-11-3553.
42) Michiels, C. C., Darchambeau, F., Roland, F. A. E., Morana, C., Llirós, M., García-Armisen, T., Thamdrup, B., Borges, A. V., Canfield, D. E., Servais, P., Descy, J.-P., & Crowe, S. A. (2017). Iron-dependent nitrogen cycling in a ferruginous lake and the nutrient status of Proterozoic oceans. Nature Geoscience, 10(3), 217–221. https://doi.org/10.1038/ngeo2886.
43) Moisander, P. H., Sexton, A. D., & Daley, M. C. (2015). Stable Associations Masked by Temporal Variability in the Marine Copepod Microbiome. PLOS ONE, 10(9), e0138967. https://doi.org/10.1371/journal.pone.0138967.
44) Moisander, P. H., Shoemaker, K. M., Daley, M. C., McCliment, E., Larkum, J., & Altabet, M. A. (2018). Copepod-Associated Gammaproteobacteria Respire Nitrate in the Open Ocean Surface Layers. Frontiers in Microbiology, 9. https://doi.org/10.3389/fmicb.2018.02390.
45) Møller, E.F., Riemann, L., & Søndergaard, M. (2007). Bacteria associated with copepods: abundance, activity and community composition. Aquatic Microbial Ecology, 47, 99–106.
46) Nejstgaard, J., Naustvoll, L., & Sazhin, A. (2001). Correcting for underestimation of microzooplankton grazing in bottle incubation experiments with mesozooplankton. Marine Ecology Progress Series, 221, 59–75. https://doi.org/10.3354/meps22105.

47) Ohman, M. D., & Runge, J. A. (1994). Sustained fecundity when phytoplankton resources are in short supply: Omnivory by *Calanus finmarchicus* in the Gulf of St. Lawrence. Limnology and Oceanography, 39(1), 21–36. https://doi.org/10.4319/lo.1994.39.1.0021.
48) Parks, D. H., Tyson, G. W., Hugenholtz, P., & Beiko, R. G. (2014). STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics (Oxford, England), 30(21), 3123–3124. https://doi.org/10.1093/bioinformatics/btu494.
49) Ploug, H., Kühl, M., Buchholz-Cleven, B., & Jørgensen, B. (1997). Anoxic aggregates - an ephemeral phenomenon in the pelagic environment? Aquatic Microbial Ecology, 13, 285–294. https://doi.org/10.3354/ame013285.
50) Proctor, LM (1997). Nitrogen-fixing, photosynthetic, anaerobic bacteria associated with pelagic copepods. Aquatic Microbial Ecology, 12, 105–113.
51) Rawlings, T. K., Ruiz, G. M., & Colwell, R. R. (2007). Association of *Vibrio cholerae* O1 El Tor and O139 Bengal with the Copepods Acartia tonsa and Eurytemora affinis. Applied and Environmental Microbiology, 73(24), 7926–7933. https://doi.org/10.1128/aem.01238-07.
52) Rocca, J. D., Simonin, M., Blaszczak, J. R., Ernakovich, J. G., Gibbons, S. M., Midani, F. S., & Washburne, A. D. (2019). The Microbiome Stress Project: Toward a Global Meta-Analysis of Environmental Stressors and Their Effects on Microbial Communities. Frontiers in Microbiology, 9. https://doi.org/10.3389/fmicb.2018.03272.
53) Saiz, E., Calbet, A., Atienza, D., & Alcaraz, M. (2007). Feeding and production of zooplankton in the Catalan Sea (NW Mediterranean). Progress in Oceanography, 74(2–3), 313–328. https://doi.org/10.1016/j.pocean.2007.04.004.
54) Scavotto, R. E., Dziallas, C., Bentzon-Tilia, M., Riemann, L., & Moisander, P. H. (2015). Nitrogen-fixing bacteria associated with copepods in coastal waters of the North Atlantic Ocean. Environmental Microbiology, 17(10), 3754–3765. https://doi.org/10.1111/1462-2920.12777.
55) Scavotto, R. E., Dziallas, C., Bentzon-Tilia, M., Riemann, L., & Moisander, P. H. (2015). Nitrogen-fixing bacteria associated with copepods in coastal waters of the North Atlantic Ocean. Environmental Microbiology, 17(10), 3754–3765. https://doi.org/10.1111/1462-2920.12777.
56) Schmidt, K., Schlosser, C., Atkinson, A., Fielding, S., Venables, H. J., Waluda, C. M., & Achterberg, E. P. (2016). Zooplankton Gut Passage Mobilizes Lithogenic Iron for Ocean Productivity. Current Biology, 26(19), 2667–2673. https://doi.org/10.1016/j.cub.2016.07.058.
57) Shoemaker, K. M., & Moisander, P. H. (2015). Microbial diversity associated with copepods in the North Atlantic subtropical gyre. FEMS Microbiology Ecology, 91(7). https://doi.org/10.1093/femsec/fiv064.
58) Shoemaker, K. M., & Moisander, P. H. (2017). Seasonal variation in the copepod gut microbiome in the subtropical North Atlantic Ocean. Environmental Microbiology, 19(8), 3087–3097. https://doi.org/10.1111/1462-2920.13780.
59) Stawiarski, B., Otto, S., Thiel, V., Gräwe, U., Loick-Wilde, N., Wittenborn, A. K., … Schmale, O. (2019). Controls on zooplankton methane production in the central Baltic Sea. Biogeosciences, 16(1), 1–16. https://doi.org/10.5194/bg-16-1-2019.
60) Steinberg, D. K., Carlson, C. A., Bates, N. R., Goldthwait, S. A., Madin, L. P., & Michaels, A. F. (2000). Zooplankton vertical migration and the active transport of dissolved organic and inorganic carbon in the Sargasso Sea. Deep Sea Research Part I: Oceanographic Research Papers, 47(1), 137–158. https://doi.org/10.1016/s0967-0637(99)00052-7.

61) Stingl, U., Desiderio, R. A., Cho, J. C., Vergin, K. L., & Giovannoni, S. J. (2007). The SAR92 clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. Applied and environmental microbiology, 73(7), 2290–2296. https://doi.org/10.1128/AEM.02559-06.
62) Stramma, L., Prince, E. D., Schmidtko, S., Luo, J., Hoolihan, J. P., Visbeck, M., Wallace, D. W. R., Brandt, P., & Körtzinger, A. (2011). Expansion of oxygen minimum zones may reduce available habitat for tropical pelagic fishes. Nature Climate Change, 2(1), 33–37. https://doi.org/10.1038/nclimate1304.
63) Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Molecular Biology and Evolution, 24(8), 1596–1599. https://doi.org/10.1093/molbev/msm092.
64) Tang, K. (2005). Copepods as microbial hotspots in the ocean: effects of host feeding activities on attached bacteria. Aquatic Microbial Ecology, 38, 31–40. https://doi.org/10.3354/ame038031.
65) Tang, K. W., Glud, R. N., Glud, A., Rysgaard, S., & Nielsen, T. G. (2011). Copepod guts as biogeochemical hotspots in the sea: Evidence from microelectrode profiling ofCalanusspp. Limnology and Oceanography, 56(2), 666–672. https://doi.org/10.4319/lo.2011.56.2.0666.
66) Tang, K. W., Visscher, P. T., & Dam, H. G. (2001). DMSP-consuming bacteria associated with the calanoid copepod *Acartia tonsa* (Dana). Journal of experimental marine biology and ecology, 256(2), 185–198. https://doi.org/10.1016/s0022-0981(00)00314-2.
67) Teuber, L., Schukat, A., Hagen, W., & Auel, H. (2013). Distribution and Ecophysiology of Calanoid Copepods in Relation to the Oxygen Minimum Zone in the Eastern Tropical Atlantic. PLoS ONE, 8(11), e77590. https://doi.org/10.1371/journal.pone.0077590.
68) Tukey–Kramer Method. (2013). In Encyclopedia of Systems Biology (pp. 2304–2304). Springer New York. https://doi.org/10.1007/978-1-4419-9863-7_101575.
69) Varaljay, V. A., Howard, E. C., Sun, S., & Moran, M. A. (2010). Deep sequencing of a dimethylsulfoniopropionate-degrading gene (dmdA) by using PCR primer pairs designed on the basis of marine metagenomic data. Applied and environmental microbiology, 76(2), 609–617. https://doi.org/10.1128/AEM.01258-09.
70) Vehmaa, A., Hogfors, H., Gorokhova, E., Brutemark, A., Holmborn, T., & Engström-Öst, J. (2013). Projected marine climate change: effects on copepod oxidative status and reproduction. Ecology and Evolution, 3(13), 4548–4557. https://doi.org/10.1002/ece3.839.
71) Wäge, J., Strassert, J. F. H., Landsberger, A., Loick-Wilde, N., Schmale, O., Stawiarski, B., … Labrenz, M. (2019). Microcapillary sampling of Baltic Sea copepod gut microbiomes indicates high variability among individuals and the potential for methane production. FEMS Microbiology Ecology, 95(4). https://doi.org/10.1093/femsec/fiz024.
72) Wirbel, J., Pyl, P. T., Kartal, E., Zych, K., Kashani, A., Milanese, A., … Zeller, G. (2019). Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. Nature Medicine, 25(4), 679–689. https://doi.org/10.1038/s41591-019-0406-6.
73) Yao, M., Henny, C., & Maresca, J. A. (2016). Freshwater Bacteria Release Methane as a By-Product of Phosphorus Acquisition. Applied and Environmental Microbiology, 82(23), 6994–7003. https://doi.org/10.1128/aem.02399-16.
74) Zimmermann, J., Wentrup, C., Sadowski, M., Blazejak, A., Gruber-Vodicka, H. R., Kleiner, M., Ott, J. A., Cronholm, B., De Wit, P., Erséus, C., & Dubilier, N. (2016).

Closely coupled evolutionary history of ecto- and endosymbionts from two distantly related animal phyla. Molecular ecology, 25(13), 3203–3223. https://doi.org/10.1111/mec.13554.

Table. 1: Details of number of Illumina files, sequences extracted, quality filtered (Phred score <25) analyzed was tabulated. RP indicate "relative proportion"

Species	**No. of files**	**RP of files (%)**	**Gross Sequences**	**RP of grs. seq. (%)**	**Net. no. of sequences after QC**	**RP after QC (%)**	**No. of OTUs**	**RP of OUT (%)**	**Number of seq lost in QC**	**RP of los (%)**
Acartia	30	6.6	2567759	15.6	2274402	16.3	1943032	16.5	293357	1.8
Calanus	244	53.9	6564419	39.7	5911821	42.2	5255849	44.7	946562	4.1
Pleuromamma	143	32.8	4310670	26.1	3020608	21.6	2995684	25.5	1290062	7.8
Centrophages	13	2.8	886314	5.3	875987	6.3	837567	7.1	10327	0.1
Temora	16	3.5	2498614	15.1	2223308	15.8	739971	6.3	275306	1.7
Total	452		16509304		13987186		11747127		2522118 (15.3%)	15.5

931 Table S1. List of sequence libraries representing the copepods associated bacteribiome. Out
932 of these only 7 libraries (highlighted in red font) where analysed in this study.

933

S. No	NCBI BioProject No	Species name	16S rDNA region	Sequencing platform	Reference
1	PRJNA383099	Details not available	Details not available	Illumina MiSeq	No
2	PRJEB23400	*Pleuromamma* sp.	V3-V4	Illumina	No
3	PRJNA416766	*Acartica* sp. and *Temora* sp.	V3-V4 & V4-V5 (archaea)	Illumina MiSeq	Wage et al., (2019)
4	PRJNA341063	*Calanus* sp.	V3-V4	Illumina MiSeq	Shoemaker and Moisander, (2017)
5	PRJNA285993	*Acartica* sp. *Centrophage* sp. and *Temora* sp.	V3-V4	Illumina MiSeq	Moisander et al., (2015)
6	PRJEB8785	*Acartia tonsa* and *Centropages hamatus*	Details not available	454/FLX-based	No
7	PRJNA248671	*Undinula vulgaris, Pleuromamma* spp., *Sapphirina metalina*, *Pseudocalanus* spp. and *Tigriopus* sp..	V5–V9	454 GS FLX Titanium	De Corte et al., (2018)
8	PRJEB14826	*Acartia tonsa* and *Temora longicornis*	V3-V4	Illumina MiSeq	Moisander et al., (2018)

9	PRJNA322089	*C. fimaarchincus*	V4	Illumina MiSeq	No
10	PRJDB5552	*Calanus* sp., *Paraeuchaeta* sp., *Themisto* sp., *Evadne* sp., and *Oncaea* sp.	V3-V4	Illumina MiSeq	No
11	PRJNA433804	*Spaniomolgus* sp.	V4-V5	Ion_Torrent	No

934

935