MaCuMBA

Marine Microorganisms: Cultivation Methods for Improving their Biotechnological Applications

Project number: 311957
Start of the project (duration): August 1st, 2012 (48 months)

Collaborative Project
Seventh Framework Programme
Cooperation, KBBE

Deliverable D6.10

Analysis of the datasets to provide a better understanding of different microbial marine communities

Organisation name of lead contractor: Univ. Miguel Hernandez (UMH – partner 11)

Due date of deliverable: July 2014 (month 24)
Actual submission date: 1 September 2015

Revision: V.1

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)	
Dissemination Level	
PU Public	
PP Restricted to other programme participants (including the Commission Services)	
RE Restricted to a group specified by the consortium (including the Commission Services)	
CO Confidential, only for members of the consortium (including the Commission Services)	**X**

List of reviewers

Issue	Date	Implemented by
v.1	August 2015	UMH (partner 11)

Indicate any document related to this deliverable (report, website, ppt etc) and give file name

** Please attach deliverable documents and any additional material if needed.*

Summary

Objective(s): Analysis of the datasets to provide a better understanding of microbial community structure, function and dynamics in the sea.

Rationale: Cultivation-independent genomic approaches, such as metagenomics, and culture-based enrichments have provided insights into the fundamental principles that drive microbial ecological and processes in marine environments.

Results: Using the huge metagenomic data set collected during the TARA-Oceans 2.5-year circumnavigation cruise (Karsenti et al., 2011; Armbrust & Palumbi, 2015) partner 16 have studied the global distribution of the marine picocyanobacteria *Prochlorococcus* and *Synechococcus* at high taxonomical resolution. This study has allowed to refine the distribution patterns of ecologically important genotypes that were largely underestimated until now but also identify new ecotype assemblages dominating huge geographic areas. By combining both metagenomic sequencing with culture-based enrichments partner 11 has investigated the selective impact of different particles and media composition on community structure. Results confirmed that the free-living Mediterranean and particle-attached microbial communities are distinct, the differences at the prokaryotic level are difficult to analyse due to the large number of eukaryotic sequences. Moreover, partner 11 presents another method to characterize microorganisms attached to particles and shows that different particle compounds drives the enrichment of distinct and specific particle-associated microbes. Some interesting novel genomes have been reconstructed including a Bacteriovorax predator bacterium and some novel Gammaproteobacteria and Bacteroidetes.

Partner(s) involved in Deliverable production: Partner 11, UMH; Partner 16, CNRS; Partner 18, UF; Partner 22 RIBO.

Partner(s) involved in Deliverable production:

Partner Name	Partner Number	Participation
UvA	2	NP
UBO	3	NP
MATIS	10	NP
UMH	11	Y (B)
UCC	12	NP
UW	13	NP
UMIL	15	NP
CNRS	16	Y (A)
UF	18	Y (D)
DSMZ	20	NP
RIBO	22	Y (C)

Y, participation in the deliverable

NP, no participation in the deliverable

D6.10 – A. Contribution of partner CNRS (Partner 16)

With the aim to analyze the diversity and biogeography of marine picocyanobacteria, the CNRS Roscoff-MaPP team (Partner 16) analyzed 109 metagenomes from the bacterial size fraction collected across the tropical to temperate world ocean at two depths: surface and deep chlorophyll maximum (sometimes the latter sample was in fact collected in subsurface). The selected approach consisted in recruiting metagenomic reads ($_{mi}$Tags; Logares et al., 2013) targeting the high-resolution marker gene *petB* (Mazard et al., 2012). We first checked that metagenomic *petB* fragments of 100 bp over most of the gene length could be assigned at sub-clade level with a high confidence level. From 119 to 14,139 picocyanobacterial *petB reads* (average: 3,371; median: 2,680) were recruited per station along the whole TARA-Oceans transect (66 stations, 109 metagenomes, 20.2 ± 9.9 Gb of metagenomic data per sample) using a non-redundant reference database of 596 high quality *petB* sequences, representing most of the genetic diversity identified so far in cultured isolates and environmental samples for the *Prochlorococcus* and *Synechococcus* genera.

To search for potential hidden genetic diversity within the marine picocyanobacterial communities sampled during the TARA-Ocean transect, we examined the percent identity of recruited reads with regard to the *petB* reference database. The diversity of the most abundant *Synechococcus* clades (I-IV) was generally well covered by reference sequences since most reads displayed more than 94% identity to their best-hit in the database, a value corresponding to the cut-off previously used to define *petB* sub-clades (Mazard et al., 2012). In contrast, for some other taxa, some of the recruited reads were quite distantly related to reference sequences (i.e., within 80-94% identity), indicating that *in situ* diversity of these groups was not fully covered by our reference database. To have a more realistic view of their *in situ* diversity, we therefore assembled about 150 near complete *petB* sequences from environmental reads by targeting samples particularly enriched in these 'missing' genotypes. These novel sequences were then added to our reference database and allowed us to assess more reliably the diversity and distribution of these groups in the field.

As expected from previous literature, *Prochlorococcus* was the dominant picocyanobacterium at the global scale, representing 91% of all *petB* reads from the bacterial size fraction, compared to 9% for *Synechococcus*. Only three main types of profiles were identified for *Prochlorococcus* in surface samples (Fig. 3), i) dominance of HLI populations in temperate waters (above 35°N and 32°S), ii) dominance of HLII in warm, tropical and iron-replete waters (30°S to 30°N), with mixed HLI-HLII profiles at intermediate latitudes and iii) co-occurrence of HLIII and IV at a ratio of ca. 1:3 in High Nutrient Low Chlorophyll (HNLC) areas. Interestingly, the latter clades contributed altogether to 16% of the *Prochlorococcus* community in TARA-Oceans samples and their distribution actually cover most of the equatorial Pacific zone from 13 °S to 14°N.

Synechococcus assemblages were much more diversified than *Prochlorococcus'* with 6 major types of profiles throughout the TARA transect (Fig. 1). Although the prevalence of clades I-IV, contributing altogether to 75% of *Synechococcus*-assigned *petB* reads in the TARA-Oceans samples, is consistent with previous studies (Zwirglmaier et al., 2008; Mazard et al., 2012) several striking elements arose from analyses of the TARA-Oceans metagenome. First, although clade III was thought to be the dominant clade in tropical and subtropical oceanic gyres, it is in fact a major contributor to the *Synechococcus* community only in the Mediterranean Sea and in the Gulf of Mexico, while it was

present only at low abundance at a number of stations of the Indian, Atlantic and Pacific Ocean. Also unexpected was the notable global abundance of sub-cluster 5.3 (5.4%) that previously had only been sporadically detected in other studies. Here, we found that this sub-cluster seems to be specifically present at a few stations of the Atlantic and Indian Ocean and at all stations of the Red Sea and Mediterranean Sea, where they contributed up to ca. 30 % of the *Synechococcus* community at the Gibraltar strait. Probably the most striking result of this study was the high contribution of clades CRD1 (9.3%), and EnvB (4.9%). these clusters dominate the whole Pacific Ocean from 33°S to 35 °N and were also locally abundant in South and North Atlantic as well as in Indian Ocean.

Altogether, this study allowed us i) to unveil novel *in situ* diversity along the TARA-Oceans transect with regard to previously known diversity, ii) to considerably refine the distribution patterns of ecologically important genotypes that were largely underestimated until now but also iii) to identify new ecotype assemblages dominating huge geographic areas and displaying distinct and complementary niches with regard to previously known *Synechococcus* ecotypes. These analyses that will be described in a paper (Farrant Doré et al., in prep.) will also very useful to decipher the links between the ecology, evolution and biology of these important members of phytoplankton.

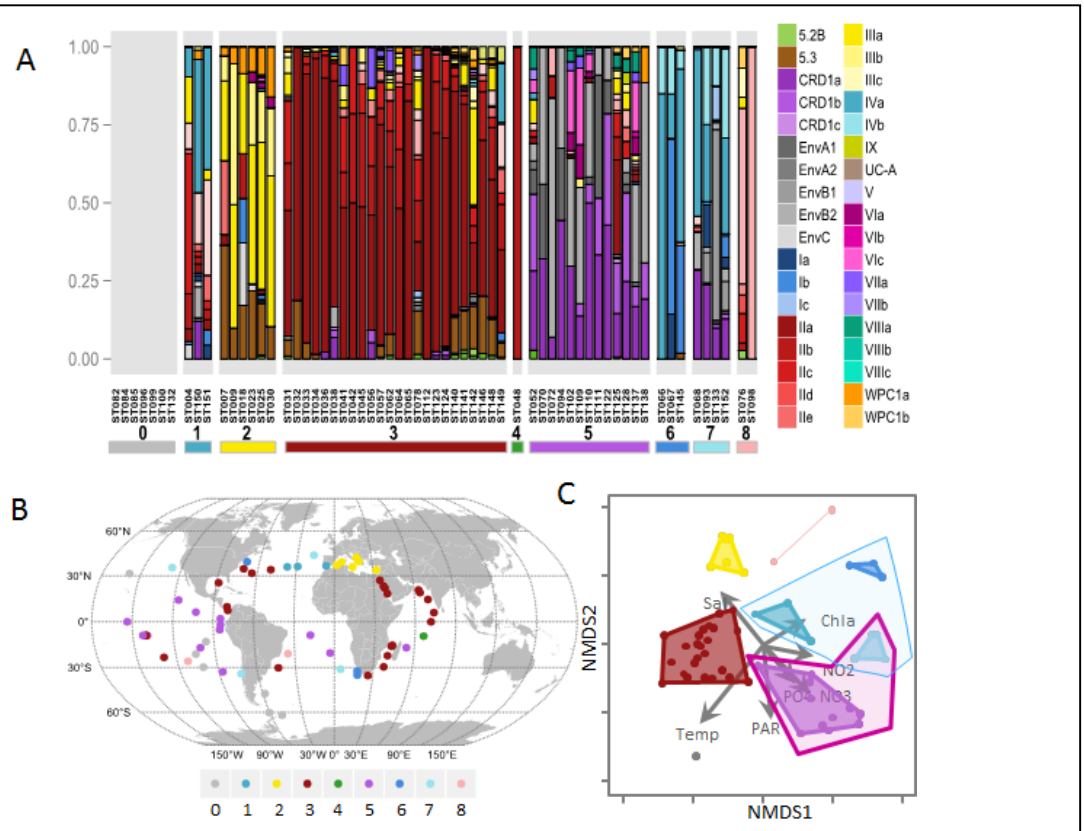

Fig. 1. Taxonomic composition and global distribution of the major *Synechococcus* assemblages detected along the TARA-Oceans transect. A. The 6 major assemblages defined using hierarchical clustering. B. Global distribution of hierarchical clusters along the TARA-Oceans transect. C. NMDS analysis showing the correlations between clusters of stations and physicochemical parameters (Farrant, Doré et al., in prep.).

D6.10 – B. Contribution of partner UMH (Partner 11)

UMH has combined direct metagenomic sequencing with culture-based enrichments to explore the role of particles in shaping microbial diversity in the Mediterranean Sea. To initially characterize the size-fractionated microbial communities of the Mediterranean, we sequenced and analysed four large metagenomic datasets of small (0.22-5.0 µm) and large (5.0-20.0 µm) filter fractions from multiple time points (September 2013 and December 2014), theoretically representing the free-living (FL) and particle-associated (PA) communities, respectively. 7-12 Gb of sequencing data were retrieved from each data set. The PA metagenomes revealed much higher percentages of eukaryotic and viral sequences and lower percentages of bacteria and archaea based on both rRNA analyses. The ratio of eukaryotic 18S rRNA hits compared to that of the prokaryotic 16S rRNA ones (0.5 and 0.63) compared to that of the FL (0.014 and 0.036). Despite the high abundance of eukaryotic sequences in the PA, differences were also noted between the FL and PA communities within the prokaryotic community metagenomes. Classification of prokaryotic rRNA reads, for example, revealed a lower percentage of *Actinobacteria* and higher percentages of *Cyanobacteria, Verrucomicrobia*, and *Planctomycetes* in the PA communities.

UMH designed also enrichment experiments in order to investigate the selective impact of different particles on community composition using metagenomic sequencing of particle-specific enrichments. Two types of particulate material were used. Sand (SWsnd) and diatomecious earth (SWde) are both inorganic forms of silicates, sand particles were larger and with flat surfaces while SWde are smaller and with perforated surfaces (larger surface/volume ratios). The other two, cellulose (SWcel) and chitin (SWchi), were organic and could be used as additional carbon source to the pyruvate supplied, both polymers are abundant in marine POM. The experimental design is summarized in Figure 2.

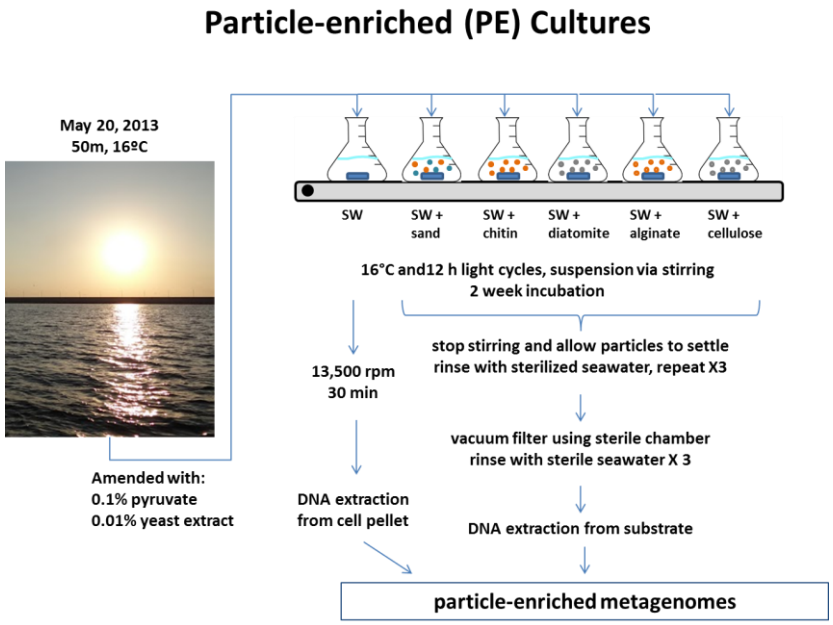

Figure 2. Particles enrichment experiment

The assembled data for each metagenome provided between 6,000-11,000 contigs > 1000 bp, with the largest contig assembled being > 350,000 bp. We used the phylogenetic classification of both annotated contigs > 10 (Figure 3) and rRNA sequences (>~100 bp) from the raw metagenomic reads to identify the dominant microorganisms in each metagenome. Even at the class level, we observed distinct community structures with both approaches in both cases, most notably the inorganic particles bore the highest similarity to the control with large a mounts of unclassified and *Roseobacter*. In the SWchi and SWde enrichments, which were dominated by Gammaproteobacteria and unclassified contigs, respectively. Using the larger contigs for classification also allowed us to identify a number of organisms, presumably the most abundant since they were assembled in large contigs, at the genus/species level, providing further evidence of variability between the particle enrichments. ANI comparisons between the assembled contigs from each metagenome and their corresponding most closely related microorganism revealed a number of species within the *Rhizobiales, Rhodobacteraceae, Flavobacteraceae, Bdellovibrionaceae*, and *Vibrionaceae* families that were not previously described.

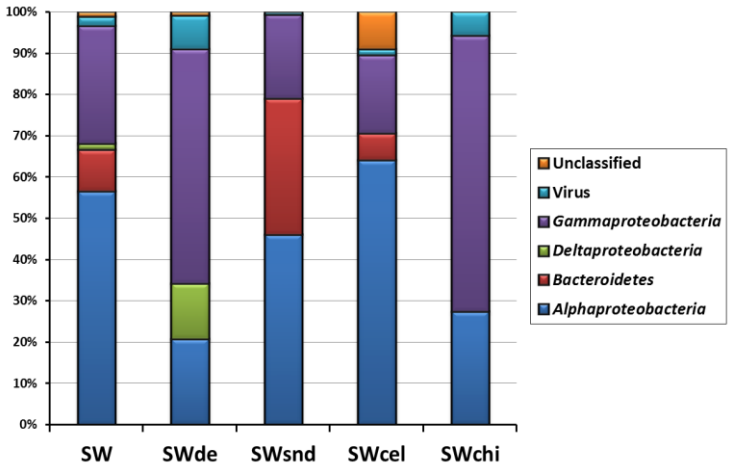

Figure 3. Phylogenetic classification of annotated contigs larger than 10 Kb

To identify whether or not the enriched microorganisms described above were an abundant part of natural microbiomes, we performed recruitment analysis by comparing each contig group to that of numerous metagenomes from a variety of seawater samples. The resulting RPKG values which provide an estimated coverage, revealed very low, if any, recruitment of the enriched microorganisms from that of natural samples. Neither the place of origin (Mediterranean Sea, Baltic Sea, Atlantic Ocean) nor the filter fraction used (ranging between 20 - 0.1 µm) for each metagenome compared had a significant effect on the low RPKG levels for each enriched microorganism and their most closely related genomes. Not even the reads from the FL or the PA metagenomes mentioned above were recruited. These results suggest that the enriched microorganisms are part of the rare biosphere, which exists at nominal levels under natural conditions.

In a second enrichment approach, UMH applied community sequencing of a marine enrichment culture using four different media composition. In the laboratory, 480 mL of seawater was added to

each of four flasks containing 20 mL of one of the following: 1. Mannitol / Peptone / Gentamicin media (50 µg/mL) (MAN), 2. Aspartic Acid/ Arginine/ Gentamicin media (50 µg/mL) (AAA), 3. Pyruvate / glycine / methionine / iron chloride (PAM) 4. Ammonium chloride / Magnesium sulfate / Sodium bicarbonate / Selenite-Tungtate / Vitamin (Ammonium oxidizers media) (AOM). Low levels of pyruvate (0.1% final concentration) and yeast extract (0.01% final concentration) were also added to each sample, which were then incubated for two weeks at 15°C with 12 h light/dark cycles with constant stirring. After two weeks of growth under ecologically relevant conditions, microbial DNA was extracted for metagenomic sequencing. Enough quality DNA was obtained from all of the samples and thus four metagenomes were sequenced. Illumina (Hi-Seq) sequencing resulted in 18-23 million reads per enrichment, totalling 1.6-2.1 Gb for each data set. The assembled data for each metagenome provided between 700-6,000 contigs > 1000 bp, with the largest contig assembled being > 350,000 bp. The picture obtained from the rRNA analysis shows that the community in the MAN and AAA metagenomes was massively dominated by the bacteriodetes *Polaribacter* sp. while *Enterovibrio* and *Pseudoalteromonas haloplanktis* were the most abundant in AOM and PAM respectively. Enrichment in these organisms has led to reconstruct the complete genome of several of them.

D6.10 – C. Contribution of partner Ribo (Partner 22)

Based on their contribution to the Deliverables D6.4 and D6.5, partner RIBO has provided bioinformatics infrastructure and expertise in genome and metatranscriptome data analysis related to Deliverables D6.8, D6.9, and D6.10 by supporting other MaCuMBA partners (within and also outside of WP6). In particular, close collaboration has been carried out with partners UvA and NIOZ, also including visits at partner Ribocon for training and joined data mining in the context of both projects.

D6.10 – D. Contribution of partner UF (Partner 18)

Partner 18 (UF) has analyzed complex metatranscriptome datasets (see **Table 2** in **Deliverable D6.9**) from samples of the marine microbial community taken during the VAHINE project in the SW Pacific (collaboration with Sophie Bonnet (IRD/MIO Noumea/Marseille)). The samples were obtained from surface waters of the open sea and from 3 replicate mesocosms. These datasets were analyzed aiming at a better understanding of microbial marine communities.
Several examples of the obtained insight are shown in **Figure 4** and **Figure 5**. The cell cycle gene *ftsZ* peaks in its expression before dark, mainly due to the likely synchronized gene expression in marine picocyanobacteria and SAR11-type alphaproteobacteria. This points at the synchronized cell cycle among these bacteria, matching earlier observations for *Prochlorococcus* populations from the Red Sea (Holtzendorff et al., 2002). In contrast, the *nrdJ* peak is solely due to the transcriptional activity of *Prochlorococcus* cyanobacteria. Their *nrdJ* gene was the only class II ribonucleotide reductase in the dataset, pointing at high DNA synthesis rate in these microorganisms towards the later afternoon. This matches well to earlier reports on synchronized DNA synthesis at the end of the light phase in *Prochlorococcus* cultures kept under a natural light/dark regime in the laboratory (Holtzendorff et al., 2001). The sequence analysis of transcripts associated to dinitrogen fixation indicated that at least two distinct populations of such microbes had been present. From these, one was mainly active during the day (identified as UCYN-A type cyanobacteria, transcription in the

morning), whereas the other one was active at night (identified as *Crocosphaera*-type cyanobacteria), with transcription occurring in the evening (**Figure 4**).

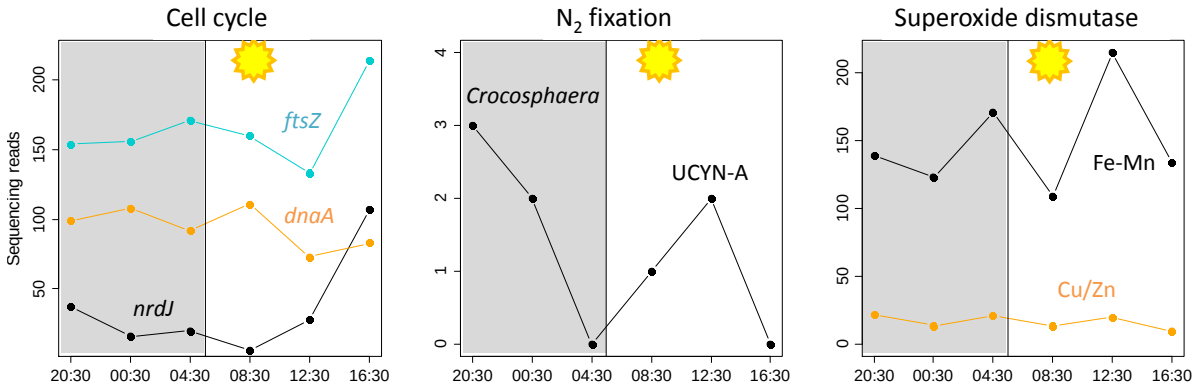

Figure 4. Time course of gene expression in the surface marine microbial community incubated in cubitainers floating at the surface, sampled every 4 hours. The examples shown demonstrate different temporal regimes for genes associated to cell cycle, nitrogen fixation or oxidative stress (superoxide dismutases). The sequencing reads from the different samples were subsampled to match the smallest dataset before BLASTX-based annotation.

In contrast, the transcript accumulation of genes associated to oxidative stress peaked in the morning, in parallel with that of photosynthesis-associated genes (**Figure 4**), pointing at an enhanced radical stress during the light phase. This observation matches reports that the cyanobacterium *Prochlorococcus* depends on the activity of hydrogen peroxide scavenging microbes for growth at the ocean's surface (Morris et al., 2011). Interestingly, this radical-scavenging activity was dominated by Fe-Mn superoxide dismutases whereas the Cu/Zn-dependent enzymes showed little expression and no variation.

To extend the insight into nitrogen assimilation to the preferred sources of fresh nitrogen, ammonia and urea, we followed the expression of genes related to ammonia uptake or urea utilization over the full course of 23 days and compared the transcript accumulation for all taxa transcribing the respective genes with > 1% of the highest observed read count (**Figure 5**).

ammonium transporters urease subunit alpha

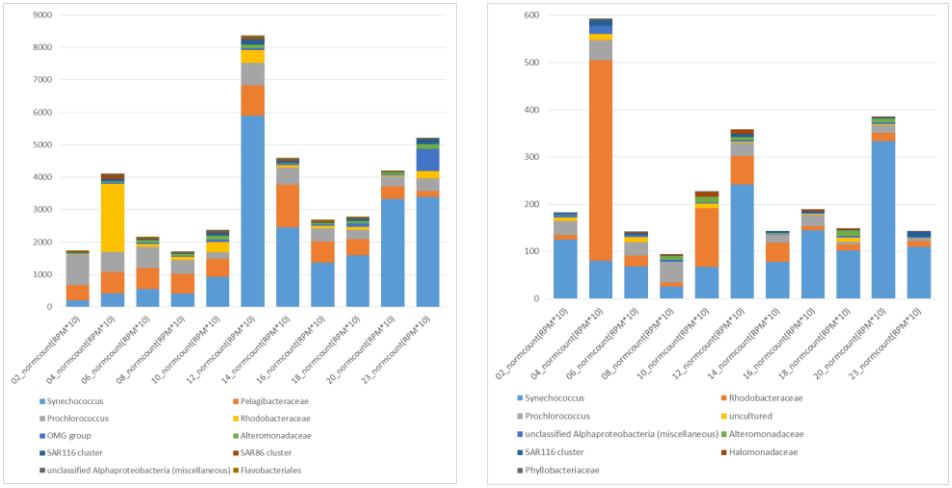

Figure 5. Transcript accumulation for ammonium transporters (left) and urease subunit alpha (right) with a taxonomic resolution for all taxa transcribing the respective genes with > 1% of the highest observed read count. Metatranscriptomic reads were obtained from samples in mesocosm 1 (compare **Table 2** in **Deliverable D6.9**), recruited to a manually curated database of ammonium transporter and urease genes. Recruited reads were then additionally filtered by BLASTn to avoid false positives. Read counts were normalized to total non-ribosomal reads per sample [(reads per million)*10].

We found genes for the transport of ammonia were highly expressed in most of the abundant bacterial groups, like *Synechococcus, Prochlorococcus*, and SAR11. Compared to their relative abundances, SAR116, SAR86, and Flavobacteriales expressed much less ammonium transporters, and members of the Rhodobacteraceae many more.
Urea is another important nitrogen source in oligotrophic marine waters. Our data show that the uptake of urea is most important for members of the Rhodobacteraceae in the first, but not the second half of the experiment. Other main groups expressing urease were *Synechococcus* and *Prochlorococcus* (**Figure 5**), whereby their expression mainly followed their abundance.

References (*papers acknowledging MaCuMBA)

*Farrant GK, Doré H, Cornejo-Castillo FM, Partensky F, Ratin M, Scanlan DJ, Acinas SG, Garczarek L and the TARA-Oceans consortium. Novel insight into the global distribution pattern of marine picocyanobacteria ecotypes and their realized niche at high taxonomical resolution based on TARA-Oceans dataset. In preparation for the Proceedings of the National Academy of Science of the USA.

Holtzendorff J., Partensky F., Jacquet S., Bruyant F., Marie D., Garczarek L., Mary I., Vaulot D., Hess W.R. (2001): Diel expression of cell cycle-related genes in synchronized cultures of *Prochlorococcus* PCC 9511. J Bacteriol 183: 915-922.

Holtzendorff J., Marie D., Post A.F., Partensky F., Rivlin A., Hess W.R. (2002): Synchronized expression of *ftsZ* in natural *Prochlorococcus* populations of the Red Sea. Environ Microbiol 4: 644-653.

Logares R and 16 coauthors (2013) Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. Environ Microbiol 16: 2659-2671.

Mazard S, Ostrowski M, Partensky F and Scanlan DJ (2012) Multi-locus sequence analysis, taxonomic resolution and biogeography of marine Synechococcus. Environ Microbiol 14: 372–386.

Morris JJ, Johnson ZI, Szul MJ, Keller M, Zinser ER. (2011) Dependence of the cyanobacterium *Prochlorococcus* on hydrogen peroxide scavenging microbes for growth at the ocean's surface. PLoS One 6(2): e16805.

Zwirglmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaulot D, Not F, Massana R, Ulloa O, and Scanlan DJ. (2008) Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. Environ Microbiol 10: 147-161.