Global analysis of gene expression dynamics within the marine microbial community during the VAHINE mesocosm experiment in the South West Pacific

Ulrike Pfreundt[1], Dina Spungin[2], Sophie Bonnet[3], Ilana Berman-Frank[2], Wolfgang R. Hess[1]

[1]{Institute for Biology III, University of Freiburg, Germany}

[2]{Mina and Everard Goodman Faculty of Life Sciences, Bar Ilan University Ramat Gan, Israel}

[3]{Aix Marseille Université, CNRS/INSU, Université de Toulon, IRD, Mediterranean Institute of Oceanography (MIO) UM110 13288, Marseille, France}

[4]{Institut de Recherche pour le Développement, AMU/ CNRS/INSU, Université de Toulon, Mediterranean Institute of Oceanography (MIO) UM110, 98848 Nouméa, New Caledonia}

Correspondence to: W. R. Hess (wolfgang.hess@biologie.uni-freiburg.de)

Submitted to Biogeosciences as a research article for the special issue “Biogeochemical and biological response to a diazotroph bloom in a low-nutrient, low-chlorophyll ecosystem: results from the VAHINE mesocosms experiment”.

Abstract. Microbial gene expression was followed for 23 days within a mesocosm (M1) isolating 50 m^3 of seawater and in the surrounding waters in the Nouméa lagoon, New Caledonia, in the South West Pacific as part of the VAHINE experiment. The aim of VAHINE was to examine the fate of diazotroph-derived nitrogen (DDN) in a Low Nutrient, Low Chlorophyll ecosystem. On day 4 of the experiment, the mesocosm was fertilized with phosphate. In the lagoon, gene expression was dominated by the cyanobacterium *Synechococcus*, closely followed by alphaproteobacteria. In contrast, drastic changes in the microbial community composition and transcriptional activity were triggered within the mesocosm within the first 4 days, with transcription bursts from different heterotrophic bacteria in rapid succession. The composition and activity of the lagoon ecosystem appeared more stable, although following similar temporal trends as in M1. We detected significant gene expression from Chromerida in M1, as well as the Nouméa lagoon, suggesting these photoautotrophic alveolates were present in substantial numbers in the open water. Other groups contributing substantially to the metatranscriptome were affiliated with marine Euryarchaeota *Candidatus* Thalassoarchaea (inside and outside) and *Myoviridae* bacteriophages likely infecting *Synechococcus*, specifically inside M1. High transcript abundances for ammonium transporters and glutamine synthetase in many different taxa (e.g., *Pelagibacteraceae*, *Synechococcus, Prochlorococcus* and *Rhodobacteraceae*) underscored the preference of most bacteria for this nitrogen source. In contrast, *Alteromonadaceae* highly expressed urease genes; *Rhodobacteraceae* and *Prochlorococcus* showed some urease expression. Nitrate reductase transcripts were detected on day 10 very prominently in *Synechococcus* and in *Halomonadaceae*. Alkaline phosphatase was expressed prominently only between day 12 and 23 in different organisms, suggesting that the microbial community was not limited by phosphate, even before the fertilization on day 4, whereas the post-fertilization community was.

We observed high expression of the *Synechococcus sqdB* gene, only transiently lowered following phosphate fertilization. *SqdB* encodes UDP-sulfoquinovose synthase, possibly enabling marine picocyanobacteria to minimize their phosphorus requirements by substitution of phospholipids with sulphur-containing glycerolipids. This result suggests a link between *sqdB* expression and phosphate availability in situ.

Gene expression of diazotrophic cyanobacteria was mainly attributed to *Trichodesmium* and *Richelia intracellularis* (diatom-diazotroph association) in the Nouméa lagoon and initially in M1. UCYN-A (*Candidatus* Atelocyanobacterium) transcripts were the third most abundant and declined both inside and outside after day 4, consistent with 16S- and *nifH*-based analyses. Transcripts related to the *Epithemia turgida* endosymbiont and *Cyanothece* ATCC 51142 increased during the second half of the experiment.

1 Introduction

In the study of natural marine microbial populations, it is of fundamental interest to identify the biota these populations consist of and to elucidate their transcriptional activities in response to biotic or abiotic changes in the environment. Metatranscriptomics gives insight into these processes at high functional and taxonomic resolution, as shown, e.g. in the analysis of a wide range of marine microbial populations (Frias-Lopez et al., 2008; Ganesh et al., 2015; Gifford et al., 2014; Hewson et al., 2010; Hilton et al., 2015; Jones et al., 2015; Moran et al., 2013; Pfreundt et al., 2014; Poretsky et al., 2009; Shi et al., 2009; Steglich et al., 2015; Wemheuer et al., 2015). Here we report the results of a metatranscriptome analysis from the VAHINE mesocosm experiment, whose overarching objective was to examine the fate of diazotroph-derived nitrogen (DDN) in a Low Nutrient, Low Chlorophyll (LNLC) ecosystem (Bonnet et al., 2016). In this experiment, three large-scale (~50 m^3) mesocosms were deployed enclosing ambient oligotrophic water from the Nouméa (New Caledonia) lagoon *in situ*. To alleviate any potential phosphate limitation and stimulate the growth of diazotrophs, the mesocosms were fertilized on day 4 with 0.8 μmol KH_2PO_4 as a source of dissolved inorganic phosphorus (DIP). The mesocosms were sampled daily for 23 days and analyzed with regard to the dynamics of carbon, nitrogen and phosphorus pools and fluxes (Berthelot et al., 2015), the diazotroph community composition on the basis of *nifH* tag sequencing (Turk-Kubo et al., 2015), N_2 fixation dynamics and the fate of DDN in the ecosystem (Berthelot et al., 2015; Bonnet et al., 2016a; Knapp et al., 2015). Furthermore, the composition, succession, and productivity of the autotrophic and heterotrophic communities were studied (Leblanc et al., 2016; Pfreundt et al., 2016; Van Wambeke et al., 2016). During days 15 to 23 of the VAHINE experiment, N_2 fixation rates increased dramatically, reaching >60 nmol N L^{-1} d^{-1} (Bonnet et al. 2015), which are among the highest rates reported for marine waters (Bonnet et al., 2016a; Luo et al., 2012). Based on the analysis of *nifH* sequences, N_2-fixing cyanobacteria of the UCYN-C type were suggested to dominate the diazotroph community in the mesocosms at this time (Turk-Kubo et al., 2015). Evidence from ^{15}N isotope labeling analyses indicated that the dominant source of nitrogen fueling export production shifted from subsurface nitrate assimilated prior to the start of the 23 day experiment to N_2 fixation by the end (Knapp et al., 2015). To link these data to the actual specific activities of different microbial taxa, here we present the community-wide gene expression based on metatranscriptomic data from one representative mesocosm (M1). Throughout the course of the experiment (23 days), we sampled water from both M1 and the surrounding Nouméa lagoon every second day from the surface (1 m) and inferred the metatranscriptomes for the plankton fraction (<1 mm).

2 Methods

2.1 Sampling, preparation of RNA and sequencing libraries

Samples were collected in January 2013 every other day at 7 am from mesocosm 1 (hereafter called M1) and from the Nouméa lagoon (outside the mesocosms) in 10 L carboys using a Teflon pump connected to PVC tubing. To ensure quick processing of samples, the carboys were immediately transferred to the inland laboratory setup on Amédée Island, located 1 nautical mile off the mesocosms. Samples for RNA were prefiltered through a 1 mm mesh to keep out larger eukaryotes and then filtered on 0.45 µm polyethersulfone filters (Pall Supor). These filters were immediately

immersed in RNA resuspension buffer (10 mM NaAc pH 5.2, 200 mM D(+)-sucrose, 100 mM NaCl, 5 mM EDTA) and snap frozen in liquid nitrogen. Tubes with filters were vortexed, then agitated in a Precellys bead beater (Peqlab, Erlangen, Germany) 2x 15s each at 6500 rpm after adding 0.25ml glass beads (0.10-0.25mm, Retsch, Frimley, UK) and 1ml PGTX (39.6g phenole, 6.9 ml glycerol, 0.1g 8-hydroxyquinoline, 0.58g EDTA, 0.8g NaAc, 9.5g guanidine thiocyanate, 4.6 g guanidine hydrochloride, H_2O to 100 ml; (Pinto et al., 2009)). We isolated RNA for metatranscriptomics and DNA for 16S tag-based community analysis (Pfreundt et al., 2016) from the same samples by adding 0.7 ml chloroform, vigorous shaking, incubation at 24 °C for 10 min and subsequent phase separation by centrifugation. RNA and DNA was retained in the aqueous phase, precipitated together and stored at -80 °C for further use.

The samples were treated by TurboDNase (Ambion, Darmstadt, Germany), purified with RNA Clean&Concentrator columns (Zymo Research, Irvine, USA), followed by Ribozero (Illumina Inc., USA) treatment for the depletion of ribosomal RNAs. To remove the high amounts of tRNA from the rRNA depleted samples, these were purified further using the Agencourt RNAClean XP kit (Beckman Coulter Genomics). Then, first-strand cDNA synthesis was primed with an N6 randomized primer. After fragmentation, Illumina TruSeq sequencing adapters were ligated in a strand specific manner to the 5' and 3' ends of the cDNA fragments, allowing the strand-specific PCR amplification of the cDNA with a proof-reading enzyme in 17 to 20 cycles, depending on yields. To secure that the origin of each sequence could be tracked after sequencing, hexameric TruSeq barcode sequences were used as part of the 3' TruSeq sequencing adapters. The cDNA samples were purified with the Agencourt AMPure XP kit (Beckman Coulter Genomics), quality controlled by capillary electrophoresis and sequenced by a commercial vendor (vertis Biotechnologie AG, Germany) on an Illumina NextSeq 500 system using the paired-end (2 x 150 bp) set-up. All raw reads can be downloaded from the NCBI Sequence Read Archive under the BioProject accession number PRJNA304389.

2.2 Pre-treatment and *de-novo* assembly of metatranscriptomic data

Raw paired-end Illumina data in fastq format was pre-treated as follows (read pairs were treated together in all steps to not produce singletons): adapters were removed and each read trimmed to a minimum Phred score of 20 using cutadapt. This left 386,010,015 pairs of good-quality raw reads for the 22 samples. Ribosomal RNA reads were removed using SortMeRNA (Kopylova et al., 2012). The resulting non-rRNA reads (corresponding to a total of 155,022,426 pairs of raw reads binned from all samples) were used as input for *de-novo* transcript assembly with Trinity (Haas et al., 2013) using digital normalization prior to assembly to even out kmer coverage and reduce the amount of input data. Remarkably, data reduction by digital normalization was only ~35 %, hinting at a high complexity of the dataset. This complexity is not surprising regarding that the sample pool contained transcripts from three weeks of experiment in two locations (mesocosm vs. lagoon), yet it also means that there will be a relatively large number of transcripts with very low sequencing coverage. This study thus misses the very rare transcripts in the analyzed community.

The transcript assembly led to 5,594,171 transcript contigs with an N50 of 285 nt, a median contig length of 264 nt, and an average of 326 nt. Transcript abundance estimation and normalization was done using scripts included in the Trinity package. *Align_and_estimate_abundance.pl* used bowtie (Langmead, 2010) to align all reads against all

transcript contigs in paired-end mode, then ran RSEM (Li and Dewey, 2011) to estimate expected counts, TPM, and FPKM values for each transcript in each sample. Only paired-end read support was taken into account. The script *Abundance_estimates_to_matrix.pl* was modified slightly to create a matrix with RSEM expected counts and a TMM-normalized (trimmed mean of M-values normalization method) TPM matrix (the original script uses FPKM here) using the R-package *edgeR*. The latter matrix was used to discard transcript contigs with very low support (maxTPM>=0.25 & meanTPM>=0.02). The remaining 3,844,358 transcript contigs were classified using the Diamond tool (Buchfink et al., 2015) with a blastX-like database search (BLOSUM62 scoring matrix, max e-value 0.001, min identity 10 %, min bit score 50) against the NCBI non-redundant protein database from 10/2015. Normalized TPM values for each transcript contig were added as a weight to the query ID in the Diamond tabular output sample-wise with a custom script, creating one Diamond table per sample which served as input to Megan 5.11.3 (Huson and Weber, 2013). Megan is an interactive tool used here to explore the distribution of blast hits within the NCBI taxonomy and KEGG hierarchy. The parameters used to import the diamond output into Megan were minimum e-value 0.01, minimum bit score 30, LCA 5% (the transcript will be assigned to last common ancestor of all hits with a bit score within 5% of the best hit), minimum complexity 0.3.

During manual analysis of the top 100 transcript contigs according to their mean expression over all samples, we found 9 transcripts to be residual ribosomal RNA or internal transcribed spacer. These contigs were removed from the count and TPM matrices for all multivariate statistics analyses. Absence of these rRNA transcripts in the Diamond output was checked and verified.

2.3 Sample clustering and multivariate analysis

The matrix with expected counts for each transcript contig (see section 2.2) was used as input for differential expression (DE) analysis with edgeR (Robinson et al., 2010) as implemented in the Trinity package script *run_DE_analysis.pl* for the set of samples taken from M1 and the Nouméa lagoon, respectively. edgeR can compute a DE analysis without true replicates by using a user-defined dispersion value, in this case 0.1. We are aware that significance values are highly dependent on the chosen dispersion, and thus only considered transcripts with at least a 4-fold expression change for further DE analysis. The script *analyze_diff_expr.pl* was used (parameters -P 1e-3 -C 2) to extract those transcripts that were at least 4-fold differentially expressed at a significance of ≤ 0.001 in any of the pairwise sample comparisons, followed by hierarchical clustering of samples and differentially expressed transcripts depending on normalized expression values ($\log_2$(TPM+1)). The resulting clustering dendrogram was cut using *define_clusters_by_cutting_tree.pl* at 20 % of the tree height, producing subclusters of similarly responding transcripts.

Non-metric multidimensional scaling (NMDS) was performed in R on the transposed matrix containing all 3,844,358 transcript contigs and their respective TMM-normalized TPM values. First, the matrix values were standardized to raw totals (sample totals) with the decostand() function of the vegan package (Oksanen et al., 2015). Then, metaMDS() was used for calculation of Bray-Curtis dissimilarity and the unconstrained ordination.

2.4 Analysis of specific transcripts

A list with genes of interest was created using the Integrated Microbial Genomes (IMG) system (Markowitz et al., 2015). First, 147 genomes close to bacteria and archaea found in the samples (based on 16S rRNA sequences; (Pfreundt et al., 2016)) were selected using “find genomes”. Then, the "find genes" tool was used to find a gene of interest (for example *nifH*) in all the pre-selected genomes and the resulting genes added to the "gene cart". This was done for all genes of interest and the full gene list including 50 nt upstream of each gene (possible 5’UTR) downloaded in fasta format. Usearch_local (Edgar, 2010) was used to find all transcripts mapping to any of the genes in the list with a minimum query coverage and minimum identity of 60 %.

From the full Diamond output (section 2.2), all matching transcripts together with their taxonomic and functional assignment were extracted and false-positives discarded (*i.e.,* transcripts that mapped to the list of specific genes but had a different Diamond hit). The top hit for each transcript was extracted, and the protein classifications manually curated to yield one common description per function (from different annotations for the same protein in different genomes). The TMM-normalized TPM counts were added to each transcript classification, as well as the full taxonomic lineage from NCBI taxonomy. These taxonomic lineages were curated manually to align taxonomic levels per entry.

The table was imported into R, all counts per sample summed up for each combination of protein and family-level taxa, and a matrix created with samples as row names and combined protein and family description as column names. Heat maps were created separately for each protein group (e.g. rhodopsins or sulfolipid biosynthesis proteins), scaling all values to the group maximum.

3 Results and discussion

The metatranscriptomic data was analysed following the strategy outlined in **Figure 1**. We obtained taxonomic assignments for 37 % of all assembled transcript contigs. This reflects the fact that the genes of complex marine microbial communities, especially from less well sampled ocean regimes like the South West Pacific (as opposed to, for example, Station ALOHA in the subtropical North Pacific), are still insufficiently covered by current databases. The data with taxonomic assignments thus give an overview about the gene expression processes during this mesocosm experiment. With this study, we aimed at identifying global differences in expression patterns between the mesocosm and the lagoon, as well as between the different sampling time points within the mesocosm. We further explored the expression of marker genes for N- and P-metabolism, and light capture in the different taxonomic groups.

3.1 Transcripts cluster into distinct groups with similar expression patterns over time in M1 and the lagoon

Gene expression changes roughly followed the timeline, within both M1 and the Nouméa lagoon, with some exceptions (**Fig. 2**). For the lagoon, samples from day 20 and 23 clustered together, the samples from day 10 to 18 formed a mid-time cluster, and those from day 2 to 8 an early cluster (**Fig. 2B**). In M1, the samples from day 6 to 10 and day 12 to 20 clustered together (**Fig. 2A**). Deviating from the timeline, the sample from day 2 was placed close to day 20, day 23 was separated from the late cluster, and day 4, exhibiting a prominent subcluster of transcripts

upregulated only that day, was the furthest apart from all other samples (**Fig. 2A, black brackets**). Closer inspection of this subcluster containing several hundred different transcripts identified >80 % of them as *Rhodobacteraceae* transcripts, correlating well with a five-fold increase of *Rhodobacteraceae* 16S tags (from 2.5 % to 12.5 % of the 16S community) and a leap in bacterial production between T2 and T4 (Pfreundt et al., 2016). The observed transcripts were broadly distributed across metabolic pathways, reflecting a general increase of *Rhodobacteraceae* gene expression on day 4. The aberrant clustering of the two earliest samples in M1 (before the DIP spike) and the tight clustering of those following the DIP spike (day 6 to 10) suggest an impact of the confinement within the mesocosm and of phosphate supplementation on gene expression.

Unconstrained ordination using non-metric multidimensional scaling (NMDS) confirmed the similar temporal distribution of samples from the Nouméa lagoon and M1 (**Fig. 3**). Yet, the samples from M1 showed a much higher variance and were more dispersed than those from the lagoon (**Fig. 3**) were. Thus, the gene expression profiles within the mesocosm were more diverse than in the lagoon waters. The comparison of the whole dataset against the KEGG database (Kanehisa et al., 2014) showed a major difference between M1 and the lagoon samples only in the category Energy Metabolism and its child categories Photosynthesis and antenna proteins. These categories comprised 22-36 %, 8-16 %, and 2.7-7.5 % in the lagoon, respectively, and were in M1 (excluding day 23) constantly below 22 %, 7 %, and 4 %, respectively (Supplement Fig. S1 and S2). This lower contribution of energy-related functions in M1 was detectable already at the earliest time point (day 2). Furthermore, diverging dynamics in the microbial community composition and transcriptional activity were triggered in M1 already within the first 48 h (before *day 2* was sampled), indicated by the large distance between M1 and lagoon samples on day 2 (**Fig. 3**). The early timing of this effect already on day 2 suggests a rapid remodelling of the microbial community's gene expression upon confinement within the mesocosm. In addition, the DIP spike on the evening of day 4 triggered distinct ecological successions in M1 further. The patterns we observed here are close to three temporal phases defined for the VAHINE experiment based on biogeochemical flux measurements (Bonnet et al., 2016a) and *nifH* amplicon analysis (Turk-Kubo et al., 2015). These were defined as follows (Bonnet et al., 2016a): Day 1-4 (phase P0): before the DIP fertilization, P deplete. Day 5-14 (P1): P availability, dominance of diatom-diazotroph-associations. Day 15-23 (P2): decreasing P availability, slightly higher temperature, increasing N_2-fixation and a dominance of UCYN-C diazotrophs in the mesocosms and increase of primary production (PP) inside and outside the mesocosms.

In the following sections, we refer to P0, P1, or P2 to describe trends and changes in gene expression when appropriate.

3.2 Succession of gene expression inside mesocosm 1 and in the Nouméa lagoon

3.2.1 Active taxonomic groups differ between M1 and the lagoon

The most striking difference between M1 and Nouméa lagoon samples was the 2- to 3-fold dominance of *Oscillatoriophycideae* transcripts over all other taxa in the lagoon over the full time of the experiment, but not in M1 (Supplement Fig. S3A, S4A). Gene expression within the *Oscillatoriophycideae* was mostly attributed to *Synechococcus*, with a substantial share of transcript reads in M1 and the lagoon coming from cyanobacteria closely related to *Synechococcus* CC9605, a strain representative of clade II within the picophytoplankton subcluster 5.1A

(Dufresne et al., 2008) and *Synechococcus* RS9916, a representative of clade IX within picophytoplankton subcluster 5.1B (Scanlan et al., 2009) (Supplement Fig. S3D, S4D). Clade II *Synechococcus* is typical for oligotrophic tropical or subtropical waters offshore or at the continent shelf, between 30°N and 30°S (Scanlan et al., 2009). Contrary to the Nouméa lagoon, *Oscillatoriophycideae* were inferior to alpha- and gammaproteobacteria in M1 during phase P0 and P1 and only gained in P2 a similar level as in the lagoon. We detected substantially higher gene expression from viruses in M1 compared to the Nouméa lagoon (**Fig. 4**). These were assigned mainly to *Myoviridae* such as S.SM2, S.SSM7 and other cyanophages of the T4-like group, which based on their known host association (Frank et al., 2013; Ma et al., 2014), suggest a viral component acting on the *Synechococcus* fraction in the mesocosm. A burst of cyanophages might have contributed to the observed low numbers and activity of *Synechococcus* in M1 compared to the lagoon during P0 and P1 (**Fig. 4**). The recovery of *Synechococcus* populations in M1 during P2 mirrors the increase in the energy and photosynthesis-related functional categories (Supplement Fig. S1, S2) and in *Synechococcus* 16S tag abundance and cell counts (Leblanc et al., 2016; Pfreundt et al., 2016).

Owing to the initial decay of *Synechococcus* in M1, alphaproteobacteria, mainly *Rhodobacteraceae*, SAR11, and SAR116, dominated the metatranscriptome during P0 (**Fig. 4**). Gammaproteobacterial transcripts increased at the beginning of P1, reaching similar levels as those of alphaproteobacteria, and dropped again towards the end of P1, when the *Synechococcus* population started recovering (Supplement Fig. S3C). This suggests that the predominant gammaproteobacteria profited from the organic matter released during bacterial decay. During P2, characterized by an abundant and very active *Synechococcus* population, alphaproteobacteria gene expression increased again. Among these, only SAR11 transcripts were decreasing, by about 75 %.

Unexpected for such a long time course, the temporal pattern of SAR11 transcription appeared tightly coordinated with that of SAR86 gammaproteobacteria (Supplement Fig. S3, S4). We tested pairwise correlations of alpha- and gammaproteobacterial groups and found that SAR11 and SAR86 transcript accumulation were highly correlated in M1 and the Nouméa lagoon (Supplement Fig. S6, Pearson correlation: 0.88/0.96, Spearman rank correlation: 0.80/0.98 for M1 and lagoon, respectively). This matches recent observations in both coastal and pelagic ecosystems for coupling of SAR11 and SAR86 gene expression throughout a diel cycle, suggesting specific biological interactions between these two groups (Aylward et al., 2015). The fact that we now see this correlation over three weeks in two replicate experiments (M1 and Nouméa lagoon sampling) strengthens this hypothesis substantially. On the other hand, transcriptional activity was decoupled from 16S based abundance estimates for both clades (Pfreundt et al., 2016). Decoupling of specific activity and cell abundance has been noted before for SAR11, with specific activity being lower than cell abundance in the North Pacific (Hunt et al., 2013). It was further reported from microcosm experiments that proteorhodopsin transcripts increased under continuous light while gene abundance decreased (Lami et al., 2009). No such information is available for SAR86 in the literature, and the reasons for this decoupling remain elusive. We found no hint in the transcriptional profile that could explain the burst in SAR11 16S tags in M1 on day 8 (from 8 % to 26 %) and in the lagoon on day 16 (from 5 % to 28 %) (Pfreundt et al., 2016).

3.2.2 Gene expression of Oligotrophic Marine Gammaproteobacteria (OMG) and *Alteromonadaceae*

A closer look into gammaproteobacterial activities (Supplement Fig. S3C, S4C) revealed a dominant pool of transcripts from the oligotrophic Marine Gammaproteobacteria (OMG) group (Cho and Giovannoni, 2004; Spring et

al., 2013) and *Alteromonadaceae,* both in M1 and the lagoon. The temporal dynamics of OMG group transcripts were very similar in both locations. The relatively high activity detected for OMG bacteria (similar to SAR11 activity) indicated that these aerobic anoxygenic photoheterotrophs (Spring et al., 2013) could thrive well both in M1 and the Nouméa lagoon at the start of the experiment. However, transcript accumulation declined constantly by 70 % - 90 % until P2, then increased again during P2 until the end of the experiment, concurrent with *Synechococcus* activity and abundance. This pattern clearly decouples OMG group activity from phosphate availability, which was much higher in M1 than in the lagoon during P1, and also from the identity of the dominant diazotroph, which differed markedly between M1 and the lagoon in P2 (Turk-Kubo et al., 2015).

Alteromonadaceae–related transcript accumulation increased >2.5-fold in M1 in the first half of phase P1, replacing the initially dominating OMG and SAR86 as the most active groups within the gammaproteobacteria, but dropping to initial values in the second half of P1 (Supplement Fig. S3C). A burst in *Alteromonas* was previously reported as a confinement effect when a marine mixed microbial population was enclosed in mesocosms of smaller volumes (Schäfer et al., 2000). Immediately following the increase of *Alteromonadaceae*-related transcripts and reaching similar abundances, *Idiomarinaceae* transcripts increased 9-fold (Supplement Fig. S3C). This group of organotrophs is phylogenetically related to *Alteromonadaceae*. Gammaproteobacteria such as the *Alteromonadaceae* occur usually at rather low abundances in oligotrophic systems, but due to their copiotrophic metabolism (Ivars-Martinez et al., 2008; López-Pérez et al., 2012) increase in numbers and activity under eutrophic conditions or when particulate organic matter becomes available (García-Martínez et al., 2002; Ivars-Martinez et al., 2008). The fact that bacterial production (measured by ^{3}H-leucine assimilation) in M1 was not limited by phosphate (Van Wambeke et al., 2016) suggests that the DIP spike in the evening of day 4 was not responsible for these observations, but rather that both, *Alteromonadaceae* and *Idiomarinaceae,* reacted to nutrients released after the *Rhodobacteraceae* bloom on day 4 and the possibly virus-induced lysis of *Synechococcus*. *Idiomarinaceae* and *Alteromonadaceae* were transcriptionally very active compared to their 16S tag-based abundance estimates (Pfreundt et al., 2016), pointing at a tight regulation of their metabolic activities as a response to the appearance of suitable energy and nutrient sources.

3.2.3 Subdominant gene expression: Microalgae, *Flavobacteria* and *Spirotricha*

Other groups following the dominant classes *Oscillatoriophycideae*, *Alpha*-, and *Gammaproteobacteria* regarding transcript abundance in M1 and in the Nouméa lagoon were *Flavobacteria*, and the eukaryotic *Haptophyceae* (*Prymnesiophyceae*), *Chromerida* and *Spirotricha* (Supplement Figs. S3A and S5B, respectively). *Chromerida* are photoautotrophic alveolates and closely related to apicomplexan parasites. *Chromerida* have been isolated only from stony corals in Australian waters thus far (Moore et al., 2008; Oborník et al., 2012). Our finding of significant gene expression from *Chromerida* in samples from M1 (Supplement Fig. S3A), as well as the Nouméa lagoon (Supplement Fig. S5B) indicates they were present in substantial numbers in the open water. These findings are consistent with the predicted wider distribution, higher functional and taxonomic diversity of chromerid algae (Oborník et al., 2012).

3.2.4 Gene expression from nitrogen-fixing cyanobacteria

We specifically examined the gene expression patterns of diazotrophic cyanobacteria (**Fig. 5**) and compared them with parallel analyses of *nifH* amplicon sequences (Turk-Kubo et al., 2015), whereas heterotrophic diazotrophs were

orders of magnitude less abundant (Pfreundt et al., 2016; Turk-Kubo et al., 2015) and not further considered. The *nifH* amplicon analyses demonstrated in M1 a shift from a diazotroph community composed primarily of *Richelia* (diatom–diazotroph associations, DDAs, (Foster et al., 2011)) and *Trichodesmium* during P0 and P1 (days 2 to 14) to approximately equal contributions from UCYN-C (unicellular N_2-fixing cyanobacteria type C) and *Richelia* in phase P2 (days 15 to 23, (Turk-Kubo et al., 2015)). This shift was not observed outside. Consistent with these findings in M1, we also measured dominant gene expression from *Trichodesmium* and *Richelia* spp. until day 14, and Candidatus *Atelocyanobacterium thalassa* (UCYN-A) until day 8, and an increase in transcripts mapping to the *Epithemia turgida* endosymbiont and *Cyanothece* sp. ATCC51142 (**Fig. 5A**), classified within the UCYN-C *nifH* group (Nakayama et al., 2014), in P2. The temporal dynamics of gene expression for *Richelia* spp. and *Trichodesmium* in M1 differed with *Trichodesmium* gene expression declining by >97 % from initiation of the experiment to day 12, while the gene expression from *Richelia* species was stable until day 10 and then declined by ~90 % until day 16 (**Fig. 5A**). Except for the high (in relation to other diazotrophs) *Trichodesmium* transcript abundances on day 2 and 4, this matched well the *nifH*-gene based reports (Turk-Kubo et al., 2015). Results are also consistent for the lagoon samples, where transcripts from *Trichodesmium*, *Richelia* and UCYN-A dominated the diazotroph transcript pool throughout the experiment and no UCYN-C transcripts were observed (**Fig. 5B**). Again, the high relative *Trichodesmium* transcript abundances between day 2 and 12 were not mirrored by *nifH*-gene counts, while the rest was (Turk-Kubo et al., 2015). Noteworthy, *Trichodesmium* transcripts were one order of magnitude lower in the lagoon than in M1.

3.3 Specific analysis of relevant transcripts

The 100 most highly expressed non-ribosomal transcripts, as identified by highest mean expression in all samples, are presented in **Supplementary Table S1**. 24 of them could not be classified with the NCBI nucleotide or protein databases and remain unknown. The most abundant transcript overall, both inside M1 and outside, was the non-protein coding RNA (ncRNA) Yfr103, discussed in section 3.4. All classified transcripts on the top 9 ranks plus 28 additional transcripts in M1 were related to *Synechococcus* and encoded mainly photosynthetic proteins or represented ncRNAs. The transcripts following on ranks 10 to 12 in M1 plus five additional transcripts were affiliated with the recently defined new class of marine *Euryarchaeota Candidatus Thalassoarchaea* (Martin-Cuadrado et al., 2015), consistent with their detection by 16S analysis (Pfreundt et al., 2016). Other top expressed transcripts were *rnpB* from various bacteria, one tmRNA, and transcripts originating from the *Rhodobacteraceae* solely due to their expression peak on day 4. We also detected three different abundant antisense-RNAs (asRNAs), among them one transcribed from the complementary strand of *Synechococcus* gene Syncc8109_1164, encoding a hypothetical protein.

3.3.1 Gene expression relevant for nitrogen assimilation

To investigate gene-specific expression patterns, we analyzed genes of interest (GOIs) from specific genera. Transcripts mapping to the respective genes from different organisms were extracted, searched against NCBI's non-redundant protein database, and the hits were manually curated. This analysis was only performed for the M1 samples.

Genes indicative of different nitrogen utilization strategies are shown in **Fig. 6**. The selected GOIs were related to nitrogen fixation, nitrate and nitrite reduction, the uptake and assimilation of ammonia (transporter AmtA and glutamine synthetase, *glnA* gene product), and the assimilation of urea. Signal transducer PII (*glnB* gene product) and NtcA (nitrogen control transcription factor) were chosen as examples for the most important regulatory factors (Forchhammer, 2008; Huergo et al., 2013; Lindell and Post, 2001).

For *Trichodesmium* and UCYN-A (*Candidatus Atelocyanobacterium*), the core genes of the nitrogenase enzyme *nifHDK* were maximally expressed around day 4, while UCYN-C and *Chromatiaceae* (gammaproteobacteria) *nif* expression peaked on day 20 (**Fig. 6,** *nitrogen fixation*). As nitrogenase gene expression and activity is under diel control, the expression patterns we obtained can only represent diazotrophs that fix N_2 during the light hours because we sampled in the morning. *nifH* phylogeny places the endosymbiotic diazotroph of *Rhopalodia gibba* within the UCYN-C group, but oppositely to other UCYN-C, these endosymbionts were shown to fix N_2 in the light (Prechtl et al., 2004), explaining why our analysis captured their *nif* transcripts, but none mapping to *Cyanothece* sp. ATCC 51142. The NifH and NifD protein sequences of the *R. gibba* endosymbiont and of *Epithemia turgida*, for which we report substantial gene expression in section 3.2.4, are 98 % identical, making it likely that the partial *nif* transcript contigs reported here could not be assigned unambiguously and probably belong to the *Epithemia turgida* symbiont or a close relative of both. Our data correlates with *nifH* gene abundance measured by Turk-Kubo et al. (2015). Note, that *nif* transcripts were generally very rare in this analysis (at maximum 2 TPM), making it very likely that below a certain expression threshold, a transcript was not sequenced at all.

Most other bacteria require ammonia, nitrate, or organic nitrogen sources such as urea, with ammonia being the energetically most favourable source. The importance of ammonia was underscored by expression of the respective uptake systems in many different taxa over long periods of the experiment and expression of glutamine synthetase (GS) (**Fig. 6**, *ammonium transporter* and *glutamine synthetase*), the enzyme forming the central point-of-entry for the newly assimilated nitrogen into the metabolism. Ammonia transporters (AMT) were highly expressed in the *Pelagibacteraceae*, throughout days 2 to 20, in *Synechococcus* from day 10 to the end, and in *Rhodobacteraceae* on day 4 (coinciding with maximum GS expression, and the global gene expression peak in this group, Supplement Fig. S3C). Interestingly, for the *Halieaceae* (OM60(NOR5) clade), the dominant family within the OMG group, AMT and GS expression did not coincide with their general expression peaks on day 2. Instead, these genes as well as the signal transducer PII were mainly expressed on day 23, indicating that *Halieaceae* were nitrogen limited toward the end of the experiment. Interestingly, *Alteromonadaceae* did not express either of these genes maximally on day 8, when they were reaching their highest abundance and transcription. Instead, they expressed urease, constituting the highest measured urease expression in the whole experiment (**Fig. 6**, *urease subunit alpha*). Urea can serve as an alternative nitrogen source and is metabolized into ammonia. A shift towards urea utilization was also seen in *Rhodobacteraceae*. While most of the N-utilization transcripts analysed here peaked on day 4 (coinciding with general expression and abundance peaks), urease expression in this group was highest from day 10 to 14 (**Fig. 6**, *urease subunit alpha*). Further, *Prochlorococcus* but not *Synechococcus* expressed urease, and both expressed ammonium transporters.

Nitrate reductase expression was detected on day 10 mainly in *Synechococcus* and in the *Halomonadaceae*. It is not clear why this gene is so strongly expressed on that single day, especially as nitrite reductase expression in *Synechococcus* was detectable over a longer period, from day 12 to day 20 (**Fig. 6**, *nitrate reductase* and *nitrite*

reductase). Other taxa showed substantial nitrite reductase expression only on day 6 (*Vibrionaceae*), day 2 and 14-16 (*Rhodobacteraceae*), or day 10 (SAR116).

The expression of the NtcA transcription factor itself can be an indicator for the nitrogen status, especially in marine picocyanobacteria (Lindell and Post, 2001; Tolonen et al., 2006). Therefore, the clear peaks for NtcA expression in *Prochlorococcus* on day 12 to 14, but in *Synechococcus* on day 20 indicate their divergent relative nitrogen demands (**Fig. 6**, *NtcA*). Noteworthy, *ntcA* expression was much stronger in *Prochlorococcus* than in *Synechococcus* compared to their 16S-based abundances and *Prochlorococcus* did not express it during its first abundance peak on day 6, but only during the second one (Van Wambeke et al., 2016). The same was true for *Synechococcus* and its first abundance peak on day 12.

3.3.2 Expression of genes involved in the assimilation of phosphate and light utilization

In addition to nitrogen, genes involved in phosphate assimilation were analysed in more detail. Expression of alkaline phosphatase (AP) was prominent between day 12 and 23 in different organisms (**Fig. 7**, *alkaline phosphatase*) and not expressed before the DIP fertilization, although phosphate levels were similar before the fertilization event and after day 13 (Pfreundt et al., 2016), and phosphate turnover time reached pre-fertilization levels after day 20 (Berthelot et al., 2015). *Alteromonadaceae* increased AP expression steadily from day 10 to 14. This suggests that the microbial community was initially adapted to the ambient phosphate levels and not phosphate limited, and that the post-fertilization community had to actively acquire P to fulfil their quota. In a companion paper N-limitation, but not P-limitation, was evident for heterotrophic bacteria throughout the experiment (Van Wambeke et al., 2016). The dominant photoautotroph, *Synechococcus*, expressed the gene for the sulfolipid biosynthesis protein SqdB, in agreement with *Synechococcus* abundance (**Fig. 7B**, *sulfolipid biosynthesis*). Van Mooy and colleagues ((Van Mooy et al., 2006, 2009)) suggested that marine picocyanobacteria could minimize their phosphorus requirements through the synthesis of sulphoquinovosyldiacyl glycerols, for which SqdB is required, and substitute phospholipids. Therefore, our finding of a high expression of the *Synechococcus sqdB* gene, especially towards the end of the experiment is consistent with this idea and with the increasing *Synechococcus* cell count towards the end of the experiment, when phosphate became limiting again (Pfreundt et al., 2016).

TonB-dependent transport allows large molecules to pass through the membrane. This strategy of exploiting larger molecules as nutrient sources is thought to be prevalent in SAR86 bacteria (Dupont et al., 2012) and indeed we found the highest expression of *tonB* genes for SAR86 at the beginning of the experiment, when phosphate was low. Despite increasing SAR86 cell numbers towards day 10 (Van Wambeke et al., 2016), *tonB* expression decreased after the DIP spike, indicating a role in phosphate acquisition for the TonB transporters in SAR86. Interestingly, *Halieaceae* expressed *tonB* genes together with patatin phospholipase on day 2 and weaker on day 23, indicating that phospholipids were utilized as a phosphate source prior to DIP fertilization when DIP availability was limited, and again at the end of the experiment when DIP was depleted again.

Proteorhodopsin was highly expressed, especially by SAR11 (*Pelagibacter*), underlining the importance of light as an additional source of ATP for this group. Bacteriorhodopsin gene expression was reported to depend on the ambient light conditions in several different bacteria, including Flavobacteria and SAR11 (Gómez-Consarnau et al., 2010; Kimura et al., 2011; Lami et al., 2009). This is consistent with our observation of upregulated proteorhodopsin

expression towards the end of the experiment in the *Pelagibacteraceae* (**Fig. 7B**). There were also archaeal rhodopsins expressed, but at a ~two orders of magnitude lower level.

3.4 Highly expressed non-coding RNA in picocyanobacteria

Although unexplored in non-model bacteria, ncRNAs can play important regulatory roles, e.g. in cyanobacteria in the adaptation of the photosynthetic appratus to high light intensities (Georg et al., 2014) or of the nitrogen assimilatory machinery to nitrogen limitation (Klähn et al., 2015). During the analysis of the 100 transcripts with the highest mean abundance, we found that 14 of these transcripts corresponded to the recently identified non-coding RNA (ncRNA) Yfr103. The Yfr103 transcripts were mapped to 14 different loci, from these 12 could be assigned to *Synechococcus*, one to *Prochlorococcus,* and one to picocyanobacteria as it was equally similar to both genera. The expression of these 14 Yfr103 isoforms was plotted over the time course of the experiment, indicating their almost constitutive expression (Supplement Fig. S7). Yfr103 was described first as the single most abundant ncRNA in laboratory cultures of *Prochlorococcus* MIT9313 (Voigt et al., 2014) and here we show Yfr103 is the single most abundant transcript over the entire microbial population. These data suggest an important function of this ncRNA in marine *Synechococcus* and *Prochlorococcus*.

4 Conclusions

Here we have studied how mesocosm confinement and DIP fertilization influenced transcriptional activities of the microbial community during the VAHINE experiment in the South West Pacific. One of the most pronounced effects we observed was transcript diversification within the mesocosm, pointing to induced transcriptional responses in several taxonomic groups compared to a more stable transcript pool in the lagoon. Despite this diversification, analysis of differentially expressed transcripts amongst time points showed that global transcriptional changes roughly followed the time line in both M1 and the lagoon. This confirms results from 16S based community analysis, where time was shown to be the factor most strongly influencing bacterial succession in both locations. Gene expression inside M1 was dominated by alphaproteobacteria until day 12, with *Rhodobacteraceae* exhibiting a prominent peak on day 4. This was followed by a burst in *Alteromonadaceae*-related gene expression on days 8 and 10 and a peak in transcript abundance from *Idiomarinaceae* on day 10 in rapid succession. In the lagoon, *Synechococcus* transcripts were the most abundant throughout the experiment, and similar abundances were reached in M1 only in P2. We further observed a tight coupling between gene expression of SAR86 and SAR11 over the whole experiment. Such coupling has been observed previously during the diel cycle (Aylward et al., 2015), whereas we have observed this phenomenon here for a longer period of time, both inside and outside (Supplement Fig. S6). Such concerted activity changes between taxonomically distinct groups should affect biogeochemical transformations and should be governed by structured ecological conditions. However, the environmental determinants driving this coupling remain to be identified.

The specific gene expression of diazotrophic cyanobacteria could be mainly attributed to *Trichodesmium* and *Richelia intracellularis* strains (diatom-diazotroph associations). UCYN-A (*Candidatus* Atelocyanobacterium) transcripts

were the third most abundant class coming from diazotrophic cyanobacteria and declined both inside and outside after day 4, consistent with both 16S- and *nifH*-based analyses. Transcripts related to the *Epithemia turgida* endosymbiont and *Cyanothece* ATCC 51142 increased from day 14 and maintained a higher share until the end of the experiment, day 23, consistent with the observed increase in UCYN-C *nifH* tags after day 14 in M1. Hence, we conclude that a relative of the *Epithemia turgida* endosymbiont is the main contributor to UCYN-C N_2 fixing cyanobacteria.

Data availability

All raw sequencing data can be downloaded from NCBI's Short Read Archive (SRA) under the accession number PRJNA304389.

Acknowledgements

The authors thank the captain and crew of the R/V Alis for their great support during the cruise, to Eyal Rahav throughout the experimental campaign, to Brian Haas (The Broad Institute) and Ben Fulton (Indiana University) for their patient help with the Trinity *de-novo* transcriptome assembler, Craig Nelson University of Hawaii at Manoa) for his valuable advice on multivariate statistics, Björn Grüning (University of Freiburg) for his immediate actions regarding tool updates and bugs on the local Galaxy server, and Steffen Lott (University of Freiburg) for his help with the visualization tool CoVennTree. The participation of UP, WRH, DS and IBF in the VAHINE experiment was supported by the German-Israeli Research Foundation (GIF), project number 1133-13.8/2011, BSF grant 2008048 to IBF and the metatranscriptome analysis by the EU project MaCuMBA (Marine Microorganisms: Cultivation Methods for Improving their Biotechnological Applications; grant agreement no: 311975) to WRH.

Author Contributions

SB conceived and designed the experiment. IBF took part in experimental planning, preparation, and implementation. UP, WRH, DS, and IBF participated in experiment and sampled. UP analysed samples and prepared all figures. WRH and UP wrote the manuscript and all authors contributed and revised the manuscript.

References

Aylward, F. O., Eppley, J. M., Smith, J. M., Chavez, F. P., Scholin, C. A. and DeLong, E. F.: Microbial community transcriptional networks are conserved in three domains at ocean basin scales, Proc. Natl. Acad. Sci. U. S. A., 112(17), 5443–5448, doi:10.1073/pnas.1502883112, 2015.

Berthelot, H., Moutin, T., L'Helguen, S., Leblanc, K., Hélias, S., Grosso, O., Leblond, N., Charrière, B. and Bonnet, S.: Dinitrogen fixation and dissolved organic nitrogen fueled primary production and particulate export during the VAHINE mesocosms experiment (New Caledonia lagoon), Biogeosciences, 12(5), 4099–4112, doi:10.5194/bg-12-4099-2015, 2015.

Bonnet, S., Berthelot, H., Turk-Kubo, K., Fawcett, S., Rahav, E., l'Helguen, S. and Berman-Frank, I.: Dynamics of N2 fixation and fate of diazotroph-derived nitrogen in a low nutrient low chlorophyll ecosystem: results from the VAHINE mesocosm experiment (New Caledonia), Biogeosciences, 13, 2653–2673, doi:10.5194/bg-13-2653-2016, 2016a.

Bonnet, S., Moutin, T., Rodier, M., Grisoni, J. M., Louis, F., Folcher, E., Bourgeois, B., Boré, J. M. and Renaud, A.: Introduction to the project VAHINE: VAriability of vertical and tropHIc transfer of fixed N2 in the south wEst Pacific, Biogeosciences Discuss., this issue, submitted, 2016b.

Buchfink, B., Xie, C. and Huson, D. H.: Fast and sensitive protein alignment using DIAMOND, Nat. Methods, 12(1), 59–60, doi:10.1038/nmeth.3176, 2015.

Cho, J.-C. and Giovannoni, S. J.: Cultivation and growth characteristics of a diverse group of oligotrophic marine Gammaproteobacteria, Appl. Environ. Microbiol., 70(1), 432–440, 2004.

Dufresne, A., Ostrowski, M., Scanlan, D. J., Garczarek, L., Mazard, S., Palenik, B. P., Paulsen, I. T., de Marsac, N. T., Wincker, P., Dossat, C., Ferriera, S., Johnson, J., Post, A. F., Hess, W. R. and Partensky, F.: Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria, Genome Biol., 9(5), R90, doi:10.1186/gb-2008-9-5-r90, 2008.

Dupont, C. L., Rusch, D. B., Yooseph, S., Lombardo, M.-J., Alexander Richter, R., Valas, R., Novotny, M., Yee-Greenbaum, J., Selengut, J. D., Haft, D. H., Halpern, A. L., Lasken, R. S., Nealson, K., Friedman, R. and Craig Venter, J.: Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage, ISME J., 6(6), 1186–1199, doi:10.1038/ismej.2011.189, 2012.

Edgar, R. C.: Search and clustering orders of magnitude faster than BLAST, Bioinformatics, 26(19), 2460–2461, doi:10.1093/bioinformatics/btq461, 2010.

Forchhammer, K.: P(II) signal transducers: novel functional and structural insights, Trends Microbiol., 16(2), 65–72, doi:10.1016/j.tim.2007.11.004, 2008.

Foster, R. A., Kuypers, M. M. M., Vagner, T., Paerl, R. W., Musat, N. and Zehr, J. P.: Nitrogen fixation and transfer in open ocean diatom-cyanobacterial symbioses, ISME J., 5(9), 1484–1493, doi:10.1038/ismej.2011.26, 2011.

Frank, J. A., Lorimer, D., Youle, M., Witte, P., Craig, T., Abendroth, J., Rohwer, F., Edwards, R. A., Segall, A. M. and Burgin, A. B.: Structure and function of a cyanophage-encoded peptide deformylase, ISME J., 7(6), 1150–1160, doi:10.1038/ismej.2013.4, 2013.

Frias-Lopez, J., Shi, Y., Tyson, G. W., Coleman, M. L., Schuster, S. C., Chisholm, S. W. and Delong, E. F.: Microbial community gene expression in ocean surface waters, Proc. Natl. Acad. Sci. U. S. A., 105(10), 3805–3810, doi:10.1073/pnas.0708897105, 2008.

Ganesh, S., Bristow, L. A., Larsen, M., Sarode, N., Thamdrup, B. and Stewart, F. J.: Size-fraction partitioning of community gene transcription and nitrogen metabolism in a marine oxygen minimum zone, ISME J., doi:10.1038/ismej.2015.44, 2015.

García-Martínez, J., Acinas, S. G., Massana, R. and Rodríguez-Valera, F.: Prevalence and microdiversity of Alteromonas macleodii-like microorganisms in different oceanic regions, Environ. Microbiol., 4(1), 42–50, 2002.

Georg, J., Dienst, D., Schürgers, N., Wallner, T., Kopp, D., Stazic, D., Kuchmina, E., Klähn, S., Lokstein, H., Hess, W. R. and Wilde, A.: The Small Regulatory RNA SyR1/PsrR1 Controls Photosynthetic Functions in Cyanobacteria, Plant Cell Online, 26(9), 3661–3679, doi:10.1105/tpc.114.129767, 2014.

Gifford, S. M., Sharma, S. and Moran, M. A.: Linking activity and function to ecosystem dynamics in a coastal bacterioplankton community, Front. Microbiol., 5, doi:10.3389/fmicb.2014.00185, 2014.

Gómez-Consarnau, L., Akram, N., Lindell, K., Pedersen, A., Neutze, R., Milton, D. L., González, J. M. and Pinhassi, J.: Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation, PLoS Biol., 8(4), e1000358, doi:10.1371/journal.pbio.1000358, 2010.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., Macmanes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., Leduc, R. D., Friedman, N. and Regev, A.: De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis, Nat. Protoc., 8(8), 1494–1512, doi:10.1038/nprot.2013.084, 2013.

Hewson, I., Poretsky, R. S., Tripp, H. J., Montoya, J. P. and Zehr, J. P.: Spatial patterns and light-driven variation of microbial population gene expression in surface waters of the oligotrophic open ocean, Environ. Microbiol., 12(7), 1940–1956, doi:10.1111/j.1462-2920.2010.02198.x, 2010.

Hilton, J. A., Satinsky, B. M., Doherty, M., Zielinski, B. and Zehr, J. P.: Metatranscriptomics of N2-fixing cyanobacteria in the Amazon River plume, ISME J., 9(7), 1557–1569, doi:10.1038/ismej.2014.240, 2015.

Huergo, L. F., Chandra, G. and Merrick, M.: P(II) signal transduction proteins: nitrogen regulation and beyond, FEMS Microbiol. Rev., 37(2), 251–283, doi:10.1111/j.1574-6976.2012.00351.x, 2013.

Hunt, D. E., Lin, Y., Church, M. J., Karl, D. M., Tringe, S. G., Izzo, L. K. and Johnson, Z. I.: Relationship between Abundance and Specific Activity of Bacterioplankton in Open Ocean Surface Waters, Appl. Environ. Microbiol., 79(1), 177–184, doi:10.1128/AEM.02155-12, 2013.

Huson, D. H. and Weber, N.: Microbial community analysis using MEGAN, Methods Enzymol., 531, 465–485, doi:10.1016/B978-0-12-407863-5.00021-6, 2013.

Ivars-Martinez, E., Martin-Cuadrado, A.-B., D'Auria, G., Mira, A., Ferriera, S., Johnson, J., Friedman, R. and Rodriguez-Valera, F.: Comparative genomics of two ecotypes of the marine planktonic copiotroph Alteromonas macleodii suggests alternative lifestyles associated with different kinds of particulate organic matter, ISME J., 2(12), 1194–1212, doi:10.1038/ismej.2008.74, 2008.

Jones, D. S., Flood, B. E. and Bailey, J. V.: Metatranscriptomic insights into polyphosphate metabolism in marine sediments, ISME J., doi:10.1038/ismej.2015.169, 2015.

Kimura, H., Young, C. R., Martinez, A. and Delong, E. F.: Light-induced transcriptional responses associated with proteorhodopsin-enhanced growth in a marine flavobacterium, ISME J., 5(10), 1641–1651, doi:10.1038/ismej.2011.36, 2011.

Klähn, S., Schaal, C., Georg, J., Baumgartner, D., Knippen, G., Hagemann, M., Muro-Pastor, A. M. and Hess, W. R.: The sRNA NsiR4 is involved in nitrogen assimilation control in cyanobacteria by targeting glutamine synthetase inactivating factor IF7, Proc. Natl. Acad. Sci. U. S. A., doi:10.1073/pnas.1508412112, 2015.

Knapp, A. N., Fawcett, S. E., Martinez-Garcia, A., Haug, G., Leblond, N., Moutin, T. and Bonnet, S.: Nitrogen isotopic evidence for a shift from nitrate- to diazotroph-fueled export production in VAHINE mesocosm experiments, Biogeosciences Discuss., 12, 19901–19939, 2015.

Kopylova, E., Noé, L. and Touzet, H.: SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data, Bioinforma. Oxf. Engl., 28(24), 3211–3217, doi:10.1093/bioinformatics/bts611, 2012.

Lami, R., Cottrell, M. T., Campbell, B. J. and Kirchman, D. L.: Light-dependent growth and proteorhodopsin expression by Flavobacteria and SAR11 in experiments with Delaware coastal waters, Environ. Microbiol., 11(12), 3201–3209, doi:10.1111/j.1462-2920.2009.02028.x, 2009.

Langmead, B.: Aligning short sequencing reads with Bowtie, Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al, Chapter 11, Unit 11.7, doi:10.1002/0471250953.bi1107s32, 2010.

Leblanc, K., Cornet, V., Caffin, M., Rodier, M., Desnues, A., Berthelot, H., Turk-Kubo, K. A. and Heliou, J.: Phytoplankton community structure in the VAHINE mesocosm experiment, Biogeosciences Discuss., 1–34, doi:10.5194/bg-2015-605, 2016.

Li, B. and Dewey, C. N.: RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, BMC Bioinformatics, 12, 323, doi:10.1186/1471-2105-12-323, 2011.

Lindell, D. and Post, A. F.: Ecological aspects of ntcA gene expression and its use as an indicator of the nitrogen status of marine Synechococcus spp, Appl. Environ. Microbiol., 67(8), 3340–3349, doi:10.1128/AEM.67.8.3340-3349.2001, 2001.

López-Pérez, M., Gonzaga, A., Martin-Cuadrado, A.-B., Onyshchenko, O., Ghavidel, A., Ghai, R. and Rodriguez-Valera, F.: Genomes of surface isolates of Alteromonas macleodii: the life of a widespread marine opportunistic copiotroph, Sci. Rep., 2, 696, doi:10.1038/srep00696, 2012.

Lott, S. C., Voß, B., Hess, W. R. and Steglich, C.: CoVennTree: a new method for the comparative analysis of large datasets, Front. Genet., 6, 43, doi:10.3389/fgene.2015.00043, 2015.

Luo, Y.-W., Doney, S. C., Anderson, L. A., Benavides, M., Berman-Frank, I., Bode, A., Bonnet, S., Boström, K. H., Böttjer, D., Capone, D. G., Carpenter, E. J., Chen, Y. L., Church, M. J., Dore, J. E., Falcón, L. I., Fernández, A., Foster, R. A., Furuya, K., Gómez, F., Gundersen, K., Hynes, A. M., Karl, D. M., Kitajima, S., Langlois, R. J., LaRoche, J., Letelier, R. M., Marañón, E., McGillicuddy, D. J., Moisander, P. H., Moore, C. M., Mouriño-Carballido, B., Mulholland, M. R., Needoba, J. A., Orcutt, K. M., Poulton, A. J., Rahav, E., Raimbault, P., Rees, A. P., Riemann, L., Shiozaki, T., Subramaniam, A., Tyrrell, T., Turk-Kubo, K. A., Varela, M., Villareal, T. A., Webb, E. A., White, A. E., Wu, J. and Zehr, J. P.: Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates, Earth Syst. Sci. Data, 4(1), 47–73, doi:10.5194/essd-4-47-2012, 2012.

Ma, Y., Allen, L. Z. and Palenik, B.: Diversity and genome dynamics of marine cyanophages using metagenomic analyses, Environ. Microbiol. Rep., 6(6), 583–594, 2014.

Markowitz, V. M., Chen, I.-M. A., Chu, K., Pati, A., Ivanova, N. N. and Kyrpides, N. C.: Ten Years of Maintaining and Expanding a Microbial Genome and Metagenome Analysis System, Trends Microbiol., doi:10.1016/j.tim.2015.07.012, 2015.

Martin-Cuadrado, A.-B., Garcia-Heredia, I., Moltó, A. G., López-Úbeda, R., Kimes, N., López-García, P., Moreira, D. and Rodriguez-Valera, F.: A new class of marine Euryarchaeota group II from the Mediterranean deep chlorophyll maximum, ISME J., 9(7), 1619–1634, doi:10.1038/ismej.2014.249, 2015.

Moore, R. B., Oborník, M., Janouskovec, J., Chrudimský, T., Vancová, M., Green, D. H., Wright, S. W., Davies, N. W., Bolch, C. J. S., Heimann, K., Slapeta, J., Hoegh-Guldberg, O., Logsdon, J. M. and Carter, D. A.: A photosynthetic alveolate closely related to apicomplexan parasites, Nature, 451(7181), 959–963, doi:10.1038/nature06635, 2008.

Moran, M. A., Satinsky, B., Gifford, S. M., Luo, H., Rivers, A., Chan, L.-K., Meng, J., Durham, B. P., Shen, C., Varaljay, V. A., Smith, C. B., Yager, P. L. and Hopkinson, B. M.: Sizing up metatranscriptomics, ISME J., 7(2), 237–243, doi:10.1038/ismej.2012.94, 2013.

Nakayama, T., Kamikawa, R., Tanifuji, G., Kashiyama, Y., Ohkouchi, N., Archibald, J. M. and Inagaki, Y.: Complete genome of a nonphotosynthetic cyanobacterium in a diatom reveals recent adaptations to an intracellular lifestyle, Proc. Natl. Acad. Sci. U. S. A., 111(31), 11407–11412, doi:10.1073/pnas.1405222111, 2014.

Oborník, M., Modrý, D., Lukeš, M., Cernotíková-Stříbrná, E., Cihlář, J., Tesařová, M., Kotabová, E., Vancová, M., Prášil, O. and Lukeš, J.: Morphology, ultrastructure and life cycle of Vitrella brassicaformis n. sp., n. gen., a novel chromerid from the Great Barrier Reef, Protist, 163(2), 306–323, doi:10.1016/j.protis.2011.09.001, 2012.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H. and Wagner, H.: vegan: Community Ecology Package. [online] Available from: http://CRAN.R-project.org/package=vegan, 2015.

Pfreundt, U., Miller, D., Adusumilli, L., Stambler, N., Berman-Frank, I. and Hess, W. R.: Depth dependent metatranscriptomes of the marine pico-/nanoplanktonic communities in the Gulf of Aqaba/Eilat during seasonal deep mixing, Mar. Genomics, 18 Pt B, 93–95, doi:10.1016/j.margen.2014.06.005, 2014.

Pfreundt, U., Van Wambeke, F., Bonnet, S., Hess, W. R. and Caffin, M.: Succession within the prokaryotic communities during the VAHINE mesocosms experiment in the New Caledonia lagoon, Biogeosciences, 13, 2319–2337, doi:10.5194/bg-13-2319-2016, 2016.

Pinto, F. L., Thapper, A., Sontheim, W. and Lindblad, P.: Analysis of current and alternative phenol based RNA extraction methodologies for cyanobacteria, BMC Mol. Biol., 10, 79, doi:10.1186/1471-2199-10-79, 2009.

Poretsky, R. S., Hewson, I., Sun, S., Allen, A. E., Zehr, J. P. and Moran, M. A.: Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre, Environ. Microbiol., 11(6), 1358–1375, doi:10.1111/j.1462-2920.2008.01863.x, 2009.

Prechtl, J., Kneip, C., Lockhart, P., Wenderoth, K. and Maier, U.-G.: Intracellular Spheroid Bodies of Rhopalodia gibba Have Nitrogen-Fixing Apparatus of Cyanobacterial Origin, Mol. Biol. Evol., 21(8), 1477–1481, doi:10.1093/molbev/msh086, 2004.

Scanlan, D. J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W. R., Post, A. F., Hagemann, M., Paulsen, I. and Partensky, F.: Ecological genomics of marine picocyanobacteria, Microbiol. Mol. Biol. Rev., 73(2), 249–299, doi:10.1128/MMBR.00035-08, 2009.

Schäfer, H., Servais, P. and Muyzer, G.: Successional changes in the genetic diversity of a marine bacterial assemblage during confinement, Arch. Microbiol., 173(2), 138–145, 2000.

Shi, Y., Tyson, G. W. and DeLong, E. F.: Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column, Nature, 459(7244), 266–269, doi:10.1038/nature08055, 2009.

Spring, S., Riedel, T., Spröer, C., Yan, S., Harder, J. and Fuchs, B. M.: Taxonomy and evolution of bacteriochlorophyll a-containing members of the OM60/NOR5 clade of marine gammaproteobacteria: description of Luminiphilus syltensis gen. nov., sp. nov., reclassification of Haliea rubra as Pseudohaliea rubra gen. nov., comb. nov., and emendation of Chromatocurvus halotolerans, BMC Microbiol., 13, 118, doi:10.1186/1471-2180-13-118, 2013.

Steglich, C., Stazic, D., Lott, S. C., Voigt, K., Greengrass, E., Lindell, D. and Hess, W. R.: Dataset for metatranscriptome analysis of Prochlorococcus-rich marine picoplankton communities in the Gulf of Aqaba, Red Sea, Mar. Genomics, 19, 5–7, 2015.

Tolonen, A. C., Aach, J., Lindell, D., Johnson, Z. I., Rector, T., Steen, R., Church, G. M. and Chisholm, S. W.: Global gene expression of Prochlorococcus ecotypes in response to changes in nitrogen availability, Mol. Syst. Biol., 2, 53, doi:10.1038/msb4100087, 2006.

Turk-Kubo, K. A., Frank, I. E., Hogan, M. E., Desnues, A., Bonnet, S. and Zehr, J. P.: Diazotroph community succession during the VAHINE mesocosms experiment (New Caledonia Lagoon), Biogeosciences, 12, 7435–7452, doi:10.5194/bg-12-7435-2015, 2015.

Van Mooy, B. A. S., Rocap, G., Fredricks, H. F., Evans, C. T. and Devol, A. H.: Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments, Proc. Natl. Acad. Sci. U. S. A., 103(23), 8607–8612, doi:10.1073/pnas.0600540103, 2006.

Van Mooy, B. A. S., Fredricks, H. F., Pedler, B. E., Dyhrman, S. T., Karl, D. M., Koblízek, M., Lomas, M. W., Mincer, T. J., Moore, L. R., Moutin, T., Rappé, M. S. and Webb, E. A.: Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity, Nature, 458(7234), 69–72, doi:10.1038/nature07659, 2009.

Van Wambeke, F., Pfreundt, U., Barani, A., Berthelot, H., Moutin, T., Rodier, M., Hess, W. R. and Bonnet, S.: Heterotrophic bacterial production and metabolic balance during the VAHINE mesocosm experiment in the New Caledonia lagoon, Biogeosciences, 13, 3187–3202, doi:10.5194/bg-13-3187-2016, 2016.

Voigt, K., Sharma, C. M., Mitschke, J., Joke Lambrecht, S., Voß, B., Hess, W. R. and Steglich, C.: Comparative transcriptomics of two environmentally relevant cyanobacteria reveals unexpected transcriptome diversity, ISME J., 8(10), 2056–2068, doi:10.1038/ismej.2014.57, 2014.

Wemheuer, B., Wemheuer, F., Hollensteiner, J., Meyer, F.-D., Voget, S. and Daniel, R.: The green impact: bacterioplankton response toward a phytoplankton spring bloom in the southern North Sea assessed by comparative metagenomic and metatranscriptomic approaches, Front. Microbiol., 6, 805, doi:10.3389/fmicb.2015.00805, 2015.

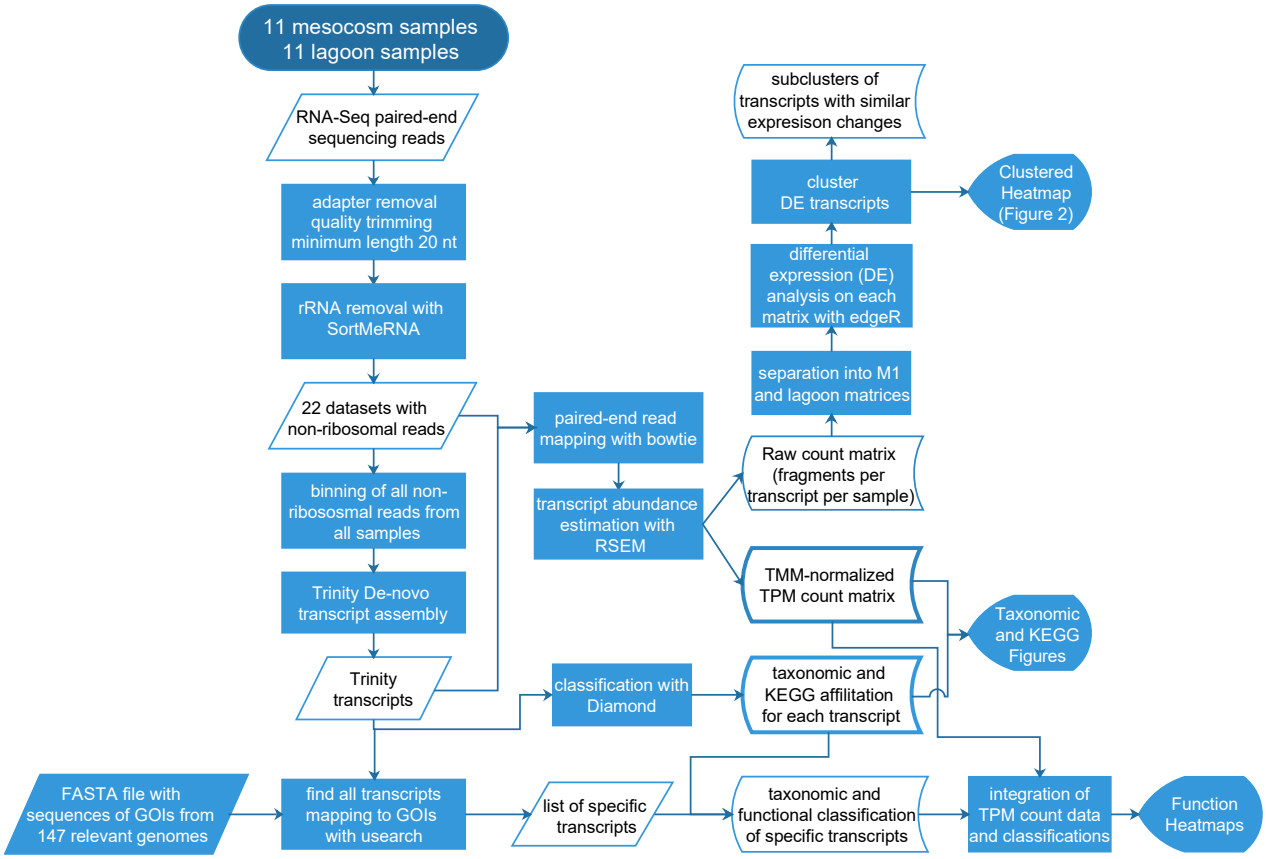

Figure 1. Flowchart describing the major steps in the bioinformatics workflow. Pre-processing of RNA-Seq reads was done separately for each dataset, leading to 22 datasets of non-ribosomal paired-end reads. These were binned and used as input for *de-novo* assembly of transcripts. The non-ribosomal reads were mapped back onto the assembled transcripts with bowtie (Langmead, 2010) to infer each transcripts abundance in each sample using RSEM (Li and Dewey, 2011). Raw abundances were used for differential expression (DE) analysis and cluster analysis with edgeR on the M1 and lagoon count matrices separately, to find transcripts which changed significantly over time. To enable direct in-between sample comparison of transcript abundances, raw abundances were converted to TPM (transcripts per kilobase million) and TMM-normalized (Trimmed Mean of M values) in RSEM, creating the final count matrix used for all figures showing transcript abundances. Classifications for these transcripts were generated using Diamond (Buchfink et al., 2015) against the RefSeq protein database. Further, a manually curated list with specific genes involved in N- and P- metabolism, as well as light capture (genes of interest, GOIs) was used to extract the corresponding transcripts, but final classifications were inferred from the Diamond output. This information was used to produce the integrated function-per-taxon heatmaps.

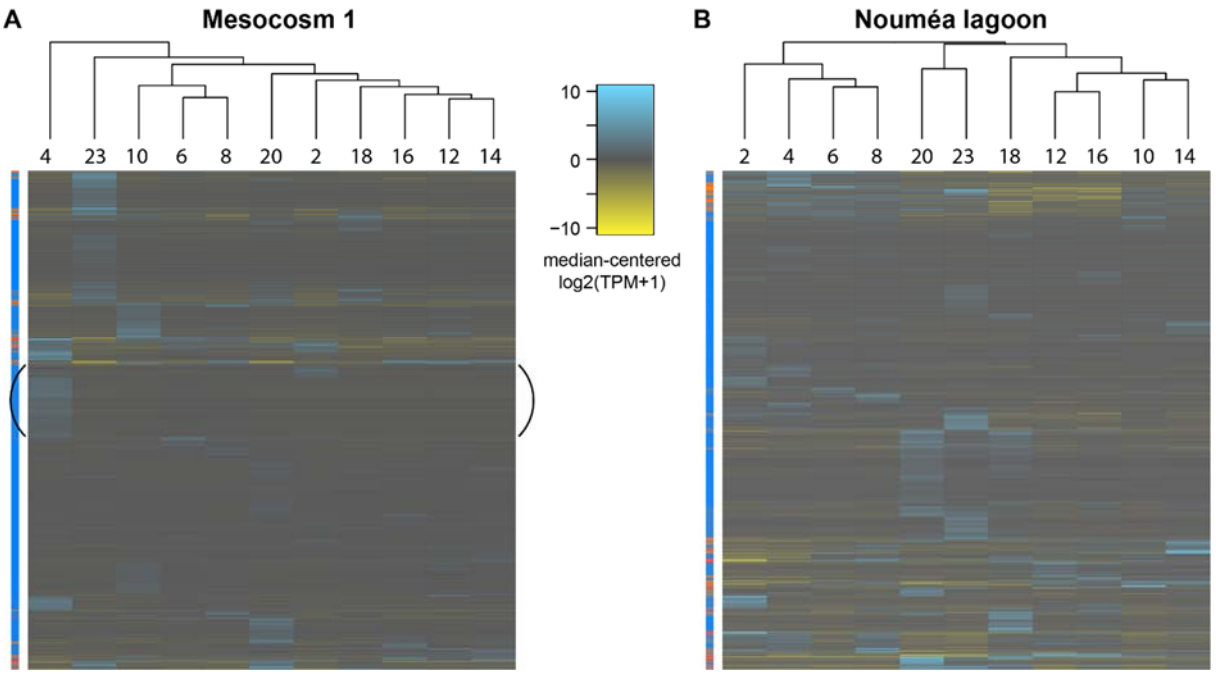

Figure 2. Heat map showing the expression (median-centered $\log_2$(TPM+1)) of all significantly differentially expressed transcripts in all samples taken from mesocosm M1 (A) and outside (B). Clustering of samples and transcripts was done using Euclidean distance measures followed by average agglomerative clustering (hclust(method=average)). Note that in (A) samples T2 and T4 cluster far away from the other samples. These were taken before the phosphate spike. In M1, T4 is distinguished by a large cluster of genes upregulated at that time point, most of which belong to the *Rhodobacteraceae* family. The general clustering along the timeline is evident inside and outside of M1.

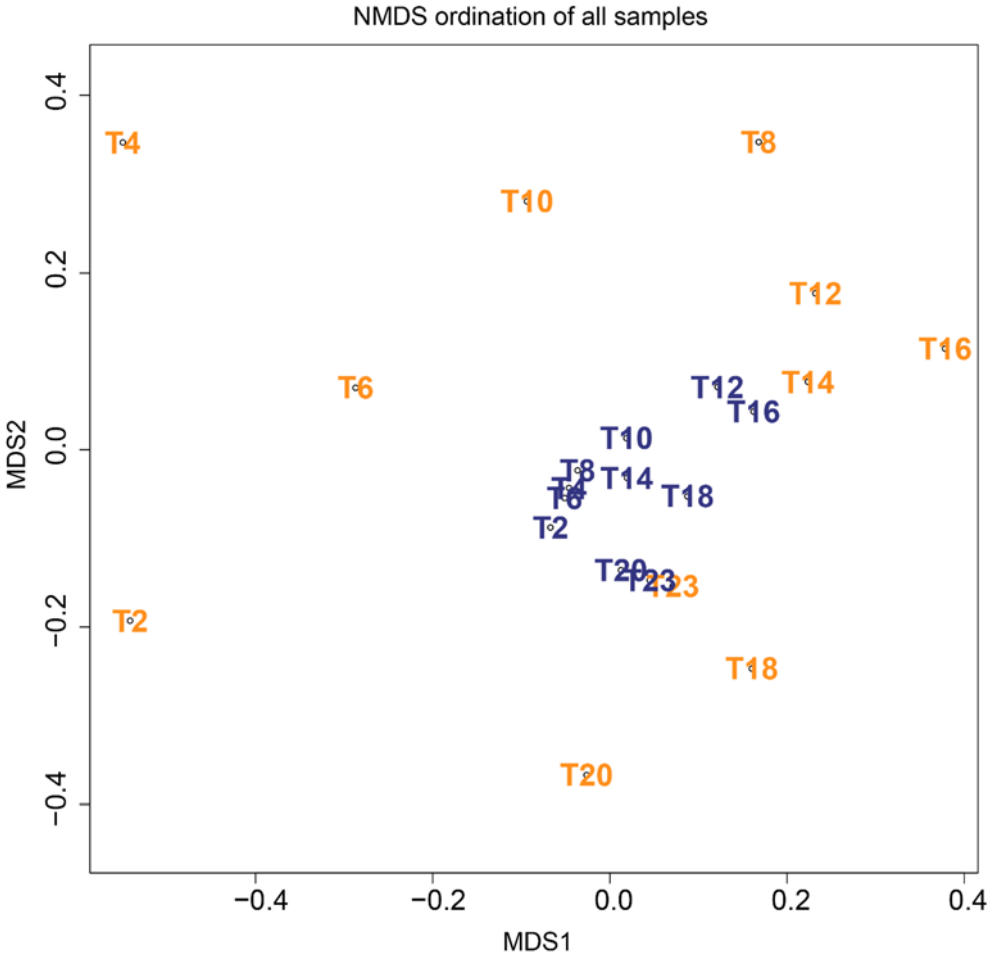

Figure 3. NMDS ordination of samples on the basis of TPM counts (transcripts per million sequenced transcripts). Outside samples are blue, samples from mesocosm M1 are orange. Note that samples from M1 are more dispersed in the plot, thus transcription profiles are more diverse than outside. This might be due to the Pi spike creating a distinct ecological succession in M1.

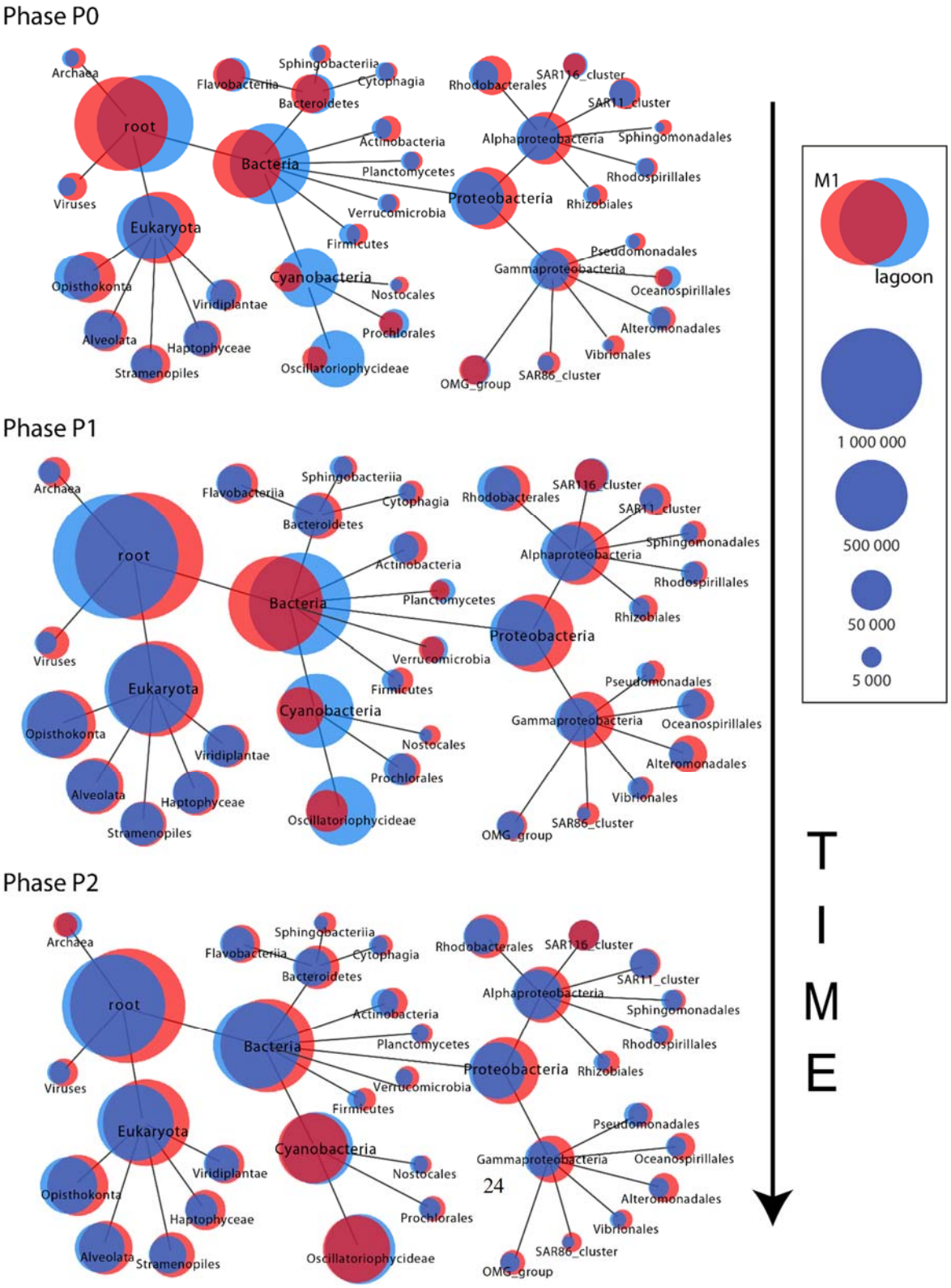
Phase P0
Phase P1
Phase P2
Archaea
root
Viruses
Eukaryota
Opisthokonta
Alveolata
Stramenopiles
Viridiplantae
Haptophyceae
Bacteria
Sphingobacteriia
Flavobacteriia
Cytophagia
Bacteroidetes
Actinobacteria
Planctomycetes
Verrucomicrobia
Firmicutes
Cyanobacteria
Nostocales
Prochlorales
Oscillatoriophycideae
Proteobacteria
Rhodobacterales
SAR116_cluster
SAR11_cluster
Alphaproteobacteria
Sphingomonadales
Rhodospirillales
Rhizobiales
Gammaproteobacteria
Pseudomonadales
Oceanospirillales
Alteromonadales
Vibrionales
SAR86_cluster
OMG_group
M1
lagoon
1 000 000
500 000
50 000
5 000
TIME

Figure 4. Comparison of the taxonomic affiliation of mRNA transcripts from M1 and the lagoon in the three chronological phases P0, P1, and P2, visualized with CoVennTree (Lott et al., 2015). Normalized transcript abundances (TMM-normalized TPM), displayed as the node area, were summed up per phase as follows. P0: day 2 - day 4, P1: day 6 - day 14, P2: day 16 - day 23. The different sizes of the root nodes occur because different transcripts with differing total read abundances may be classifiable in the different datasets, and the data set normalization included all transcripts (also non-classifiable). The overlap of the red (M1) and blue (lagoon) circles denotes the amount of transcripts present in both locations during the respective phase. The diagrams were reduced to show only major nodes and thus raise no claim to completeness. Yet, each node contains the information from all its children nodes, also those not shown. Archaea are scarcely represented in the current RefSeq protein database, thus their transcript abundances are underestimated here.

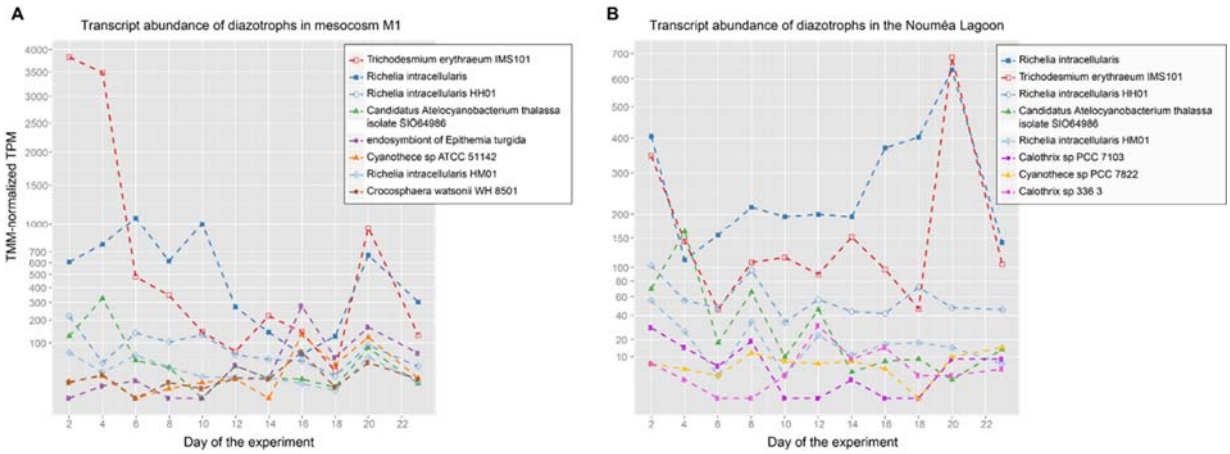

Figure 5. Gene expression in putative diazotrophic cyanobacteria inside mesocosm M1 and in the Nouméa Lagoon. Note the square-root scale for both plots and the generally higher transcript abundances inside M1. Transcriptional activity is presented in TPM (transcripts per million transcripts sequenced), normalized in between samples by TMM normalization (edgeR). Thus, plots can be directly compared, but values are relative.

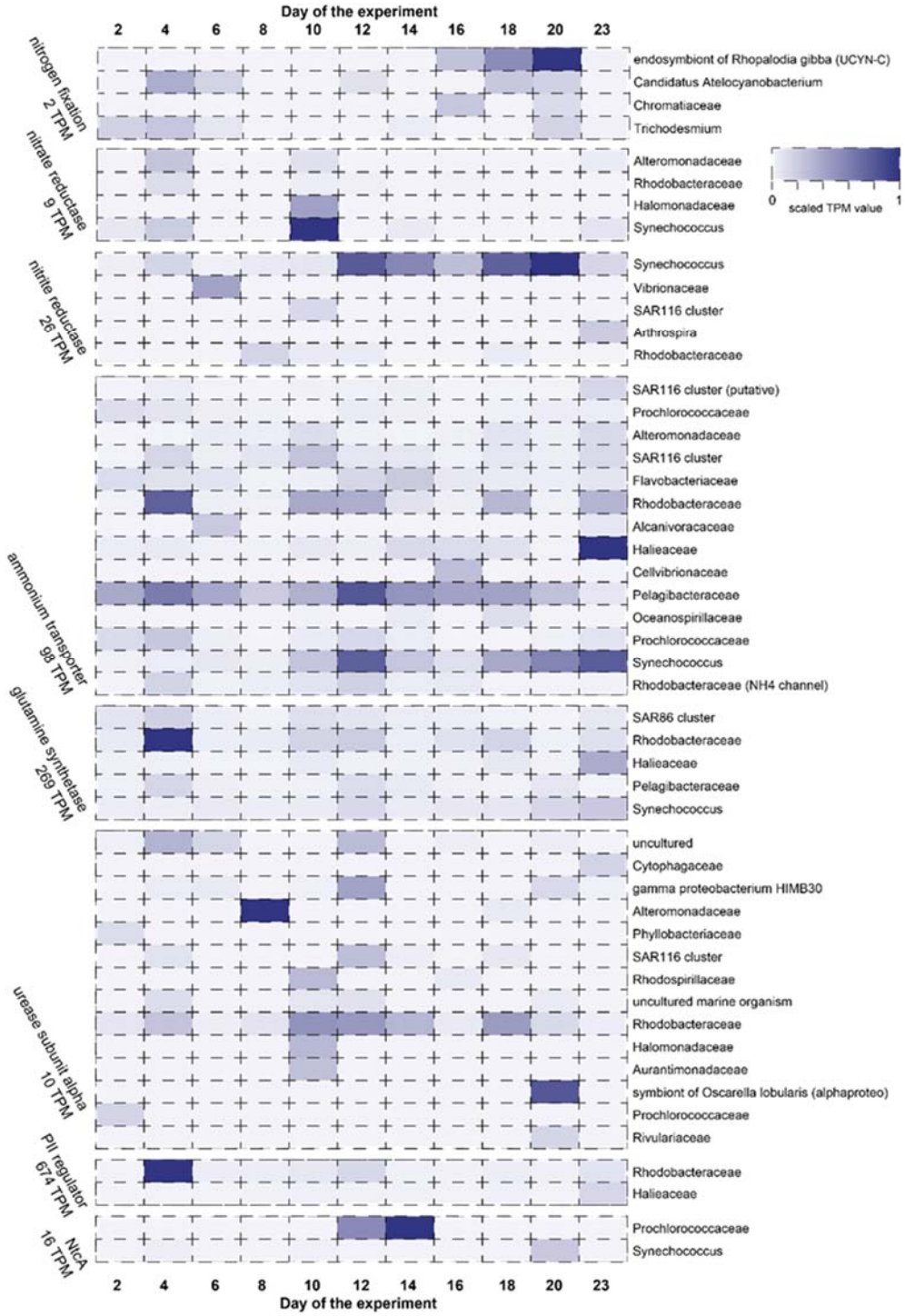

Figure 6. Expression of selected genes indicative for different nitrogen acquisition strategies. TPM counts were summed per taxonomic family, the names of which are denoted to the right of each line. The maximum TPM value for each group is written below the name of that group. For plotting, values were scaled within each functional group, but not for each line, resulting in

the maximum color density always representing the maximum TPM. After the name, in brackets, is additional annotation information, if deviant from the name of the functional group. NifHDK transcript counts were summarized for each taxon to evade possible classification biases due to multicistronic transcripts (i.e. a multicistronic transcript might be classified as either of these depending on the best BLAST match).

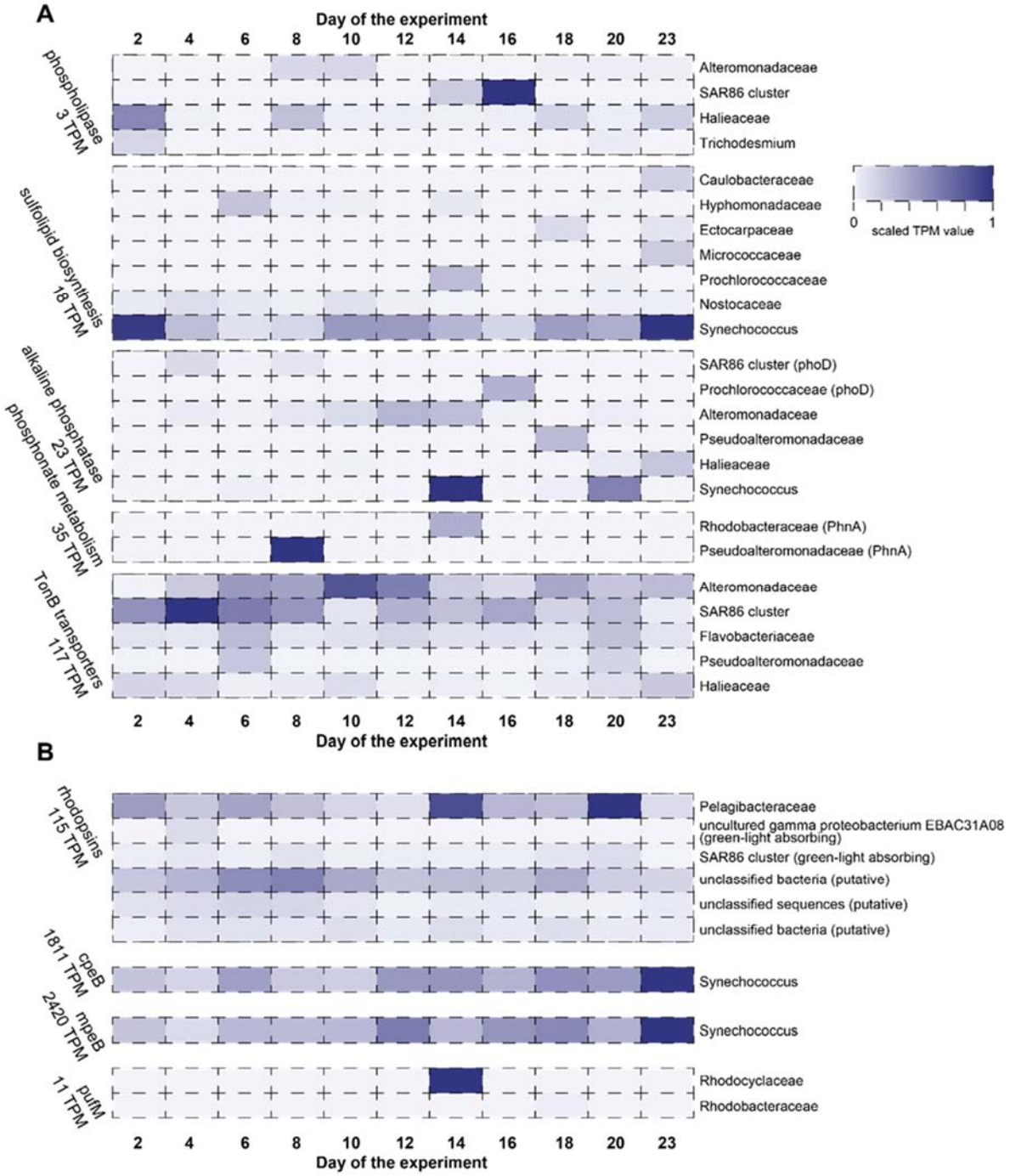

Figure 7. Heat map showing selected genes indicative for different phosphorus acquisition strategies (A) and genes for light-absorbing proteins (B). TPM counts were summed up per taxonomic family, the names of which are denoted to the right of each line. The maximum TPM value for each group is written below the name of that group. For plotting, values were scaled within each functional group, but not for each line, resulting in the maximum color density always representing the maximum TPM. After the name, in brackets, is additional annotation information, if deviant from the name of the functional group.