Marine Microbial Biodiversity, Bioinformatics & Biotechnology

Grant agreement n°287589

Acronym: Micro B3
Start date of project: 01/01/2012, funded for 48 month

Deliverable 6.8

Description of Community Structure and Composition in its Environmental Context

Version: 1.0
Circulated to: WP6 (Dec 2013)
Approved by: Prof. Frank Oliver Glöckner (January 14)

Expected Submission Date: 31.12.2013
Actual submission Date: 15.01.2014

Lead Party for Deliverable: CNRS

Mail: cbowler@biologie.ens.fr Tel.: +331 4432 3525

Dissemination level:	
Public (PU)	X
Restricted to other programme participants (including the Commission Services) (PP)	
Restricted to a group specified by the consortium (including the Commission Services) (RE)	
Confidential, only for members of the consortium (including the Commission Services) (CO)	

The Micro B3 project is funded from the European Union's Seventh Framework Programme (Joint Call OCEAN.2011-2: Marine microbial diversity – new insights into marine ecosystems functioning and its biotechnological potential) under the grant agreement no 287589. The Micro B3 project is solely responsible for this publication. It does not represent the opinion of the EU. The EU is not responsible for any use that might be made of data appearing herein.

Table of Content

Summary

Marine plankton ecosystems comprise a range of highly diverse microbes that are crucial for the regulation of Earth's biogeochemical cycles and climate (Arrigo, 2005; Falkowski *et al.*, 2004). Yet their organization, evolution and dynamics remain poorly understood. In this Deliverable, methods have been developed to systematically describe the relationships between viral, prokaryotic and protist diversity patterns taken from data sets derived from sediments and a range of different pelagic environments. Metagenomes and metatranscriptomes have allowed to investigate the phylogenetic variation among samples using marker gene-based approaches. Such methodologies have been complemented with phylogenetic information arising from comparing community composition based on key nuclear, chloroplastic, and mitochondrial genes at both the DNA and RNA levels from various organismal groups that cover the entire community. Where available, flow cytometry data have also been used to interpret the values of DNA and RNA markers from protists. Subsequently, the intrinsic complexity, structure, and dynamical properties of each community have been assessed using network statistics and stochastic modelling based on the presence, amount, and nature of biodiversity markers.

Altogether, this first-order analysis of marine microbial community structure and dynamics serves as a reference framework to interpret the metagenomics and metatranscriptomics data analysed in the Micro B3 project, and will orient the choice of high-throughput technologies to further explore highly-complex marine eukaryotic metagenomes and metatranscriptomes.

1. Community diversity analysis

Viruses, bacteria, archaea, protists and planktonic metazoans form the bulk of biomass throughout the oceans and drive the global biogeochemical cycles that fuel planet Earth (Arrigo, 2005; Falkowski *et al*., 2008; Karl, 2007). For instance, in carbon cycling marine microbes produce nearly as much oxygen through photosynthesis as land plants (Field *et al*., 1998). However, this system includes not only photosynthetic bacteria, but also heterotrophic and mixotrophic microbes, as well as their associated viruses. Photosynthetic protists and their grazers are key elements of this system, and ocean viruses modulate photosynthesis through mortality of photosynthetic cells as well as by encoding core photosynthesis genes that are expressed during infection (Lindell *et al*., 2005; Sharon *et al*., 2007; Sullivan *et al*., 2006). Thus, both groups directly impact global photosynthesis and carbon recycling. Therefore, only an ecosystem-wide approach, from viruses to fish larvae, will enable us to start disentangling the functioning of the Earth system. This ranges from mapping organismal diversity across scales spanning multiple orders of magnitude in size and abundance to generating empirical datasets that inform conceptual models about the complex interplay between organisms driving fluxes of energy, biogeochemical and

molecular 'currencies' in ocean ecosystems (e.g., see Nature Insight in Microbial Oceanography (Nature (2009) 459: 179-212)).

Global-scale studies of morphological, genetic and functional biodiversity of plankton organisms in relation to the changing physico-chemical parameters of the oceans (Bowler *et al*., 2009; Fuhrman, 2009; Irigoien *et al*., 2004; Rohwer and Thurber, 2009) are required to understand and manage sustainably our fragile oceans. Specifically, such datasets will improve our understanding of evolution, principles governing marine ecology, prediction of fish stock distribution, ecosystem services monitoring, and prediction of global climate variations (see Science Special Section on Changing Oceans (Science (2010) 328: 1497-1528)). Planktonic organisms are also an enormous but largely untapped source (Bowler *et al*., 2009; Sarmento *et al*., 2010) of bio-active compounds for the pharmaceutical, food, and cosmetics industries, as well as metabolic pathways which may provision our future energy needs (Arnaud-Haond *et al*., 2011).

A wide range of resources have been used by the Micro B3 partners contributing to Deliverable 6.8 to perform in-depth analyses of plankton community composition, as summarized below.

2. Generating a pipeline for community diversity analysis

2.1 Resources utilized

A range of datasets that can be used to address plankton community structure are now available, and include data from the Global Ocean Sampling (GOS) expedition (Rusch *et al*., 2007), the Earth Microbiome (Gilbert *et al*., 2010) project, the *Tara* Oceans expedition (Karsenti *et al.*, 2011), the Malaspina expedition, as well as the Census of Marine Life (notably ICOMM; International Census of Marine Microbes), long-term observation sites, and arrays of remote sensors on floats that provide physical, chemical and biological data.

2.2 Methodologies

2.2.1 Sampling methodologies

An ideal dataset for analysing microbial community diversity should be derived from a representative sampling of the entire community. This is a major challenge given that these organisms cover several orders of magnitude in size and are present in vastly differing amounts in seawater, and the only attempt so far to systematically sample complete microbial ecosystems (from the smallest viruses to the largest protists) during one entire expedition has been performed by the *Tara* Oceans consortium, in particular the CNRS and CSIC partners of Micro B3 (Partners 6 and 7, respectively). An overview of the sampling methodologies has been reported previously (Karsenti *et al.*, 2011) and full details of the protocols have recently been submitted for publication by CNRS-Roscoff (Not *et al.*). Briefly, the protocols are based on the collection of different volumes of seawater using Niskin

bottles, peristaltic pumping, or net tows, followed by the size fractionation of organisms to preferentially recover viruses, giruses, prokaryotes, pico-, nano- and micro-plankton.

The *Tara* Oceans datasets are therefore particularly precious for analysis of the entire plankton community, which is not always the case for other datasets, which typically prioritize the collection of one particular component of the ecosystem, e.g., prokaryotes. For each analysis it is therefore important to recognize the limitations of what insights can be obtained as a consequence of the sampling protocols performed.

The analyses that have been performed have been instrumental in establishing an appropriate sampling methodology that can be implemented for Ocean Sampling Day in Work Package 2 to recover entire microbial communities.

2.2.2 Analysis methodologies

The very broad skills available within the different WP6 partner laboratories has permitted the exploration of a wide array of statistical methodologies to assess community structure, which includes analyses of diversity, richness, stability, endemism, rank abundance, and co-occurrence, amongst others.

2.3 Results

In the following sections a brief summary has been provided about the work that has been performed with specific organismal groups, viruses, prokaryotes, and protists. An example of cross-kingdom data analysis has subsequently been presented based on the Agulhas retroflection region below South Africa.

2.3.1 Viruses

A study coordinated by Genoscope (Partner 17) was undertaken to describe end-to-end genomics of Mediterranean plankton communities collected during the *Tara* Oceans expedition. The analysis combines metagenomes, metatranscriptomes and metabarcodes to compare three Mediterranean stations in terms of environmental biogeochemical parameters, global genetics (eukaryotes, prokaryotes and viruses) and functional content. The CNRS-IGS laboratory (Partner 6) contributed to this study by analyzing viral distributions in the 29 metagenome samples of this dataset, focusing on comparisons of the 5 different size fractions studied.

As observed previously (Angly *et al.*, 2006), phages were highly abundant in samples from smaller size fractions (e.g. <5µm), representing more than 3% of the overall mapped metagenomic reads. However, we were also able to observe viral sequences in samples from larger fractions (representing around 0.2% of the mapped reads), even though the larger pore sizes of these fractions (e.g. >5µm) are not expected to retain such viruses.

Taxonomic annotation of these sequences revealed a higher heterogeneity in viral taxonomy in the large fractions than in the smaller fractions. A similar heterogeneity was

also observed for the distribution of host types in the larger size fractions. For instance, in larger fraction sizes, samples from station 30 displayed a high proportion of eukaryotic viruses, but very few bacteriophages. Interestingly, in the samples where we observed low relative abundance of bacteriophages, we also noticed low relative abundance of bacteria.

Co-occurence of viruses with their hosts could be illustrated by the case of a Bacillariodnaviridae, a Chaetoceros diatom infecting virus (Nagasaki, 2008). Sequences from Bacillariodnaviridae were uniquely found in the sample coming from station 23, DCM depth, fraction 20-180µm (2.4kb mapped), while their diatom host sequences were detected in exactly the same location, in the two fractions consistent with its cell size (5-20 and 20-180µm). This suggested we might have sampled an infected host cell in the larger size fractions. We hypothesize that most viral sequences captured in the larger size fractions (> 5µm) might have originated from viruses replicating in their host cells. Such occasional co-detection might explain the observed viral type heterogeneity in large fractions, whilst the more homogeneous viral type distribution could be explained by the detection of free viral particles which could be less coupled to their host activity.

2.3.2 Prokaryotes

In direct interaction with Work Package 5, the MBA team (Partner 9) have looked at bacterial OTU assignments using various pipelines. A range of different bioinformatics protocols are available and are used by researchers to cluster, classify, and assign 16S rRNA short read amplicon sequences with taxonomic information. Increasing numbers of environmental samples around the globe are being sequenced using amplicon sequencing in attempt to characterise and compare microbial community diversity. The MBA team have specifically considered the impact of different bioinformatics tools and databases (specifically the latest release of the Greengenes and Silva databases) using an Illumina HiSeq 16S rRNA amplicon dataset from an Indian Ocean marine sample totalling more than 7 million reads. The triplicate independent PCR experimental design of this dataset together with the high depth of sequencing allows the implications of bioinformatics protocols on rarefaction sampling requirements and taxonomic assignments to be assessed at different levels.

On the other hand, CSIC (Partner 7) has validated the use of miTags (metagenomic reads of the 16S rDNA fragments from Illumina metagenomes) as an alternative approach to explore microbial diversity and community structure at a global scale using Tara Oceans datasets. A first paper validating this approach in comparison with amplicon Tags (454 pyrosequencing; 454Tags) and metagenomes performed by 454 pyrosequencing (m454Tags) has recently been published (Logares *et al.*, 2013).

From the Malaspina circumnavigation, CSIC has used another approach to analyze microbial diversity and biogeography of prokaryotic communities at global scale in the dark oceans, specifically in bathypelagic ocean waters (Salazar *et al.*, unpublished). In this study, they have used and compared amplicon tags from Illumina (16S iTags) with ARISA profiles based on ITS length variation to compare the structure of deep ocean prokaryotic communities. A

good correlation was found between the prokaryotic community structure based on Bray-Curtis distance matrix using the length of the ITS from ARISA profiles compared with the iTags, and the results uncovered not only a significant fraction of novelty within the prokaryotes but also described the biogeographical patterns and barriers as well as the processes involved in shaping prokaryotic communities in this underexplored realm (Figure 1).

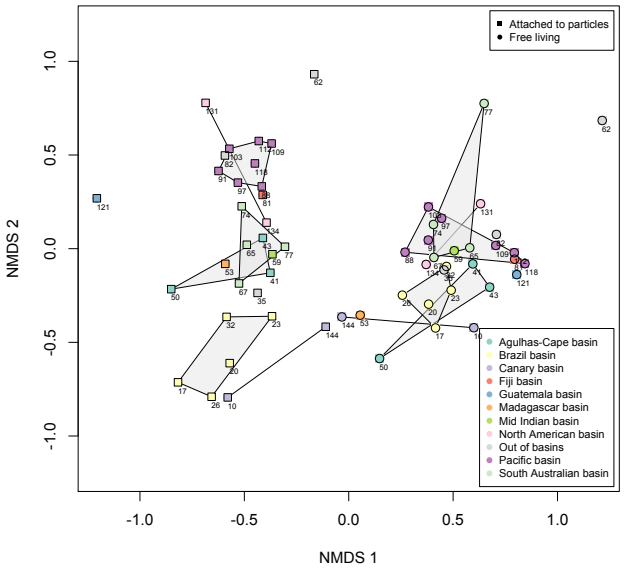


Figure 1. *NMDS plot showing how prokaryotic communities from particle attached fractions are clustered based on specific ocean basins.*

At the Blanes Bay Observatory, the CSIC team has assessed DGGE fingerprinting and CARD-FISH quantifications during two years (Diez-Vives *et al.*, 2013) to assess changes in prokaryotic diversity. The results show that at a temporal scale, the relative abundance of Bacteroidetes exhibits a marked seasonality, being higher in spring and decreasing in winter (Figure 2).

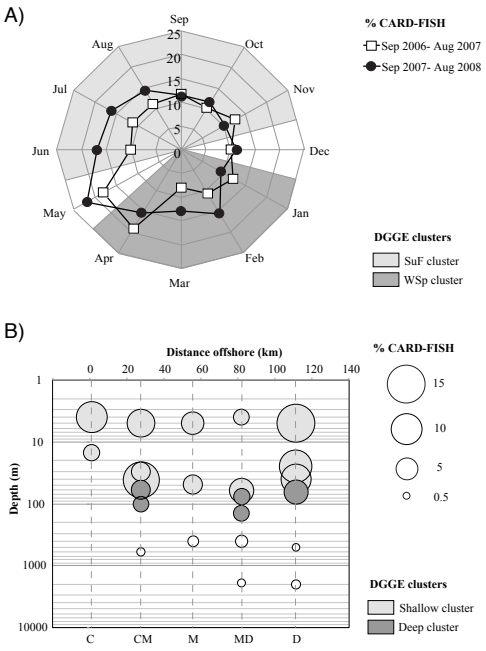


Figure 2*. Percentage of Bacteroidetes in CARD-FISH counts: A) over two consecutive years in the temporal study, represented as circles and squares and B) in the spatial study as circle areas for each sample point. The groups defined by the DGGE clustering analyses (SuF/WSp or Shallow/Deep) are also represented as different grey tones. Since transitional months (May and Dec) fell into different clusters in the 2 years analyzed, these are coloured in white. Samples under 200 m (not included in the DGGE analysis) are also in white (from Díez-Vives et al., 2013).*

Finally, because most microbal diversity studies encountered in WP6 are spatially explicit (i.e., geolocalized), the AWI team (Partner 5) has examined the typical issues found at the technical (sequencing depth, presence of rare taxa) and sampling (number of samples) levels when describing diversity patterns of bacteria in marine environments. The taxa-area relationship (TAR) and the distance-decay relationship (DDR) both describe spatial turnover of taxa, and are central patterns of biodiversity. Using the ICoMM database, consisting of one of the most comprehensive sets of 16S rRNA gene sequences available to date, and gathering hundreds of samples collected across the globe, they have compared TAR and DDR of bacterial communities across different marine realms and ecosystems at the global scale. They found that slope coefficients of bacterial TARs and DDR are mostly affected by removing rare taxa and by the number of sampling sites considered in the calculations. Furthermore, TAR and DDR slope coefficients were found to be overestimated at sequencing depth < 4,000 sequences per sample. They also found marine bacterial TAR and DDR to be steeper in ecosystems associated with high environmental heterogeneity or spatial isolation, namely marine sediments and coastal environments compared to pelagic ecosystems (Figure 3). Overall, this study provides methodological and conceptual insights to biodiversity surveys that increasingly make use of high-throughput sequencing technologies. This study has recently been accepted for publication in Molecular Ecology (Zinger *et al.*, 2014).

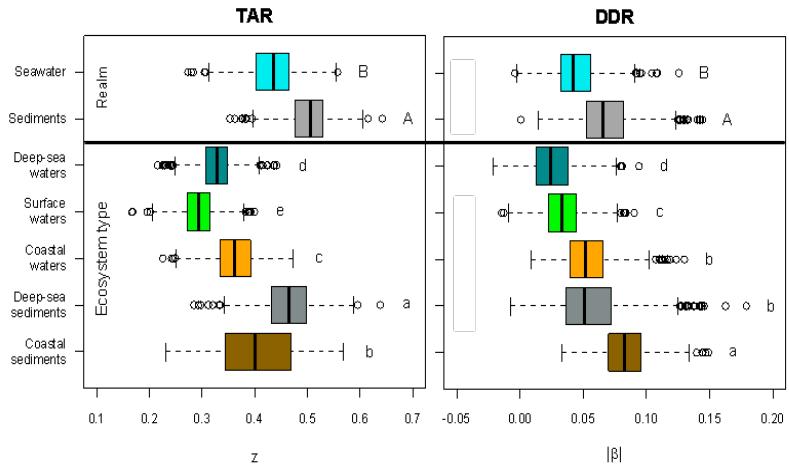


Figure 3. *Variation of the TAR and DDR slope coefficients (z and β respectively) according to realms and ecosystem types. Values per ecosystem type were standardized by randomly resampling 40 samples and 5,000 sequences per sample 1,000 times. Upper/lower case letters indicate significant differences (Mann-Whitney test, Holm-corrected, P<0.05) between realms or ecosystem types (from Zinger et al., 2014).*

2.3.3 Protists

In order to obtain an overall estimation of protist diversity within marine microbial communities, the CNRS-Roscoff team (Partner 6) has developed amplicon sequencing based on the hypervariable V4 and V9 sequences of 18S rRNA. The V9 sequence was found to be particularly amenable to current Illumina Hi-Seq sequencing technologies and so has been employed to assess eukaryotic diversity in samples collected during the *Tara* Oceans expedition. The approach has been used by Genoscope (Partner 17) to sequence more than 230 samples from 46 sampling stations, generating close to 500 million sequences from which 2.3 million unique ribotypes have been obtained. The sequences have been

taxonomically assigned using the PR2 18S reference database for eukaryotes (Guillou *et al.*, 2012). As expected, the opisthokonts (including metazoans and fungi) are the most abundant, whereas the most highly represented protists are from the Alveolata (e.g., dinoflagellates), Rhizaria (eg, radiolarians and foraminifera), and Stramenopiles (e.g., diatoms) (Figure 4).

One way to identify only the active fraction of the community is to compare the diversity obtained using rDNA with the diversity represented by reverse transcribed rRNA, based on the assumption that cell activity is proportional to the concentration of rRNA within the cell. Very few studies have explored the rDNA versus rRNA diversity and abundance in marine microbial eukaryotic communities. CNRS-Roscoff explored this issue using approximately 100 samples from the BioMarks programme derived from European coasts from Norway to the Black Sea, including 3 organismal size fractions and 3 depths (2 water column and sediment). Specific patterns obtained with both templates highlighted the importance of using the rRNA approach to complement the use of the rDNA barcode for a complete description of the actual metabolically active eukaryotic microbial diversity.

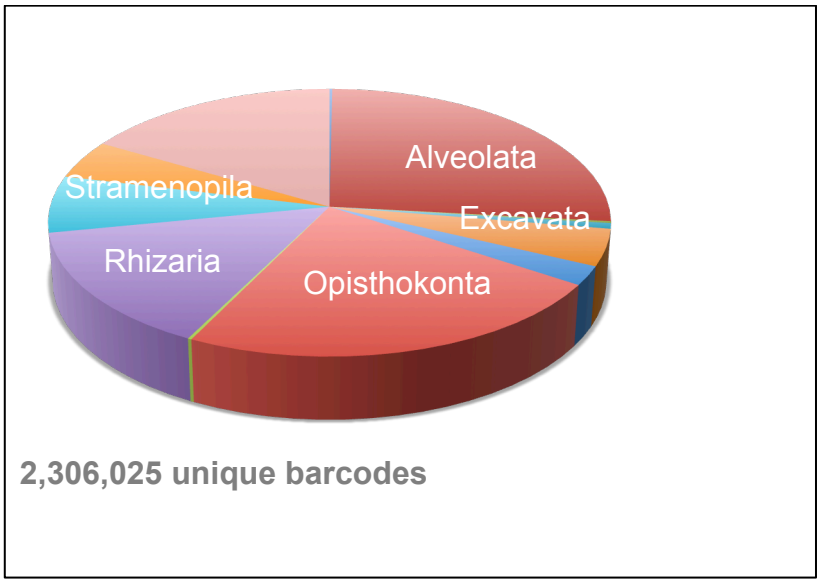


Figure 4. *Taxonomic assignments of barcodes obtained from 46* Tara *Oceans sampling sites distributed worldwide, indicating alpha-diversities of different eukaryote groups.*

A crucial issue with all DNA-based studies of diversity is to be able to relate sequence diversity and abundance with individuals. While this is still an unresolved issue for many of the smaller organisms that cannot be observed microscopically, the larger-sized diatoms can be assessed using both methods. Using samples from *Tara* Oceans, CNRS-ENS (Partner 6) and SZN (Partner 8) have compared microscope-based diatom counts with diatom abundance data based on V9 tags at a range of sampling sites and the results generally show a relatively good correspondence (e.g., Figure 5). This result is an important validation of the V9 metabarcoding approach, which should permit detailed studies of diatom abundance and diversity worldwide for the first time.

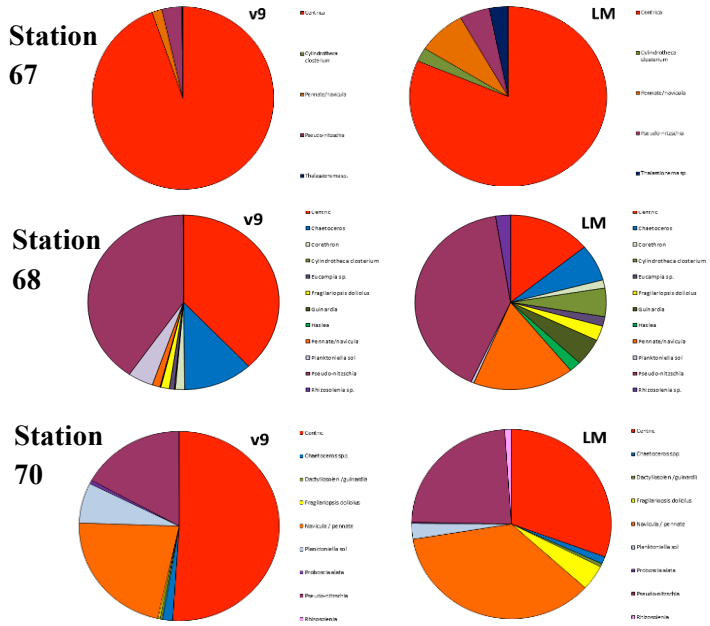


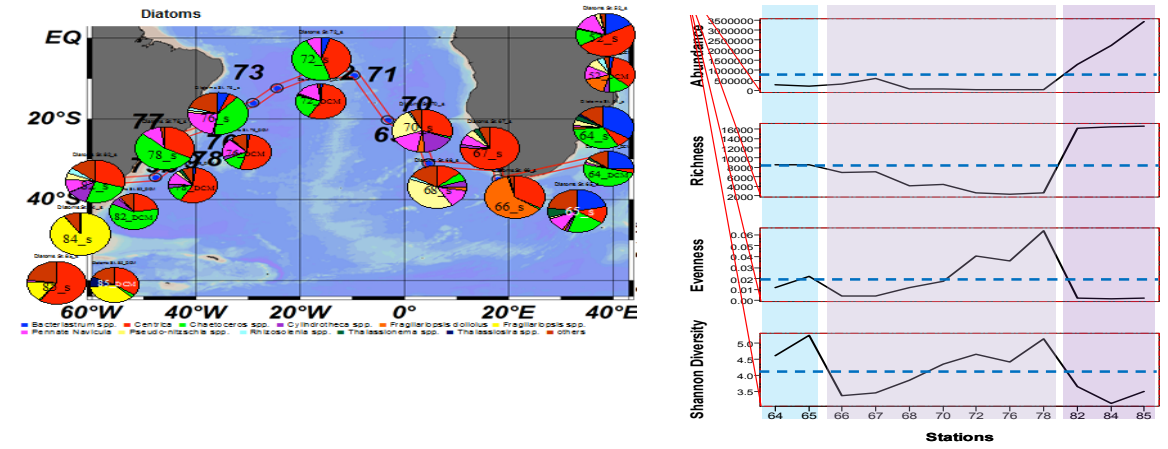
Figure 5. *Correspondence of diatom abundance determined by V9 ribotype counts (left) and light microscopy observations (right) at three Tara Oceans sampling sites. Each colour in the pie chart indicates a different diatom genus.*

CNRS-Roscoff has used the same BioMarks and *Tara* Oceans datasets of 18S ribotypes to examine the green plastid lineage (Chlorophyta), which has been relatively understudied until now within marine phytoplankton. Their studies have allowed for the first time to assess the importance of the different groups of green algae in different environments across various size fractions. In coastal waters Mamiellophyceae appear largely dominant (50% of V4 metabarcodes), especially within the pico-plankton. In contrast, in open ocean waters, prasinophyte clade VII, which has been very little studied, represents almost half of the sequences. A deeper analysis of prasinophyte clade VII (Lopes, unpublished) based on environmental sequences and cultured strains demonstrate that this group is formed of two major sub-clades (A and B) each containing different ecotypes, in particular linked to light levels (surface vs. DCM ecotypes), some of which are not yet available in culture.

2.3.4 Cross-kingdom analyses

A major objective of this Deliverable is to describe community structure across different kingdoms, and in an environmental context. At the current time, the *Tara* Oceans dataset provides the best opportunity to do this, and an example is shown here from the Agulhas retroflection region off South Africa. This region is an important chokepoint for ocean circulation between the Indian and Atlantic Oceans, and so it is of interest to examine how plankton communities change in this region of the global ocean.

Figure 6 shows how diatom communities change in this region, and a major shift is apparent, reflected also in a significant loss of diversity in the western side of Cape Agulhas with respect to the eastern side. Interestingly, this same pattern has been observed in the Cyanobacterial communities (Figure 6, lower panel; CNRS-Roscoff) but not for other bacteria (CSIC, data not shown) nor for viruses (CNRS-IGS, data not shown).



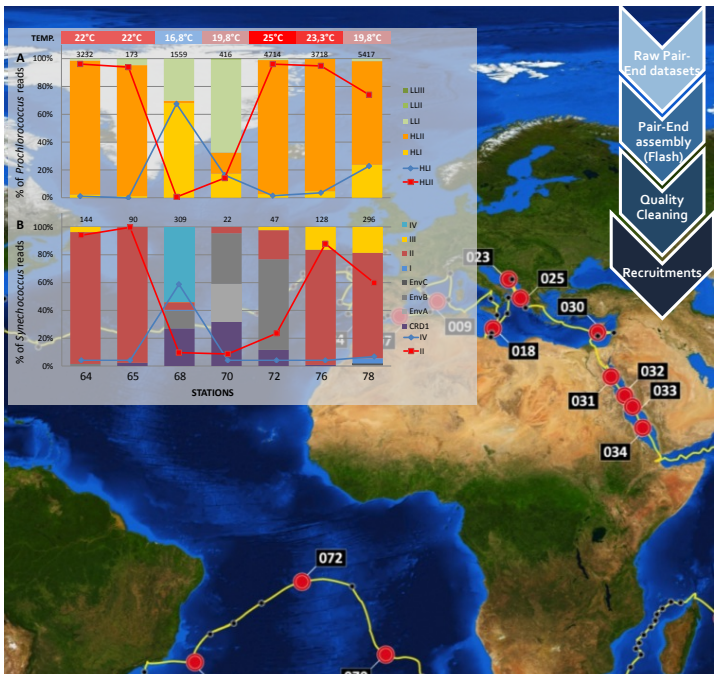


Figure 6. *Diatom and cyanobacterial community changes around Cape Agulhas. The top left panel shows diatom community changes as revealed by light microscopy counts (Partner 8-SZN). The top right panel shows the diversity shifts based on analysis of diatom-specific V9 ribotypes. In particular, the lower panel shows the dramatic reduction in the Shannon diversity between Stations 65 and 66, east and west of the Agulhas Cape, respectively (Partner 6-CNRS-ENS). The lower left panel shows the changes in Prochlorococcus (top graph) and Synechococcus (bottom plot) around the same region (Partner 6-CNRS-Roscoff).*

Additional examples of cross-kingdom analyses relate to studies of co-occurrence. CNRS-IGS and VIB (Partner 10) have already reported a previously unknown co-occurrence between an oomycete and a girus (Hingamp *et al.*, 2012).

Finally, the Deliverable also involved the refinement of global maps of marine microbe distributions by assessing their correlations with real data from the field. CNRS-ENS (Partner 6) has produced a global map of phytoplankton diversity based on analysis of spectral anomalies in remote sensing data (Figure 7). The relationship between such biodiversity proxy and sea surface temperature follows macroecological patterns that have been observed previously in classical studies of biodiversity and that are predicted by evolutionary models. The patterns are now being validated with data from the Atlantic Meridional Transect (AMT) and from *Tara* Oceans.

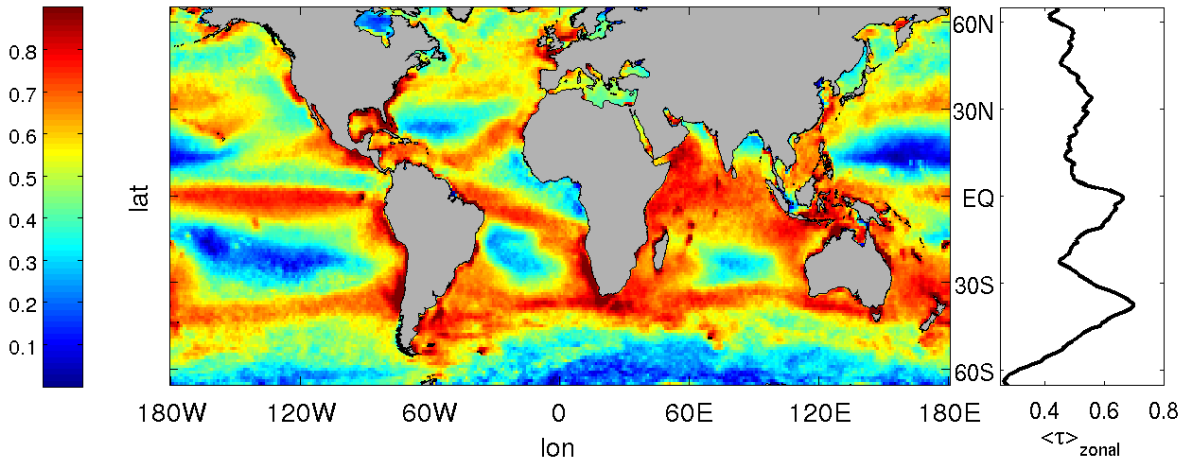


Figure 7. *Annual average of the area-based Shannon index computed from remote-sensing observations that is proposed as an indicator of global planktonic biodiversity distribution and hotspots (De Monte et al., 2013). To the right, the zonal mean of the index.*

2.4 Conclusions

The breadth of the analyses reported here reveal the tremendous progress that has been made by Micro B3 researchers aiming towards a comprehensive description of marine microbial community structure and composition in its environmental context, and indicates that the Deliverable has been accomplished with success.

The sampling protocols that have allowed the achievement of Deliverable 6.8 have been instrumental in establishing the appropriate methodologies for the Ocean Sampling Day (Work Package 2).

The statistical methods that have been employed to examine microbial community structure, although not yet sufficiently robust to permit the development of a standardized pipeline for such studies, nonetheless illustrate the enormous progress that has been made for the achievement of this Deliverable.

Reference list

Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, Liu H, et al. 2006. The marine viromes of four oceanic regions. PLoS Biol 4: e368

Arnaud-Haond S, Arrieta JM, Duarte CM. 2011. Global genetic resources. Marine biodiversity and gene patents. Science 331: 1521-1522

Arrigo KR. 2005. Marine microorganisms and global nutrient cycles. Nature 437: 349-355

Bowler C, Karl DM, Colwell RR. 2009. Microbial oceanography in a sea of opportunity. Nature 459: 180-184

De Monte S, Soccodato A, Alvain S, d'Ovidio F. 2013. Can we detect oceanic biodiversity hotspots from space? ISMEJ 1-3 doi: 10.1038/ismej.2013.72

Díez-Vives C, Gasol JM, Acinas SG. 2013. Spatial and temporal variability among *Bacteroidetes* populations in the NW Mediterranean Sea. Systematic Applied and Microbiology; doi 10.1016/j.syapm.2013.08.006

Falkowski PG, Fenchel T, Delong EF. 2008. The microbial engines that drive Earth's biogeochemical cycles. Science 320: 1034-1039

Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, et al. 2004. The evolution of modern eukaryotic phytoplankton. Science 305: 354-360

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. Science 281: 237-240

Fuhrman JA. 2009. Microbial community structure and its functional implications. Nature 459: 193-199

Gilbert JA, Meyer F, Antonopoulos D, Balaji P, Brown CT, et al. 2010. Meeting report: the terabase metagenomics workshop and the vision of an Earth microbiome project. Stand Genomic Sci 3: 243-248

Hingamp P, Grimsley N, Acinas SG, Clerissi C, Subirana L, Poulain J, Ferrera I, Sarmento H, Villar E, Lima-Mendez G, et al. 2013. Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. ISME J 7, 1678–1695. doi:10.1038/ismej.2013.59

Irigoien X, Huisman J, Harris RP. 2004. Global biodiversity patterns of marine phytoplankton and zooplankton. Nature 429: 863-867

Karl DM. 2007. Microbial oceanography: paradigms, processes and promise. Nat Rev Microbiol 5: 759-769

Karsenti E, Acinas SG, Bork P, Bowler C, De Vargas C, Raes J, Sullivan M, Arendt D, Benzoni F, Claverie J-M, et al. 2011. A holistic approach to marine eco-systems biology. PLoS Biol 9, e1001177

Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW. 2005. Photosynthesis genes in marine viruses yield proteins during host infection. Nature 438: 86-89

Logares R, Sunagawa S, Salazar S, Cornejo-Castillo FM, Ferrera I, Sarmento H, Hingamp P, Ogata H, de Vargas C, Lima-Mendez G, Raes J, Poulain J, Jaillon O, Wincker P, Kandels-Lewis S, Karsenti E, Bork P, Acinas SG. 2013. Metagenomic 16S rDNA Illumina-Tags as a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. Environmental Microbiology, doi 10.1111/1462-2920.12250

Nagasaki K. 2008. Dinoflagellates, diatoms, and their viruses. Journal Microbiology 46: 235–243

Rohwer F, Thurber RV. 2009. Viruses manipulate the marine environment. Nature 459: 207-212

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, et al. 2007. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol 5: e77

Sarmento H, Montoya JM, Vázquez-Domínguez E, Vaqué D, Gasol JM. 2010. Warming effects on marine microbial food web processes: how far can we go when it comes to predictions? Philos Trans Royal Soc London B Biol Sci 365: 2137-2149

Sharon I, Tzahor S, Williamson S, Shmoish M, Man-Aharonovich D, et al. 2007. Viral photosynthetic reaction center genes and transcripts in the marine environment. ISME J 1: 492-501

Sullivan MB, Lindell D, Lee J, Thompson LR, Bielawski JP, et al. 2006. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. PLoS Biol 4: e234

Zinger A, Boetius A, Ramette A. 2014. Bacterial taxa-area and distance-decay relationships in marine environments. Molecular Ecology, in press.