Baseline information on prokaryotic and microeukaryotic plankton communities inside and outside of Indonesian marine lakes

D.F.R. Cleary, N.C.M. Gomes

PII: S1385-1101(18)30224-7
DOI: https://doi.org/10.1016/j.seares.2019.04.001
Reference: SEARES 1757

To appear in: *Journal of Sea Research*

Received date: 14 July 2018
Revised date: 5 April 2019
Accepted date: 10 April 2019

Please cite this article as: D.F.R. Cleary and N.C.M. Gomes, Baseline information on prokaryotic and microeukaryotic plankton communities inside and outside of Indonesian marine lakes, Journal of Sea Research, https://doi.org/10.1016/j.seares.2019.04.001

Baseline information on prokaryotic and microeukaryotic plankton communities inside and outside of Indonesian marine lakes

Cleary DFR[1,2,*] cleary@ua.pt; dfrcleary@gmail.com and Gomes NCM[1]

[1]Department of Biology, CESAM, Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

[2]Tropical Island Sustainable Development Research Center, National Penghu University of Science and Technology, 300 Liu-Ho Rd., Magong City, Penghu 880, Taiwan

*Corresponding author at: CESAM and Department of Biology, Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal.

Abstract

Marine lakes represent a unique and globally rare aquatic environment characterised by lower salinities and pH and higher temperatures than the surrounding open water environment. Here we provide baseline data on planktonic communities of Archaea, Bacteria and microeukaryotes inside and outside (open water habitat) of three marine lakes (Kakaban, Haji Buang and Tanah Bamban) in the Berau region of Indonesia. Compositional variation was highly congruent with the major axis of variation separating open water from marine lake samples for all three domains. Planktonic Archaea mainly consisted of OTUs assigned to the Euryarchaeota that were closely related to organisms in Genbank previously obtained from seawater samples. The majority of archaeal OTUs were most abundant in open water habitat with a few OTUs abundant in all habitats. Most bacterial sequences were assigned to Proteobacteria, Cyanobacteria and Bacteroidetes with the percentage of Cyanobacteria highest in two of the marine lakes and lowest in the remaining lake (Tanah Bamban). In contrast to Archaea, there

were a number of bacterial OTUs that were markedly more abundant in marine lake habitat. Most microeukaryote sequences were assigned to the Alveolata, Stramenopiles, Opisthokonta, Archaeplastida and Hacrobia. As was the case with Bacteria, a number of abundant microeukaryote OTUs were more abundant in marine lake habitat. Our results thus indicate similar compositional responses to the environmental conditions in marine lake habitat across the major domains of life and point to marine lakes harbouring distinct microbial communities.

Keywords: marine lakes; ordination; next generation sequencing

Introduction

Marine lakes, also known as anchialine lakes, are bodies of saline water surrounded by land and connected to the open water marine environment by fissures and channels (Tomascik et al., 1997; Becking et al., 2013). The size of these fissures and channels is a strong determinant of the local environmental characteristics. Lakes connected by large channels tend to be similar to the open water marine environment whereas lakes with reduced connectivity have more distinct environmental parameters (Becking et al., 2011). In general, marine lakes are characterised by having elevated temperatures, lower salinities and lower pH compared to the surrounding open water marine environment (Becking et al., 2011). These characteristics, make the study of marine lakes particularly interesting given future predictions of ocean acidification whereby ocean pH is expected to decrease by 0.3 units to levels approaching those in present day marine lakes (Caldeira and Wickett 2003; Orr et al. 2005).

There are an estimated 200 known marine lakes across the globe although more may be discovered in the future (Dawson and Hamner, 2005; Becking et al., 2011; Becking et al., 2015). The most recent lakes were discovered in the remote Misool archipelago of Indonesia (Becking

et al., 2015). Most marine lakes are found in Indonesia, Palau and Vietnam (Becking et al., 2011; Becking et al., 2015). Marine lakes have received some degree of notoriety as jellyfish lakes due to the presence of very large populations of various jellyfish species (Cleary et al., 2015). The lakes, however, also host numerous rapidly evolving and endemic species of ascidians, shrimps, sponges and mussels (Holthuis, 1973; Tomascik and Mah, 1994; Dawson and Hamner, 2005; Becking et al., 2016). The benthic fauna of the lakes differs strongly from that of the open water marine environments and is dominated by sponges, mussels and oysters (Tomascik et al., 1997). Corals are generally absent except in lakes connected by relatively large channels. The sponge fauna also differs strongly from that found in the open water marine environment and includes a large number of presumably endemic species (Becking et al., 2013).

In the present study, we focused on marine lakes located within the islands of Kakaban and Maratua in the Berau region of Indonesia and samples from coastal sites surrounding these islands (open water habitat). The lakes have been estimated to have formed 7000 - 12000 years before present (Becking et al., 2011). Previous studies of these lakes have highlighted the distinct flora and fauna and the high degree of endemicity (Holthuis 1973; Tomascik et al. 1997). Recent studies showed that sponges (Cleary et al., 2013), mussels (Cleary et al., 2015) and jellyfish (Cleary et al., 2016) in these lakes hosted diverse and distinct symbiotic prokaryote communities. In the present study, our main goal was to investigate the composition and diversity of prokaryotic (Archaea and Bacteria) and microeukaryotic plankton communities of marine lakes located in the Berau region of East Kalimantan, Indonesia. The specific aims of the present study were to 1. compare higher taxon abundance of Archaea, Bacteria and Eukaryota and their compositional differences in marine lake and open water habitats 2. identify closely related organisms to abundant operational taxonomic units (OTUs) of Archaea, Bacteria and Eukaryota and 3. test for significant compositional concordance among Archaea, Bacteria and Eukaryota sampled in different habitats. To the best of our knowledge, this is the first study to

characterise and compare the planktonic microbiome (Archaea, Bacteria and microeukaryotes) of marine lake and open water habitats.

Material and Methods

Study site

Sampling took place in marine lakes located within Kakaban and Maratua islands and the surrounding sea in the Berau region of East Kalimantan, Indonesia (Fig. 1 and Supplementary Table 1). Annual rainfall from 1987 - 2007 in Tanjung Redeb, Berau ranged from 1700 to 3350 mm year-1 (average 2084 mm); monthly precipitation in Berau ranges from 110-250 mm with lowest rainfall in August (average 117 mm) and highest from November-January (average 223 mm) (http://www.bmkg.go.id/BMKG_Pusat/; Becking et al. 2013). Becking et al. (2011) provided a description of the marine lakes of Kakaban and Maratua. Kakaban is a raised atoll on a relatively flat, 200-300 m, submarine platform. It is a relatively large island with a very large marine lake in the centre (the c. 4 km^2 lake Kakaban). Southern, western and eastern shores of lake Kakaban are fringed by mangroves. The northern shore is predominantly rocky. Tidal amplitude in Kakaban is dampened to 11% of the surrounding sea and the tidal phase has a 3h30 delay indicating limited connection with the surrounding open water environment (Becking et al., 2011). Salinity in lake Kakaban varied from 23-24 ppt, pH varied from 7.0 to 7.8 and temperature from 29 to 31.5 °C (Becking et al., 2011). Maratua is a horseshoe-shaped uplifted atoll with a large open lagoon (29.5 x 6.5 km) ranging in depth from 0.5 – 5 m at low tide. The island is further offshore from the main island of Borneo than Kakaban. A number of anchialine lakes (at least 9), including Haji Buang and Tanah Bamban are found on the inner side of the raised rim. Haji Buang is an elongated 0.14 km^2 lake located on the western arm of Maratua.

Most of the coastline of Haji Buang consists of limestone rock with a small area of mangrove fringing the southern shore. Tidal amplitude in Haji Buang was 48% of the adjacent sea with a tidal delay of 2h30 indicating a limited connection to the sea but higher than lake Kakaban. Salinity in Haji Buang ranged from 26-28.5 ppt, pH varied from 7.3 to 7.8 and temperature from 29 to 30 °C (Becking et al., 2011). Tanah Bamban is located just to the north of Haji Buang lake, from which it is separated by a limestone cliff and mangrove swamp, and is an elongated 0.02 km^2 lake with a maximum length 600 m (Becking et al., 2011). On the east shore of the lake, rock was covered in patches by mussels interspersed with sponges at a lower diversity and abundance than in lake Kakaban and Haji Buang. Salinity in Tanah Bamban ranged from 29 to 30 ppt; temperature and pH were not measured. Kakaban had the lowest degree of connectivity to the open sea followed by Haji Buang; Tanah Bamban had the greatest connectivity to the open sea as indicated by the relatively high salinity and was more affected by human perturbation than the other two lakes, which included harvesting of the local mussel populations (Cleary et al., 2015; Becking et al., 2016). In addition to sampling in marine lakes, we also sampled water from coastal open water areas, thus surrounding the islands of Kakaban and Maratua. Salinity in the open water ranged from 33 – 34 ppt, pH from 8.2 – 8.5 and temperature from 28 - 30 °C.

Sampling

Eleven Water samples were collected from the lakes and the surrounding coastal environment using snorkeling from the 17th to 25th of August 2012 (Fig. 1). These included four samples collected in open water surrounding the islands in which the marine lakes were located, three samples from lake Kakaban, and two samples each from lakes Haji Buang and Tanah Bamban. Water was collected between the depths of 1 - 2 m with a 1.5 L bottle and subsequently 1 L of water was filtered (Sogin et al., 2006; Bowen et al., 2012) through a Millipore® White Isopore Membrane Filter (0.22 µm pore size) to obtain the bacterio-, archaeo- and microeukaryote-

plankton. The filter was subsequently preserved in 96% EtOH. All samples were kept cool (< 4 ºC) immediately after collection and during transport. In the laboratory, samples were stored at -80 ºC until DNA extraction.

DNA extraction and pyrosequencing

We isolated PCR-ready total community DNA (TC-DNA) from water samples using the FastDNA® SPIN Kit (MP Biomedicals) following the manufacturer's instructions. Briefly, the whole membrane filter was transferred to Lysing Matrix E tubes containing a mixture of ceramic and silica particles. The microbial cell lysis was performed in the FastPrep® Instrument (Q Biogene) for 80 seconds at speed 6.0. The extracted DNA was eluted into DNase/Pyrogen-Free Water to a final volume of 50 µl and stored at -20°C until use. Prior to pyrosequencing, the amplicons of the archaeal and bacterial 16S rRNA gene were obtained using the Archaea and Bacteria specific primers ARC344f-mod and Arch958R-mod (Pires et al., 2012) and 27F and 1494R (Gomes et al. 2001) respectively. Using the amplicons of the archaeal and bacterial 16S rRNA gene as template, the V3-V4 regions were amplified using barcoded fusion primers [524F-10-ext, Arch958R-mod (Pires et al., 2012) and V3, V4 (Cleary et al., 2015) respectively] with Roche-454 A Titanium sequencing adapters.

For microeukaryotes, the V4 hypervariable region of the 18S rRNA gene was amplified using the primers EukV4F 5′-CCAGCASCYGCGGTAATTCC-3′ and EukV4R 3'-ACTTTCGTTCTTGATYRA-5' with barcode on the forward primer in a 28 cycle PCR assay (5 cycle used on PCR products) using the HotStarTaq Plus Master Mix Kit (Qiagen, USA) under the following conditions: 94°C for 3 minutes, followed by 28 cycles of 94°C for 30 seconds, 53°C for 40 seconds and 72°C for 1 minute, after which a final elongation step at 72°C for 5 minutes was performed. After amplification, PCR products were checked in 2% agarose gel to determine

the success of amplification and the relative intensity of bands. Multiple samples were pooled together in equal proportions based on their molecular weight and DNA concentrations. Pooled samples were purified using calibrated Ampure XP beads. Pooled and purified PCR product was used to prepare the DNA library following the Illumina TruSeq DNA library preparation protocol. Next-generation, paired-end sequencing was performed at mrDNA Molecular Research LP (http://www.mrdnalab.com/; last checked 2016 11 18) on an Illumina MiSeq device (Illumina Inc, San Diego, CA, USA) following the manufacturer's guidelines. Sequences from each end were joined following Q25 quality trimming of the ends followed by reorienting any 3'-5' reads back into 5'-3', and removal of short reads (< 150 bp). Following previous studies (Cleary et al., 2015; de Voogd et al., 2015), the resultant files were analysed using the QIIME (Quantitative Insights Into Microbial Ecology; (Caporaso et al., 2010) software package (http://www.qiime.org/; last checked 2017-01-20).

Separate fasta and qual files were used as input for the split_libraries.py script. Default arguments were used except for the minimum sequence length, which was set at 250 bps after removal of forward primers and barcodes. In addition to user-defined cut-offs, the split_libraries.py script performs several quality filtering steps (http://qiime.org/scripts/split_libraries.html). OTUs were selected using UPARSE with usearch7 (Edgar, 2013). The UPARSE sequence analysis tool (Edgar, 2013) provides clustering, chimera checking and quality filtering on de-multiplexed sequences. Chimera checking was performed using the UCHIME algorithm (Edgar et al., 2011). The quality filtering as implemented in usearch7 filters noisy reads and preliminary results suggest it gives results comparable to other denoisers such as AmpliconNoise, but is much less computationally expensive (http://drive5.com/usearch/features.html; last checked 2014-01-20). First, reads were filtered with the -fastq_filter command and the following arguments -fastq_trunclen 250 -fastq_maxee 0.5 -fastq_truncqual 15. Sequences were then dereplicated and sorted using the -derep_fulllength and -sortbysize commands. OTU clustering (using a sequence similarity

threshold of 97%) was performed using the -cluster_otus command. An additional chimera check was subsequently applied using the -uchime_ref command with the gold.fa database (http://drive5.com/uchime/gold.fa). AWK scripts were then used to convert the OTU files to QIIME format. In QIIME, representative sequences were selected using the pick_rep_set.py script in QIIME using the 'most_abundant' method. Taxonomy was assigned to reference sequences of OTUs using default arguments in the assign_taxonomy.py script in QIIME with the rdp method (Wang et al. 2007). In the assign_taxonomy.py function, we used a fasta file containing reference sequences from the Greengenes 13_8 release and the rdp classifier method for Bacteria and Archaea. For microeukaryotes, we used the Protist Ribosomal Reference Database 'PR2' to map sequences to the assigned taxonomy (Guillou et al., 2013). Finally, we used the make_otu_table.py script in QIIME to generate a square matrix of OTUs x samples. This was subsequently used as input for further analyses using the R package (R Core Team 2013). Sequence Identifiers of closely related taxa of numerically dominant OTUs for Archaea (≥ 100 sequences), Bacteria (≥ 150 sequences) and microeukaryptes (≥ 2000 sequences) were downloaded using the NCBI Basic Local Alignment Search Tool (BLAST) command line 'blastn' tool with the -db argument set to nt (Zhang et al., 2000). BLAST identifies locally similar regions between sequences, compares sequences to extant databases and assesses the significance of matches; functional and evolutionary relationships can subsequently be inferred. Each run produces a list of hits based on significant similarity between pairs of sequences, i.e., the target sequence and taxa present in the database (or no hits if no significantly similar sequences are found). A discussion of how significance is determined can be found at http://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-1.html. The DNA sequences generated in this study can be downloaded from the NCBI SRA: SRP081069 and SRP068454.

Higher taxon abundance

We tested for significant differences in the relative abundance of selected higher taxa and dominance (the relative abundance of the most abundant OTU in each sample) among habitats with an analysis of deviance using the glm() function in R (R Core Team, 2013). Because the data was proportional, we first applied a glm with the family argument set to binomial. The ratio, however, of residual deviance to residual d.f. in the models substantially exceeded 1 so we set family to 'quasibinomial'. In the 'quasibinomial' family the dispersion parameter is not fixed at one so that it can model over-dispersion. Using the glm model, we tested for significant variation among habitats (open water, Kakaban, Haji Buang, Tanah Bamban) using the anova() function in R with the F test, which is most appropriate when dispersion is estimated by moments as is the case with quasibinomial fits. Detailed descriptions of the functions used here can be found in R (e.g. ?cmdscale) and online in reference manuals (http://cran.r-project.org/web/packages/vegan/index.html; Accessed 27-02-2015).

Composition

Tables containing counts of all bacterial, archaeal and microeukaryote OTUs per sample were imported into R using the read.table() function. For the bacterial OTU table, OTUs not classified as bacteria, unclassified at the level of phylum or classified as chloroplasts and mitochondria were removed prior to statistical analysis. For the archaeal and eukaryote OTU tables, OTUs not classified as Archaea or Eukaryota, respectively, were removed prior to statistical analysis. All tables were $\log_e (x+1)$ transformed (in order to normalise the distribution of data) and distance matrices constructed using the Bray-Curtis index with the vegdist() function in the vegan package (Oksanen et al., 2009) in R. The Bray-Curtis index is one of the most frequently applied (dis)similarity indices used in ecology (Legendre and Gallagher, 2001; Cleary, 2003). Variation in OTU composition among habitats (open water, Kakaban, Haji Buang and Tanah Bamban) was assessed with Principal Coordinates Analysis (PCO) using the cmdscale() function in R with the Bray-Curtis distance matrix as input. Variation among habitats was tested

for significance using the adonis() function in vegan. In the adonis analysis, the Bray-Curtis distance matrix of species composition was the response variable with habitat as independent variable; the strata (block) argument was set to site so that randomisations were constrained to occur within each habitat and not across all habitats. We used the procrustes() function in vegan to test for significant congruence among PCO ordinations of Archaea, Bacteria and Eukaryota with default values used for the arguments in the function. This included scaling, which adjusts one configuration 'Y' to maximum similarity with another configuration 'X'. The scaling is non-symmetric given that Y is scaled to fit X. In addition to the procrustes() function, the protest() function in vegan was used to estimate the significance of the Procrustes statistic. The latter function uses a statistic (r = sqrt(1-ss)) derived from the symmetric Procrustes sum of squares 'ss' and calls the procrustes() function a given number of times (1000 permutations in the present case). Detailed descriptions of the functions used here can be found in R (e.g., ?cmdscale) and online in the reference manuals (e.g., http://cran.r-project.org/web/packages/vegan/index.html; checked 2014 09 21).

Results

Sequencing yielded 48028 sequences, assigned to 113 archaeal OTUs, 25749 sequences assigned to 816 bacterial OTUs and 359058 sequences assigned to 3176 microeukaryote OTUs.

Archaea

Most archaeal sequences belonged to OTUs assigned to Euryarchaeota (68 OTUs and 47753 sequences) followed by Crenarchaeota (40 OTUs and 268 sequences) and Parvarchaeota (5

OTUs and 7 sequences). The percentage of Euryarchaeota in the archaeal community ranged from 99.00 ± 0.54% in Tanah Bamban to 99.88 ± 0.03% in Kakaban. The percentage of Crenarchaeota in turn ranged from 0.09 ± 0.10% in Haji Buang to 0.98 ± 0.54% in Tanah Bamban. The Parvarchaeota were not recorded in open water habitat and represented less than 0.05% of the community in lake habitats. The total number of archaeal phyla varied from 2 in open water habitat to 3 in all lake habitats. The number of archaeal classes varied from 4 in open water habitat and Kakaban to 7 in Tanah Bamban and 8 in Haji Buang while the number of archaeal orders varied from 4 in Kakaban to 5 in open water habitat, 9 in Tanah Bamban and 10 in Haji Buang(Fig. 2).

All water samples from all habitats consisted largely of OTUs assigned to the class Thermoplasmata (Phylum: Euryarchaeota). OTUs assigned to the Crenarchaeota were most abundant in open water habitat and Tanah Bamban while OTUs assigned to the MCG (Miscellaneous Crenarchaeotal Group) class were most abundant in Haji Buang and Tanah Bamban (Fig. 3). Rarefied richness (*n* = 3000 sequences) and evenness were both higher in open water than marine lake habitat.

The PCO analysis of archaeal communities showed that the main axis of variation separated open water samples from marine lake samples and the second axis samples from different marine lakes with the greatest difference between Kakaban and Tanah Bamban (Fig. 4). Overall, there was a highly significant difference in archaeal composition among habitats (adonis: $F_{3,7} = 18.07$, $P < 0.001$, $R^2 = 0.886$). Habitat thus explained close to 90% of the variation in composition. The majority of abundant OTUs were mainly found in open water (Supplementary Table 2). The most abundant OTUs (OTUs 4 and 7), however, were relatively abundant in all habitats. None of the abundant OTUs were mainly found in marine lake habitat. All abundant OTUs, with the exception of OTU-174 were assigned to the Marine group II family (Phylum: Euryarchaeota). OTU-174 was assigned to the Cenarchaeaceae family (Phylum:

Crenarchaeota). Most of the abundant OTUs were closely related (sequence similarity ≥ 99%) to organisms in Genbank, the majority of which were obtained from seawater. Two OTUs (22 and 1082) were closely related to organisms obtained from a hydrothermal vent.

Bacteria

Most bacterial sequences belonged to OTUs assigned to Proteobacteria (538 OTUs and 15913 sequences) followed by Cyanobacteria (17 OTUs and 5284 sequences), Bacteroidetes (118 OTUs and 3110 sequences), Actinobacteria (21 OTUs and 560 sequences) and GN02 (11 OTUs and 409 sequences). Proteobacteria were most abundant in Tanah Bamban at 73.56 ± 4.04% of the local community and least abundant in Kakaban at 54.60 ± 5.15% of the local community (Supp Fig. 1).

The total number of bacterial phyla recorded per habitat varied from 12 in Kakaban and Tanah Bamban to 15 in Haji Buang and 16 in open water. The number of classes varied from 27 in Tanah Bamban to 30 in Kakaban, 33 in open water and Haji Buang while the number of orders varied from 40 in Tanah Bamban to 44 in Kakaban and 54 in open water and Haji Buang. As with Archaea, rarefied OTU richness and evenness were higher in open water than marine lake habitat (Fig 2). Using an alpha-adjusted P value of 0.0025 for multiple comparisons, the relative abundance of most bacterial and archaeal phyla and the major proteobacterial classes did not differ significantly among habitats. The only significant differences were in the relative abundances of Cyanobacteria (highest in lake Kakaban) and Tenericutes (highest in open water habitat). The percentage of Cyanobacteria varied from 0.06 ± 0.08% in Tanah Bamban to 34.21 ± 4.16% in Kakaban while the percentage of Tenericutes varied from 0.00 ± 0.00% in all three marine lake habitats to 0.11 ± 0.08% in open water.

As was the case with Archaea, the main axis of variation of the PCO separated open water samples from marine lake samples (Fig. 4). The second axis separated samples from different lakes with the greatest difference between samples from Kakaban and Haji Buang versus Tanah Bamban. There was also a highly significant difference in bacterial composition among habitats (adonis: $F_{3,7} = 4.56$, $P < 0.001$, $R^2 = 0.662$). Habitat thus explained more than 60% of the variation in composition. Most of the abundant OTUs were distributed throughout all habitats with the most abundant, OTU-7 assigned to the genus *Synechococcus*, relatively abundant in all habitats except Tanah Bamban. This OTU was closely related (sequence similarity = 100%) to an organism obtained from seawater in the Arabian Sea (Supplementary Table 3). There were, however, a number of OTUs that were more abundant in marine lake habitat including OTUs 60, 82, 91 and 94 (Supplementary Table 3), but which had high sequence similarities (100%) to organisms found in open water, estuarine and marine habitats. OTU-60, for example, assigned to the Chromatiales, was closely related to an organism obtained from an estuary in China while OTU-82, assigned to the family Saprospiraceae, was closely related to an organism obtained from hypersaline sediment in the gulf of Mexico. A number of OTUs were also relatively abundant in all habitats except Tanah Bamban including the previously mentioned OTU-7 in addition to OTUs 69, 156, 271 and 2125, assigned to the Rhodobacterales and Rickettsiales orders. The remaining OTUs were either found in all habitats (27, 48, 57, 71, 80, 105, 106, 109, 168, 5055) or were found in all habitats but with highly variable abundances among samples within habitats (59, 72, 75, 86, 176, 4360, 5168).

As was the case with Archaea, most abundant OTUs were closely related (≥ 99% of similarity) to organisms previously detected in marine environments. OTU-140 was the only bacterial OTU with relatively low similarity to sequences deposited in Genbank. This OTU was assigned to the order Rickettsiales, which had a sequence similarity of 97% to an organism previously detected in dinoflagellate (*Karenia brevis*) bloom and nonbloom water in Florida (Supplementary Table 3). This OTU was also only found in Tanah Bamban.

Eukaryota

Most eukaryote sequences belonged to OTUs assigned to the Alveolata (1614 OTUs and 154833 sequences) followed by Stramenopiles (626 OTUs and 83969 sequences), Opisthokonta (352 OTUs and 50105 sequences), Archaeplastida (179 OTUs and 4502 sequences), Hacrobia (102 OTUs and 16540 sequences) and Rhizaria (245 OTUs and 409 sequences). The total number of microeukaryote phyla recorded per habitat varied from 9 in Kakaban to 10 in all other habitats. The number of classes varied from 68 in Kakaban to 74 in Haji Buang, 72 in Tanah Bamban and 97 in open water while the number of orders varied from 104 in Kakaban to 109 in Tanah Bamban and 117 in Haji Buang and 161 in open water. As was the case with Archaea and Bacteria, rarefied OTU richness and evenness were higher in open water than all marine lake habitats and higher in lake Kakaban than Haji Buang and Tanah Bamban (Fig 3).

Alveolata were most abundant in Haji Buang at 64.16 ± 4.12% of the local community and least abundant in Tanah Bamban at 34.44 ± 18.60% of the local community (Fig. 3). The percentage of Stramenopiles varied from 5.56 ± 1.19% in Haji Buang to 61.69 ± 17.69% in Tanah Bamban. Tanah Bamban had the lowest percentages of Archaeplastida, Opisthokonta, Hacrobia and Rhizaria. The percentages of Opisthokonta and Rhizaria were highest in open water. Using an alpha-adjusted P value of 0.0023 for multiple comparisons, the relative abundance of most eukaryote taxa did not differ significantly among habitats. The only significant differences were for Bacillariophyta (highest in Tanah Bamban), Arthropoda (highest in open water), Trebouxiophyceae (highest in Haji Buang) and Platyhelminthes (highest in Haji Buang and Tanah Bamban). In open water, the Arthropoda made up more than 27% of the microeukaryote community, but this was less than 4% in all marine lake habitats and only 0.51% in Tanah Bamban. The Trebouxiophyceae were rare (< 0.35%) in all habitats except Haji Buang where

they made up 6.24% of the community on average. Bacillariophyta made up more than 60% of the eukaryote community of Tanah Bamban on average, but only 6.56 ± 5.75 of open water habitat and 3.10 ± 0.59 of Haji Buang (Fig. 3).

In line with the results obtained for Archaea and Bacteria, there was also a highly significant difference in microeukaryote composition among habitats (adonis: $F_{3,7} = 4.22$, $P < 0.001$, $R^2 = 0.644$). Habitat thus explained more than 60% of the variation in composition. The main axis of variation of the PCO analysis separated open water from marine lake samples (Fig. 4). The second axis separated samples from different marine lakes with the greatest difference between samples from Kakaban and Tanah Bamban. Three of the four most abundant OTUs in open water samples were grazers. These included OTUs 4 and 21, both of which were assigned to the Crustacea. OTU-4 was assigned to the copepod genus *Acrocalanus* and was closely related (sequence similarity = 99%) to an organism obtained from the Atlantic Ocean. OTU-21 had 100% sequence similarity to an organism identified as *Paracalanus aculeatus* (Supplementary Table 3). OTU-23 was assigned to the hydrozoan genus *Nanomia* and was closely related (sequence similarity = 100%) to an organism obtained from the South China Sea. OTU-7 was an abundant (≥ 2000 sequences) non-grazer mainly associated with open water samples and was assigned to the green algae genus *Ostreococcus* and closely related to an organism obtained from seawater in the South China Sea.

In contrast to Bacteria, most of the eukaryote OTUs were mainly found in a single habitat or had variable abundance within habitats (Supplementary Table 3). There were, however, a subset of OTUs that were relatively abundant in Kakaban and Haji Buang including OTUs 3, 13, 17, 18, 20 and 27 assigned to the Mamiellophyceae, Dinophyceae, Syndiniales and Arthropoda.

A number of abundant OTUs were mainly found in Tanah Bamban. These included the most abundant OTU overall, OTU-1, which was assigned to the species *Skeletonema grevillei* and closely related (sequence similarity = 100%) to an organism obtained from seawater in Japan (Supplementary Table 3). Although the most abundant OTU overall, there was a pronounced difference in the relative abundance of OTU-1 among habitats, which was much higher in Tanah Bamban (44.39 ± 12.14) than open water (0.43 ± 0.08), Kakaban (0.60 ± 0.16) or Haji Buang (0.72 ± 0.35). OTU-22 was another diatom, assigned to the genus *Chaetoceros*, that was mainly found in Tanah Bamban. The other abundant diatoms, OTUs 8, 14 and 25 were mainly found in Kakaban. OTUs 15 and 30 were both assigned to the dinoflagellate class Dinophyceae. OTU-25, assigned to the Raphid-pennate group, had 97% sequence similarity to an organism identified as *Cylindrotheca closterium* from the Beaufort Sea (Supplementary Table 3).

As shown in Fig. 3, all habitats had relatively high abundances of OTUs assigned to the marine dinoflagellate class Syndiniales. OTU-27, mainly found in Kakaban had low sequence similarity (95%) to an organism previously detected in seawater from the South China Sea. OTU-10 only had 87% sequence similarity to an organism from the Atlantic Ocean. OTU-26 mainly found in Haji Buang had 100% sequence similarity to an organism obtained from the South China Sea. It also had 96% sequence similarity to an organism identified as *Euduboscquella costata* (Acc: KP749831), which is an intracellular parasite of the tintinnid ciliate *Schmidingerella arcuata* (class: Spirotrichea).

Procrustes analysis was used to investigate if the compositional variation observed in the microbial plankton communities from different habitats were congruent among different microbial domains. Our results showed highly significant correlations ($P < 0.001$; corrs > 0.93) between the community structure of Archaea and Bacteria, Archaea and Eukaryota and Bacteria and Eukaryota in marine lake and open water habitats thus indicating significant congruence in the spatial, compositional variation of all three microbial domains (Fig. 4).

Discussion

Marine lake habitat differed from open sea habitat in OTU richness, evenness, composition and differences in the relative abundance of selected taxa including, Cyanobacteria, Tenericutes, Arthropoda, Bacillariophyta and Trebouxiophyceae. Richness and evenness were lower in marine lake than open water habitat in all three domains. In line with the low evenness in marine lakes, there was pronounced dominance of particular OTUs. OTU-1 assigned to the diatom species *Skeletonema grevillei*, for example made up 44.1 ± 12.1% of the microeukaryote community in Tanah Bamban while OTU-4, assigned to the euryarchaeotal Marine_group_II, made up 70.4 ± 1.0% of the archaeal community in Kakaban. Autotrophic groups including diatoms (Bacillariophyta) were particularly enriched in Kakaban and Tanah Bamban, while Cyanobacteria were more abundant in Kakaban and Haji Buang and Trebouxiophyceae in Haji Buang. In contrast, heterotrophic groups Tenericutes and Arthropoda were more abundant in open water habitat. Overall, these results suggest that despite lower richness, these lakes have potentially higher primary productivity than open water habitat. The marine lakes of Kakaban and Maratua are known to contain diverse plant and animal assemblages and are characterised by high primary productivity. In Kakaban, for example, there are extensive meadows of benthic macrophytes in the lagoon largely consisting of the green calcareous algae species *Halimeda opuntia* and *H. tuna*. Net primary production of *H. opuntia* was 2.47 mg C·g organic dry wt^{-1} h^{-1}. The stilt-root rhizosphora habitats, sandy and mud-bottom habitats also team with marine life including densely packed assemblages of sponges, ascidians and mussels (Tomascik et al., 1997). Sponge density is also far higher than that recorded outside the marine lakes in open water coral reef or mangrove habitat (de Voogd et al., 2009; Becking et al., 2013). Kakaban lake was reported by Tomascik et al. (1997) to have an extremely low N:P ratio of 2.0 indicating N limitation within the lake. They reported the main sources of nitrogen as runoff, rain, nitrogen fixation and limited seawater flushing. Phosphate concentrations were also reported to be

relatively high in this lake. In general, coral atolls, such as Kakaban, are areas of high productivity despite being in regions characterised by low to very low nutrient levels. Tomascik et al. (1997) suggested that nitrogen fixation contributed to the high productivity of lake Kakaban. This fits well with the high relative abundance of autotrophic phytoplankton communities in the lakes

Spatial variation in community composition was, furthermore, similar for all three microbial domains as shown with the Procrustes analysis. The major axis of variation in the ordinations of all three domains separated open water from marine lake habitat. Open water habitat had greater evenness and relative abundances of Tenericutes and Arthropoda than marine lake habitat as mentioned previously. Interestingly, the relative abundance of Arthropoda was much lower in all marine lakes than open sea habitat. The main OTUs assigned to the Arthropoda included OTUs 4, 18 and 21, all of which were assigned to the Crustacea. The most abundant of these, OTU-4, was assigned to the genus *Acrocalanus* and closely related to an organism obtained from seawater in the tropical eastern Atlantic Ocean. Members of the genus *Acrocalanus* are often found in a variety of habitats in the Indopacific region and are one of the most dominant copepods detected in parts of the Gulf of Thailand, north-eastern South China Sea and an estuarine region in Northern Taiwan (Tseng et al., 2008; Hwanget al., 2009; Maiphae and Sa-ardrit, 2011). The community structure and abundance of copepods has been shown to be sensitive to water physicochemical parameters (Lee et al., 2009). For example, Jagadeesan et al. (2013) showed that ocean-current induced shifts in salinity in Palk Bay, Sri Lanka significantly affected copepod composition and abundance. In addition to this, the interaction of low water pH and increased temperature adversely affected the fecundity of female copepods (Foo and Byrne, 2017). Zooplankton grazing is one of the most important processes for transferring phytoplankton carbon to higher trophic levels in the marine environments. The relative lack of predators such as copepods may have an important effect on

the trophic structure of marine lakes, which highlights the importance of further studies on the ecology of these unique environments.

The unique environmental conditions of marine lakes (lower salinity and pH and higher temperature) may act as important drivers of plankton community structure. In line with this hypothesis, Brannock et al. (2016) showed that spatio-temporal variation in the composition of pelagic microeukaryotes in Mobile Bay, Alabama and along the continental shelf were mainly driven by variation in salinity, temperature and nutrient levels. In addition to these variables, reduced pH values of marine lake water may also have played an important role in structuring the composition and abundance of lake plankton communities. Recently, Coelho et al. (2016) showed that under laboratory controlled conditions (microcosm experiments), small reductions in pH (~0.3) affected the composition of archaeal, bacterial and microeukaryote communities inhabiting estuarine surface sediment.

The relative abundances of certain groups varied widely both within and among habitats. This was, for example, the case with the Mamiellophyceae, a class of green algae. Members of the Mamiellophyceae group are often abundant organisms in the phytoplankton of several marine environments across the globe (Monier et al. 2016). Monier et al. (2015) showed that members of this family can be sensitive to minor disturbances in nutrient and light regimes, which may lead to strong shifts in their abundance. Mamiellophyceae are often represented in the phytoplankton by three genera *Bathycoccus*, *Micromonas* and *Ostreococcus* (Monier et al., 2015). Previous studies suggested that different ecotypes within these genera may be specifically adapted to different marine environments (e.g., coastal habitats and pelagic/deep waters) (Slapeta et al. 2006). In line with this, we observed a high abundance of an OTU assigned to the genus *Ostreococcus* (OTU-7) in open water. OTU-3, assigned to the genus *Bathycoccus*, was far more abundant in Kakaban and Haji Buang than open water and Tanah Bamban. OTU-12, assigned to the genus *Micromonas*, was moderately abundant in all habitats.

Trebouxiophyceae, another class of green algae, was specifically enriched in Haji Buang and was mainly represented by OTU-9 assigned to the genus *Nannochloris*. In general, there is a lack of information about the ecology of this genus in marine and brackish waters. Members of this genus, however, are closely related to *Chlorella* spp., which complicates the task of differentiating them using only a short fragment of the 18S rDNA (Henley et al., 2004).

OTUs assigned to the phylum Bacillariophyta (diatoms) were most abundant in Kakaban and Tanah Bamban. However, the OTU analysis indicated that different microeukaryotic populations within this group were specifically enriched in these lakes. While Tanah Bamban was mainly dominated by OTUs assigned to the species *Skeletonema grevillei*, (OTUs 1 and 979) and genus *Chaetoceros* (OTU-22), Kakaban was dominated by OTUs 8, 14 and 25. OTU-8 was unassigned at generic level and OTU-14 was assigned to the genus *Chaetoceros*. OTU-25, assigned to the Raphid-pennate group, had 97% sequence similarity to an organism identified as *Cylindrotheca closterium*. All of these OTUs, with the exception of OTU-25, were assigned to diatoms belonging to the Polar-centric-Mediophyceae group.

Tanah Bamban is somewhat of an enigma. Cyanobacteria, for example, were virtually absent from this lake. Although having the highest probable connectivity to the open water and highest salinity of the marine lakes, the abundance of Cyanobacteria was much lower even than that recorded in open water (Becking et al., 2011). Instead, bacterial plankton communities in Tanah Bamban had relatively high abundances of Gammaproteobacteria, particularly Oceanospirillales members. Most microeukaryote groups were also largely underrepresented in Tanah Bamban while the Bacillariophyta (diatoms) were greatly overrepresented at > 60% of the microeukaryote community. The most abundant of these, OTU-1, was assigned to the diatom species *S. grevillei* and was 100% similar to an organism obtained from water in Onagawa Bay, Japan and identified as strain ‘FON073’ of *Skeletonema grevillei*. *Skeletonema* species belong to the picoplankton fraction, are widely distributed, and can occur over a broad range of

salinities (10 to 35 ppt) (Kooistra et al., 2008; Balzano et al., 2011). Their ability to grow at low salinities, however, varies among different species and ecotypes (Balzano et al., 2011). *Skeletonema grevillei* is also a bloom forming species that thrives in large rivers, coastal embayments and eutrophic, saline inland lakes with water column mixing (Spaulding and Edlund, 2008). The species has also apparently been introduced to various regions of the globe including the Mediterranean Sea (Marić Pfannkuchen et al., 2018). The most abundant non-Bacillariophyte in Tanah Bamban was OTU-15 assigned to the species *Protoperidinium tricingulatum*. This OTU was 100% similar to an isolate obtained from seawater in the Dutch Wadden Sea (Kawami et al., 2009).

Tanah Bamban also had a greater, albeit non-significant, abundance of Thaumarchaeota compared to the other marine lakes. Archaea play an important role in the nitrogen cycle (Radax et al., 2012) with all known cultivated members of the Thaumarchaeota (Mesophilic Crenarchaeota) obtaining energy through ammonia oxidation (Offre et al. 2013). The relative abundance of Crenarchaeota in seawater appears to be related to the role they play in the nitrogen metabolism. Polónia et al. (2014) suggested that Crenarchaeota sequences in reef seawater can be a pollution or nutrient indicator. In a study of three reef systems, the percentage of Crenarchaeota was highest in the most polluted system, Jakarta (Polónia et al., 2014, 2015, 2016). The open water Berau system itself is considered relatively pristine (Buschman et al., 2012; Christianen et al., 2012; Fauzi et al., 2014; van Katwijk et al., 2011) with limited eutrophication although there can be local upwelling and transport of nutrient-rich waters to the surface (Tomascik et al., 1997).

Conclusions

Marine lakes harbored a relatively low diversity of microorganisms, particularly with respect to microeukaryotes, when compared to open water habitat. They were also characterised by higher dominance and compositionally distinct communities of Archaea, Bacteria and microeukaryotes compared to open water. Abundant OTUs of all three domains found within lake habitat were, however, either also found in open water habitat or had high sequence similarity to organisms previously detected in open water. There were also pronounced differences among lakes with one of the lakes, Tanah Bamban, characterised by a very low abundance of Cyanobacteria and relatively high abundance of Thaumarchaeota and Bacillariophyta compared to the other lakes. Overall, different lakes appear to favour the enrichment of specific members of phytoplankton communities. Future studies should focus on the environmental factors and ecological interactions that shape diversity and productivity of microbial plankton communities in marine lakes and their influence on ecosystem functioning and services.

The following are the supplementary data related to this article.

Supplementary Table 1. List of samples used in the present study including the sample code (Sample), Habitat, sampling date (Date), sampling depth (Depth), sampling site (Site), Latitude (Lat) and Longitude (Lon). The percentages of the most abundant phyla, classes and orders are given as is the dominance (Dom; relative abundance of the most abundant OTU), Dom3 (relative abundance of the three most abundant OTUs), rarefied richness (Richness), Pielou's J (J), Shannon's H' diversity index (H), coordinates for the PCO ordination (PC1, PC2, PC3 and PC4) for Archaea (arc), Bacteria (bac) and microeukaryotes (euk).

Supplementary Table 2. List of abundant (≥ 100 sequence reads) OTUs assigned to Bacteria and closely related organisms identified using BLAST search. OTU: OTU number; Sum: number of sequence reads; Preference: habitat where OTU was most abundant, Wide: abundant in all habitats, Variable: locally abundant in various habitats, Opn: open water, Kak: Kakaban, Mab:

Haji Buang, Tan: Tanah Bamban. If not present in all samples of a given habitat, the number between brackets after the code indicates the number of samples in which the OTU was found; Acc: Accession sequence identifiers of closely related organisms identified using BLAST; Seq: sequence similarity of these organisms with our representative OTU sequences; Source: isolation source of organisms identified using BLAST.

Supplementary Table 3. List of abundant (≥ 150 sequence reads) OTUs assigned to Archaea and closely related organisms identified using BLAST search. OTU: OTU number; Sum: number of sequence reads; Preference: habitat where OTU was most abundant, Wide: abundant in all habitats, Variable: locally abundant in various habitats, Opn: open water, Kak: Kakaban, Mab: Haji Buang, Tan: Tanah Bamban. If not present in all samples of a given habitat, the number between brackets after the code indicates the number of samples in which the OTU was found; Acc: Accession sequence identifiers of closely related organisms identified using BLAST; Seq: sequence similarity of these organisms with our representative OTU sequences; Source: isolation source of organisms identified using BLAST.

Supplementary Table 4. List of abundant (≥ 2000 sequence reads) OTUs assigned to Eukaryota and closely related organisms identified using BLAST search. OTU: OTU number; Sum: number of sequence reads; Preference: habitat where OTU was most abundant, Wide: abundant in all habitats, Variable: locally abundant in various habitats, Opn: open water, Kak: Kakaban, Mab: Haji Buang, Tan: Tanah Bamban. If not present in all samples of a given habitat, the number between brackets after the code indicates the number of samples in which the OTU was found; Acc: Accession sequence identifiers of closely related organisms identified using BLAST; Seq: sequence similarity of these organisms with our representative OTU sequences; Source: isolation source of organisms identified using BLAST.

Acknowledgements

The research was sponsored by the Indonesian Institute of Sciences (LIPI) and the Portuguese Foundation for Science and Technology, FCT, project LESS CORAL, PTDC/AAC-AMB/115304/2009 and Ecotech-Sponge (PTDC/BIAMIC/6473/2014 – POCI-01-0145-FEDER-016531). Thanks are also due, for the financial support to CESAM [UID/AMB/50017/2019], to FCT/MEC through national funds, and co-funding by FEDER, within the PT2020 Partnership Agreement and Compete 2020. We are grateful to the Indonesian State Ministry of Research and Technology (RISTEK) for providing research permits. We thank the following people for their help in various ways: Leontine Becking, Rossana Freitas, Suharsono, Nicole de Voogd, Ana Rita Polónia, Yos Tuti, E. Oberhauser, R. Suhr, and the staff of Nabucco Island Dive Resort.

References

1. Balzano, S., Sarno, D., Kooistra, W., 2011. Effects of salinity on the growth rate and morphology of ten Skeletonema strains. J Plankton Res 33, 937–945.

2. Becking, L.E., Cleary, D.F.R., de Voogd, N.J., 2013. Sponge species composition, abundance, and cover in marine lakes and coastal mangroves in Berau, Indonesia. Mar Ecol Prog Ser 481, 105–120. doi: 10.3354/meps10155.

3. Becking, L.E., de Leeuw, C.A., Knegt, B., Maas, D.L., de Voogd, N.J., Abdunnur, Suyatna, I., Peijnenburg, K.T.C.A. 2016. Highly divergent mussel lineages in isolated Indonesian marine lakes. PeerJ 4, e2496. doi: 10.7717/peerj.2496.

4. Becking, L.E., de Leeuw, C., Vogler, C., 2015. Newly discovered “jellyfish lakes” in Misool, Raja Ampat, Papua, Indonesia. Mar Biodivers 45, 597-598. doi: 10.1007/s12526-014-0268-6.

5. Becking, L.E., Renema, W., Santodomingo, N.K., Hoeksema, B.W., Tuti, Y., de Voogd, N.J., 2011, Recently discovered landlocked basins in Indonesia reveal high habitat diversity in anchialine systems. Hydrobiologia 677, 89–105. doi: 10.1007/s10750-011-0742-0.

6. Brannock, P.M., Ortmann, A.C., Moss, A.G., Halanych, K.M., 2016, Metabarcoding reveals environmental factors influencing spatio-temporal variation in pelagic micro-eukaryotes. Mol Ecol 25, 3593–3604.

7. Buschman, F.A., Hoitink, A.J.F., De Jong, S.M., Hoekstra, P., Hidayat, H., Sassi, M.G., 2012. Suspended sediment load in the tidal zone of an Indonesian river. Hydrol Earth Syst Sci 16, 4191-4204.

8. Caldeira, K., Wickett, M.E., 2003. Anthropogenic carbon and ocean pH. Nature 425, 365.

9. Christianen, M.J., Govers, L.L., Bouma, T.J., Kiswara, W., Roelofs, J.G., Lamers, L.P., van Katwijk, M.M., 2012. Marine megaherbivore grazing may increase seagrass tolerance to high nutrient loads. J Ecol 100, 546–560.

10. Cleary, D.F.R., 2003. An examination of scale of assessment, logging and ENSO-induced fires on butterfly diversity in Borneo. Oecologia 135, 313-321.

11. Cleary, D.F.R., Becking, L.E., Pires, A.C.C., de Voogd, N.J., Egas, C., Gomes, N.C.M., 2013. Habitat and host related variation in sponge bacterial communities in Indonesian coral reefs and marine lakes. FEMS Microbiol Ecol 85, 465–482.

12. Cleary, D.F.R., Becking, L.E., Polónia, A.R.M,, Freitas, R., Gomes, N.C.M., 2015. Composition and putative functional ecology of mussels inhabiting Indonesian marine lakes. Anton Leeuw Int J G 107, 821-834. doi: 10.1007/s10482-014-0375-1.

13. Cleary, D.F.R., Becking, L.E., Polónia, A.R.M., Freitas. R., Gomes, N.C.M., 2016. Jellyfish associated microbiomes of Indonesian Marine lakes. FEMS Microbiol Ecol 92, fiw064. doi: 10.1093/femsec/fiw064.

14. Dawson, M.N., Hamner, W.M., 2005. Rapid evolutionary radiation of marine zooplankton in peripheral environments. P Natl Acad Sci USA 102, 9235-9240. doi: 10.1073/pnas.0503635102.

15. de Voogd, N.J., Becking, L.E., Cleary, D.F.R., 2009. Sponge community composition in the Derawan Islands, NE Kalimantan, Indonesia. Marine Ecology Progress Series 396, 219-230. doi: 10.3354/meps08349.

16. Edgar, R.C., 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10, 996–998.

17. Edgar, R., Haas, B., Clemente, J., Quince, C., Knight, R., 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27, 2194–2200.

18. Fauzi, A., Skidmore, A.K., Heitkönig, I.M., van Gils, H., Schlerf, M., 2014. Eutrophication of mangroves linked to depletion of foliar and soil base cations. Environ Monit Assess 186, 8487–8498.

19. Foo, S.A., Byrne, M., 2017. Marine gametes in a changing ocean: Impacts of climate change stressors on fecundity and the egg. Mar Environ Res 128, 12–24.

20. Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C. et al., 2013. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Res 41, (D1) D597-D604. doi: 10.1093/nar/gks1160 [PMID: 23193267].

21. Henley, W.J., Hironaka, J.L., Guillou, L., Buchheim, M.A., Buchheim, J.A., Fawley, M.W., Fawley, K.P., 2004. Phylogenetic analysis of the 'Nannochloris-like' algae and diagnoses of Picochlorum oklahomensis gen. et sp. nov. (Trebouxiophyceae, Chlorophyta). Phycologia 43, 641-652.

22. Hilty, M., Burke, C., Pedro, H., Cardena,s P., Bush, A., Bossley, C., Davies, J., Ervine, A., Poulter, L., Moffatt, M.F., Cookson, W.O., 2010. Disordered microbial communities in asthmatic airways. PloS One 5, e8578.

23. Holthuis, L.B., 1973. Caridean shrimps found in land-locked saltwater pools at four indo-west pacific localities (Sinai Peninsula, Funafuti Atoll, Maui and Hawaii Islands), with the description of one new genus and four new species. Zool Verhandel 128, 1-48.

24. Hwang, J.S., Souissi, S., Dahms, H.U., Tseng, L.C., Schmitt, F.G., Chen, Q.C., 2009. Rank-abundance allocations as a tool to analyze planktonic copepod assemblages off the Danshuei river estuary (Northern Taiwan). Zool Stud 48, 49-62.

25. Jagadeesan, L., Jyothibabu, R., Anjusha, A., Arya, P.M., Madhu, N.V., Muraleedharan, K.R., Sudheesh, K., 2013. Ocean currents structuring the mesozooplankton in the Gulf of Mannar and the Palk Bay, southeast coast of India. Prog Oceanogr 110, 27-48.

26. Kawami, H., van Wezel, R., Koeman, R.P.T., Matsuoka, K., 2009. *Protoperidinium tricingulatum* sp. nov. (Dinophyceae), a new motile form of a round, brown, and spiny dinoflagellate cyst. Phycol Res 57, 259-267. DOI: 10.1111/j.1440-1835.2009.00545.x.

27. Kooistra, W., Sarno, D., Balzano, S., Gu, H., Andersen, R., Zingone A., 2008. Global diversity and biogeography of *Skeletonema* species (Bacillariophyta). Protist 159, 177-193.

28. Legendre, P., Gallagher, E.D., 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129, 271–280. doi: 10.1007/s004420100716.

29. Lee, C.Y., Liu, D.C., Su, W.C., 2009. Seasonal and spatial variations in the planktonic copepod community of Ilan Bay and adjacent Kuroshio waters off Northeastern Taiwan. Zool Stud 48, 151-161.

30. Maiphae, S., Sa-ardrit, P., 2011. Marine copepods at Mo KoThale Tai, Gulf of Thailand. Songklanakarin J Sci Technol 33, 641-651.

31. Marić Pfannkuchen, D., Godrijan, J., Smodlaka Tanković, M., Baričević, A., Kužat, N., Djakovac, T., Pustijanac, E., Jahn, R., Pfannkuchen, M., 2018. The ecology of one cosmopolitan, one newly introduced and one occasionally advected species from the genus *Skeletonema* in a highly structured ecosystem, the northern Adriatic. Microb Ecol 75, 674-687. doi: 10.1007/s00248-017-1069-9.

32. Monier, A., Comte, J., Babin, M., Forest, A., Matsuoka, A., Lovejoy, C., 2015. Oceanographic structure drives the assembly processes of microbial eukaryotic communities. ISME J 9, 990–1002.

33. Monier. A., Worden, A.Z., Richards, T.A., 2016. Phylogenetic diversity and biogeography of the Mamiellophyceae lineage of eukaryotic phytoplankton across the oceans. Environ Microbiol Rep 8, 461–469.

34. Offre, P., Spang, A., Schleper, C., 2013. Archaea in biogeochemical cycles. Annu Rev Microbiol 67, 437-457.

35. Orr, J.C., Fabry, V.J., Aumont, O. et al. (2005) Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. Nature 437, 681–686.

36. Polónia, A.R.M., Cleary, D.F.R., Duarte, L.N., de Voogd, N.J., Gomes, N.C.M., 2014. Composition of Archaea in seawater, sediment and sponges in the Kepulauan Seribu reef system, Indonesia. Microb Ecol 67, 553–567.

37. Polónia, A.R.M., Cleary, D.F.R., Freitas, R., Coelho, F.J.R.C., de Voogd, N.J., Gomes, N.C.M., 2016. Comparison of archaeal and bacterial communities in two sponge

species and seawater from an Indonesian coral reef environment. Mar genom 29, 69-80. doi: 10.1016/j.margen.2016.04.014.

38. Polónia, A.R.M., Cleary, D.F.R., Freitas, R., de Voogd, N.J., Gomes, N.C.M., 2015. The putative functional ecology and distribution of archaeal communities in an Indonesian coral reef environment. Mol Ecol 24, 409–423. DOI: 10.1111/mec.13024.

39. R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from http://www.R-project.org/.

40. Radax, R., Hoffmann, F., Rapp, H.T., Leininger, S., Schleper, C., 2012. Ammonia-oxidizing archaea as main drivers of nitrification in cold-water sponges. Environ Microbiol 14, 909–923.

41. Slapeta. J,, López-García, P., Moreira, D., 2006. Global dispersal and ancient cryptic species in the smallest marine eukaryotes. Mol Biol and Evol 23, 23–29.

42. Spaulding, S., Edlund, M., 2008\. *Skeletonema*. In: Diatoms of the United States. Retrieved April 05, 2018, from http://westerndiatoms.colorado.edu/taxa/genus/Skeletonema.

43. Tomascik, T., Mah, A.J., 1994. The ecology of 'Halimeda lagoon': an achialine lagoon of a raised atoll, Kakaban Island, East Kalimantan, Indonesia. Tropical Biodiversity 2, 385-399.

44. Tomascik, T., Mah, A.J., Nontji, A., Moosa, M.K., 1997. The Ecology of the Indonesian Seas. Part II. The Ecology of Indonesia Series, Volume VIII. 756 Pages, Periplus, Singapore, ISBN: 9780198501862.

45. Tseng, L.C., Kumar, R., Dahms, H.U., Chen, C.T., Souissi, S., Chen, Q.C., Hwang, J.S., 2008. Copepod community structure over a marine outfall area in the northeastern South China Sea. J Mar Biol Assoc UK 88, 955-966.

46. Van Katwijk, M.M., Van der Welle, M.E.W., Lucassen, E.C.H.E.T., Vonk, J.A., Christianen, M.J.A., Kiswara, W., Hakim, I.I., Arifin, A., Bouma, T.J., Roelofs, J.G.M., Lamers, L.P.M., 2011. Early warning indicators for river nutrient and sediment loads in tropical seagrass beds: a benchmark from a near-pristine archipelago in Indonesia. Mar Pollut Bull 62,1512-1520.

47. Zhang, Z., Schwartz, S., Wagner, L., Miller, W., 2000., A greedy algorithm for aligning DNA sequences. J Comput Biol 7, 203–214.

Figure 1. Map of the study area showing the location of the marine lake Kakaban, Haji Buang and Tanah Bamban sampled on the islands of Kakaban and Maratua and the location of these islands in Indonesia. Open water samples were sampled in the sea surrounding both islands.

Figure 2. Mean relative abundance of the most abundant bacterial and archaeal phyla and classes, OTU richness and evenness. Error bars represent a single standard deviation. a) Proteobacteria, b) Cyanobacteria, c) Bacteroidetes, d) Actinobacteria , e) GN02, f) Verrucomicrobia, g) Firmicutes, h) ZB3, i) Planctomycetes, j) Acidobacteria, k) Tenericutes, l) TM6, m) SBR1093, n) Thermoplasmata, o) Thaumarchaeota, p) MCG, q) Gammaproteobacteria, r) Alphaproteobacteria, s) Deltaproteobacteria, t) Betaproteobacteria and diversity components u) archaeal evenness, v) archaeal OTU richness w) bacterial evenness and x) bacterial OTU richness in the following habitats: open water (Opn), Kakaban (Kak), Haji Buang (Mab) and Tanah Bamban (Tan). Results of the GLM analyses for each taxon are presented in the top right of each subfigure.

Figure 3. Mean relative abundance of the most abundant microeukaryote XXX, phyla and classes, OTU richness and evenness. Error bars represent a single standard deviation. a) Alveolata, b) Stramenopiles, c) Opisthokonta, d) Archaeplastida, e) Hacrobia, f) Rhizaria, g)

Apusozoa, h) Excavata, i) Amoebozoa, j) Dinophyceae, k) Bacillariophyta, l) Syndiniales, m) Arthropoda , n) Mamiellophyceae, o) Cryptophyceae, p) Trebouxiophyceae, q) Picobiliphyta_X, r) Spirotrichea, s) Cnidaria, t) MAST u) Choanoflagellatea, v) Platyhelminthes and diversity components w) Evenness and x) Richness in the following habitats: open water (Opn), Kakaban (Kak), Haji Buang (Mab) and Tanah Bamban (Tan). Results of the GLM analyses for each taxon are presented in the top right of each subfigure.

Figure 4. Ordination showing the first two axes of the PCO analysis for a) Archaea, c) Bacteria and e) Eukaryota. Symbols represent samples from from open water (Opn), Kakaban (Kak), Haji Buang (Mab) and Tanah Bamban (Tan). Numbers represent abundant (≥ 100 sequence reads for Archaea, ≥ 150 sequence reads for Bacteria and 2000 reads for Eukaryota) OTUs referred to in Supplementary Table 2 for Archaea, Supplementary Table 3 for Bacteria and Supplementary Table 4 for Eukaryota. The circle size of OTUs is proportional to the abundance (number of sequences). Procrustes analysis comparing (b) Archaea and Bacteria, (d) Archaea and Eukaryota and (f) Bacteria and Eukaryota. For (b) the arrow bases indicates the positions of the samples in the bacteria map while the arrowheads indicated the corresponding positions of the samples in the Archaea map. For (d) the arrow bases indicates the positions of the samples in the Eukaryota map while the arrowheads indicated the corresponding positions of the samples in the Archaea map. For (f) the arrow bases indicates the positions of the samples in the Eukaryota map while the arrowheads indicated the corresponding positions of the samples in the Bacteria map.

Highlights

- Archaea, Bacteria and Eukaryota in marine lakes and open water habitats
- Compositional variation was highly congruent for all three groups
- There were marked compositional differences between marine lake and open water habitat

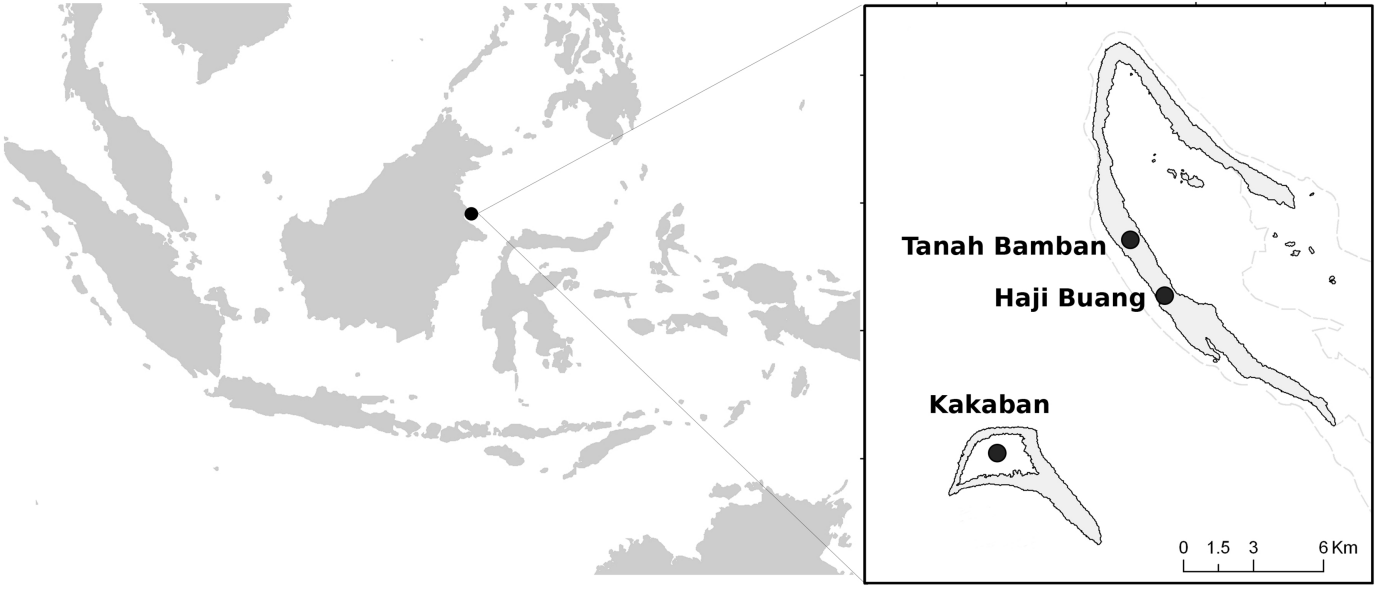

Figure 1

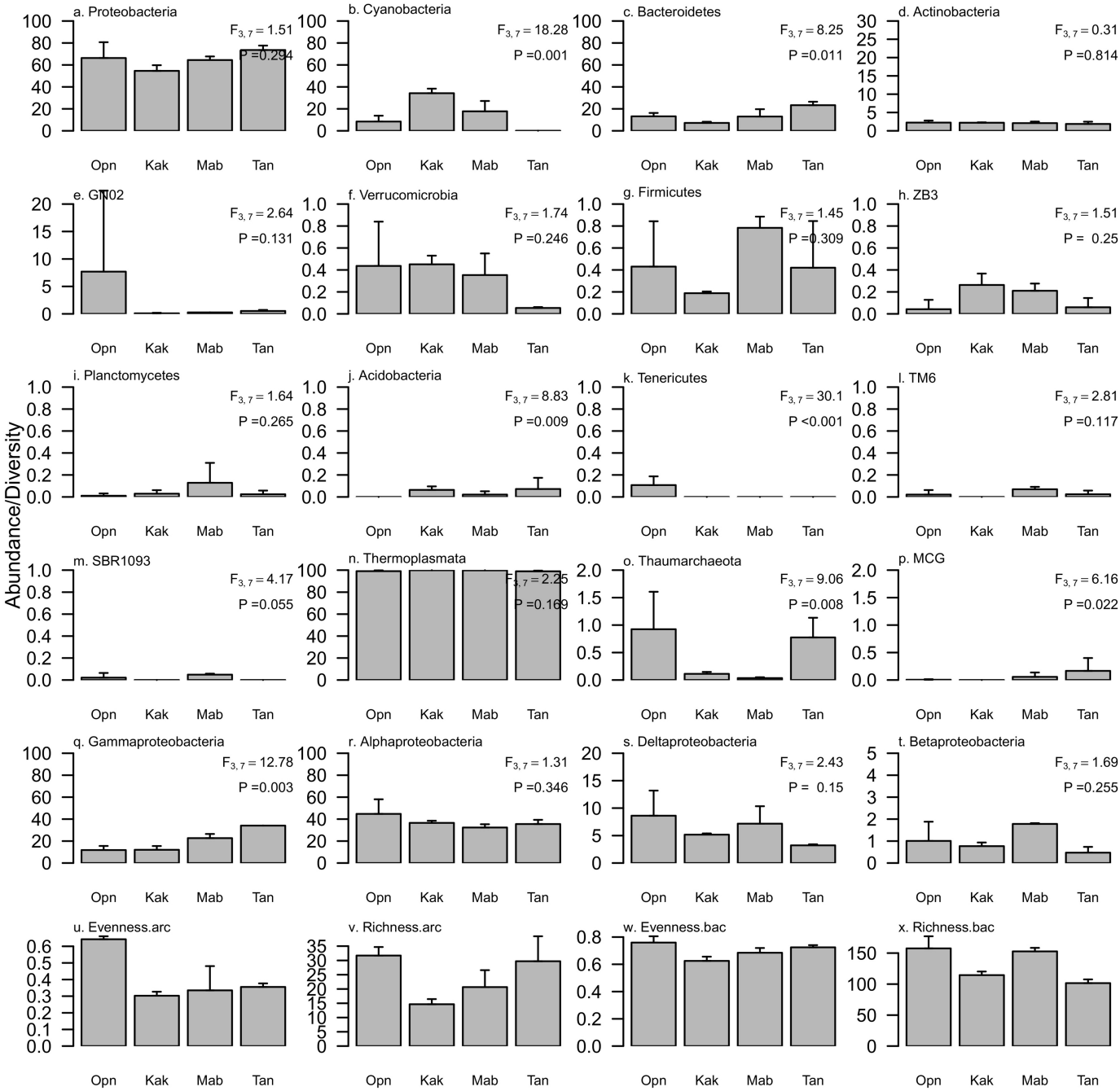
a. Proteobacteria
$F_{3,7} = 1.51$
P =0.294
b. Cyanobacteria
$F_{3,7} = 18.28$
P =0.001
c. Bacteroidetes
$F_{3,7} = 8.25$
P =0.011
d. Actinobacteria
$F_{3,7} = 0.31$
P =0.814
e. GN02
$F_{3,7} = 2.64$
P =0.131
f. Verrucomicrobia
$F_{3,7} = 1.74$
P =0.246
g. Firmicutes
$F_{3,7} = 1.45$
P =0.309
h. ZB3
$F_{3,7} = 1.51$
P = 0.25
i. Planctomycetes
$F_{3,7} = 1.64$
P =0.265
j. Acidobacteria
$F_{3,7} = 8.83$
P =0.009
k. Tenericutes
$F_{3,7} = 30.1$
P <0.001
l. TM6
$F_{3,7} = 2.81$
P =0.117
m. SBR1093
$F_{3,7} = 4.17$
P =0.055
n. Thermoplasmata
$F_{3,7} = 2.25$
P =0.169
o. Thaumarchaeota
$F_{3,7} = 9.06$
P =0.008
p. MCG
$F_{3,7} = 6.16$
P =0.022
q. Gammaproteobacteria
$F_{3,7} = 12.78$
P =0.003
r. Alphaproteobacteria
$F_{3,7} = 1.31$
P =0.346
s. Deltaproteobacteria
$F_{3,7} = 2.43$
P = 0.15
t. Betaproteobacteria
$F_{3,7} = 1.69$
P =0.255
u. Evenness.arc
v. Richness.arc
w. Evenness.bac
x. Richness.bac
Opn
Kak
Mab
Tan
Abundance/Diversity

Figure 2

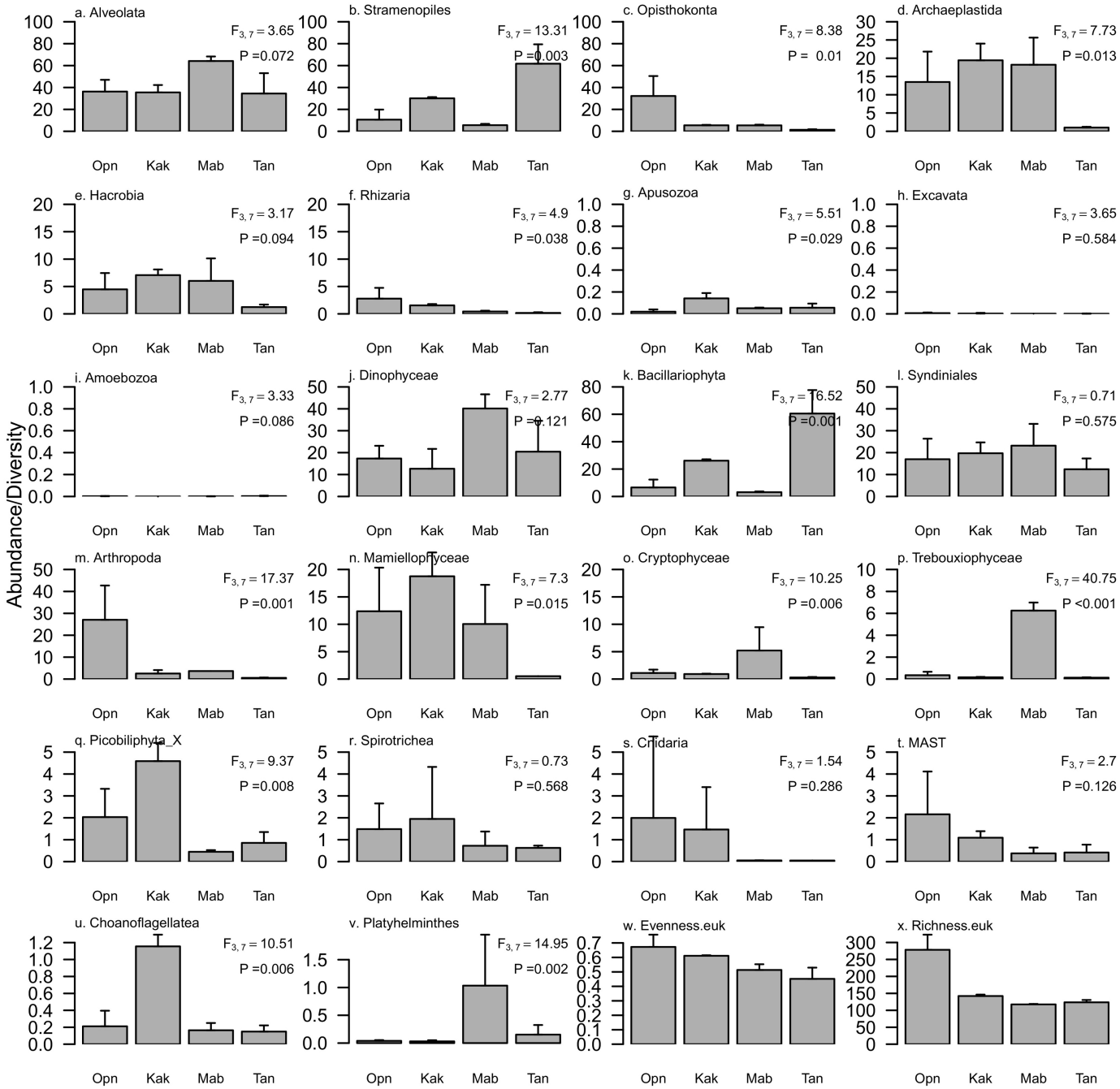
Abundance/Diversity
a. Alveolata
$F_{3,7} = 3.65$
P =0.072
b. Stramenopiles
$F_{3,7} = 13.31$
P =0.003
c. Opisthokonta
$F_{3,7} = 8.38$
P = 0.01
d. Archaeplastida
$F_{3,7} = 7.73$
P =0.013
e. Hacrobia
$F_{3,7} = 3.17$
P =0.094
f. Rhizaria
$F_{3,7} = 4.9$
P =0.038
g. Apusozoa
$F_{3,7} = 5.51$
P =0.029
h. Excavata
$F_{3,7} = 3.65$
P =0.584
i. Amoebozoa
$F_{3,7} = 3.33$
P =0.086
j. Dinophyceae
$F_{3,7} = 2.77$
P =0.121
k. Bacillariophyta
$F_{3,7} = 16.52$
P =0.001
l. Syndiniales
$F_{3,7} = 0.71$
P =0.575
m. Arthropoda
$F_{3,7} = 17.37$
P =0.001
n. Mamiellophyceae
$F_{3,7} = 7.3$
P =0.015
o. Cryptophyceae
$F_{3,7} = 10.25$
P =0.006
p. Trebouxiophyceae
$F_{3,7} = 40.75$
P <0.001
q. Picobiliphyta_X
$F_{3,7} = 9.37$
P =0.008
r. Spirotrichea
$F_{3,7} = 0.73$
P =0.568
s. Cnidaria
$F_{3,7} = 1.54$
P =0.286
t. MAST
$F_{3,7} = 2.7$
P =0.126
u. Choanoflagellatea
$F_{3,7} = 10.51$
P =0.006
v. Platyhelminthes
$F_{3,7} = 14.95$
P =0.002
w. Evenness.euk
x. Richness.euk
Opn
Kak
Mab
Tan

Figure 3

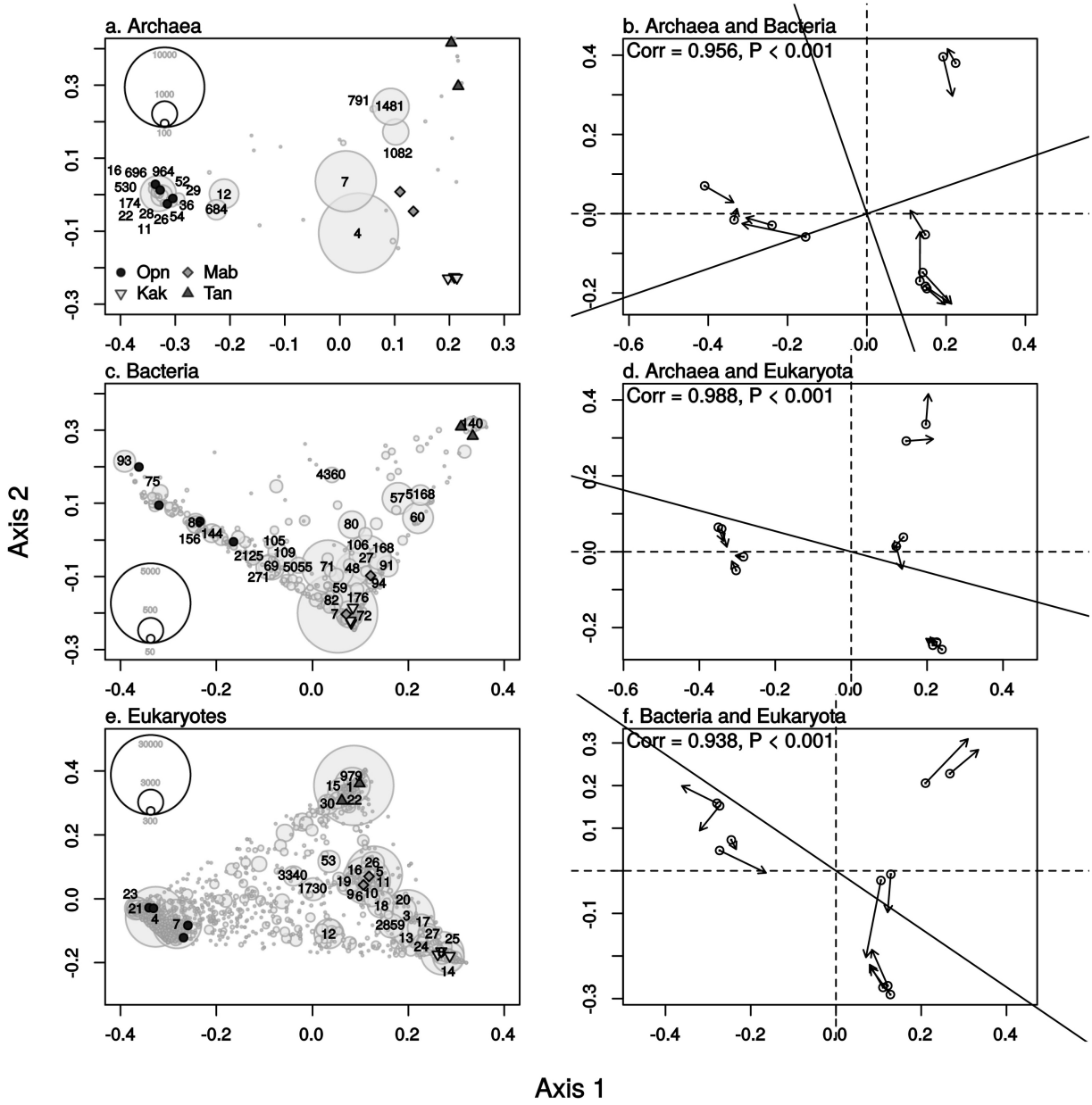
a. Archaea
Opn
Mab
Kak
Tan
b. Archaea and Bacteria
Corr = 0.956, P < 0.001
c. Bacteria
d. Archaea and Eukaryota
Corr = 0.988, P < 0.001
e. Eukaryotes
f. Bacteria and Eukaryota
Corr = 0.938, P < 0.001
Axis 1
Axis 2

Figure 4