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 D6.8

Genome and metagenome datasets from all the samples and microbes analyzed

Organisation name of lead contractor: UF (Partner 18)

Authors: Wolfgang R. Hess (UF, Partner 18); Laurence Garczarek and Frédéric Partensky (CNRS Roscoff, MaPP team, Partner 16); S. L'Haridon, G. Burgaud, G. Le Blay (UBO S., Partner 3); Jörg Peplies (RIBO, Partner 22); Mario López-Pérez and Francisco Rodriguez-Valera (UMH, partner 11).

Due date of deliverable: M36
Actual submission date: M36

Revision: V.2

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)	
Dissemination Level	
PU Public	**x**
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)	

Summary

Write a short informative summary of your Deliverable (2 pages maximum), which should include the following elements:

Objective(s): The aim of this deliverable was to obtain genome and metagenome datasets from all the samples and microbes analysed. The obtained information then was to be used to infer information about the potential for secondary metabolite synthesis, growth specificities and ecophysiological adaptation. Moreover, comparative genomics is critical to understand the evolution and differential biosynthesis potential of the different taxonomical groups within marine picocyanobacteria, which are the most abundant and widespread representatives of marine phytoplankton.

Rationale: Total DNA was extracted, in case of marine fungi using the CTAB (hexadecyltrimethylammonium bromide) and Genomic-tip (Qiagen) method. The genomic DNA was used to generate shotgun and mate-pair libraries. Libraries were sequenced on an Illumina HiSeq 2500 sequencer, in case of newly isolated bacterial and archaeal strains using the Illumina MiSeq platoform, with 2x 250 bp-reads (MS44). Whole genomes of marine picocyanobacteria were assembled using the CLC *de novo* assembler and scaffolded using a custom-designed semi-automatic scaffolder (WiseScaffolder). Whole genome sequences of marine fungi were assembled using the ALLPATHS-LG whole-genome shotgun assembler. The genome sequences of selected marine picocyanobacteria were then closed manually. The information for marine picocyanobacteria was integrated into the information system Cyanorak v2 and all clusters were annotated automatically using a very rich functional annotation, including COG/EggNOG, EC and K numbers, GO terms, protein domains (Pfam, ProSite and InterPro) and TIGR roles as well as a custom-designed 'Cyanorak roles' detailing photosynthetic processes.

Results: Genome sequences of 2 newly isolated fungal strains, *Cadophora malorum* (UBOCC-A-108058) and *Cryptococcus* sp. (UBOCC-A-208024) were obtained. The search for gene clusters encoding enzymes for secondary metabolites revealed within the genome of *Cadophora malorum* 6 type I polyketide synthases, 5 non-ribosomal peptide synthetases, 2 hybrids PKS-NRPS and 2 terpene synthases genes and in the genome of *Cryptococcus* sp. one sequence of a type III polyketide synthase and also one terpene synthase gene.

The genome sequences of 97 marine picocyanobacteria, including 32 unpublished *Synechococcus* strains that were isolated from a variety of habitats and depths, were obtained and analyzed. Differences in the set of enzymes involved in the biosynthesis of lipids have been unveiled between *Synechococcus* clades that might be related to recently revealed differences in thermal preferences of strains representative of these clades. Moreover, the draft genomes of 4 newly isolated bacterial and of 3 archaeal strains were generated and analyzed.

Partner(s) involved in Deliverable production: UBO S., Partner 3; UMH, Partner 11; CNRS Roscoff, Partner 16; UF, Partner 18; RIBO, Partner 22

Genomes of marine cyanobacteria

Comparative genomics is critical to understand the evolution and differential biosynthesis potential of the different taxonomical groups within marine picocyanobacteria, which are the most abundant and widespread representatives of marine phytoplankton. In the last six month period (Month 30-36), **CNRS-Roscoff-MaPP** (Partner 16) has pursued the comparison of 95 marine picocyanobacteria, including 32 unpublished *Synechococcus* strains that were isolated from a variety of habitats and depths, essentially by **Partners 13 (UW) and 16** (**CNRS-Roscoff**; see **Table 1**). All but 3 of the latter 32 genomes were sequenced by the Genoscope (Evry, France) or the Center for Genome Research (Liverpool, UK), assembled using the CLC *de novo* assembler, scaffolded using a custom-designed semi-automatic scaffolder (WiseScaffolder) then closed manually (see technical details in our previous report for deliverable 6.5 and Farrant et al. BMC Bioinformatics, in revision). After inclusion in the information system Cyanorak v2 (application.sb-roscoff.fr/cyanorak/), the 252,974 genes were split between 27,832 cluster of orthologs (CoGs) and all clusters were annotated automatically using a very rich functional annotation, including COG/EggNOG, EC and K numbers, GO terms, protein domains (Pfam, ProSite and InterPro) and TIGR roles as well as a custom-designed 'Cyanorak roles' detailing photosynthetic processes. More than 900 of these CoGs were further manually annotated and more than 600 genes missed by ORF finding softwares were manually created and added to CoGs. The database also includes rRNA and tRNA. Comparative genomics of marine picocyanobacteria using Cyanorak v2 was used to refine the sets of core, accessory and unique genes for these microorganisms and has enlightened the key role of genomic islands that are shared between distinct clades of strains (Farrant, Doré et al., in prep. for Genome Biology).

Differences in the set of enzymes involved in the biosynthesis of lipids have been unveiled between *Synechococcus* clades (Piterra et al., in prep.) that might be related to recently revealed differences in thermal preferences of strains representative of these clades (Pittera et al., 2014; see Task 3.1).

Table 1: Genome characteristics of the 32 novel Synechococcus strains annotated using the Cyanorak information system.

Strain Name	SubCluster	Clade	SubClade	Pigment type	Sequencing center	Genome status	Genome size	CDS	rRNA	tRNA
Syn_A15-127	5.1	WPC1	WPC1	3c	Genoscope	Complete	2 543 463	2 836	6	44
Syn_A15-24	5.1	III	IIIa	3c	CGR, Liverpool	Complete	2 305 373	2 739	6	45
Syn_A15-28	5.1	III	IIIb	3c	Genoscope	Complete	2 341 586	2 773	6	45
Syn_A15-44	5.1	II	IIa	2	Genoscope	Complete	2 621 965	3 170	9	46
Syn_A15-60	5.1	VII	VIIa	3c	Genoscope	Complete	2 543 402	3 086	6	44
Syn_A15-62	5.1	II	IIc	3dB	Genoscope	Complete	2 294 140	2 801	6	44
Syn_A18-25c	5.1	VII	VIIa	3c	CGR, Liverpool	Complete	2 511 360	3 050	6	44
Syn_A18-40	5.1	III	IIIa	3dB	CGR, Liverpool	Complete	2 401 547	2 801	6	45
Syn_A18-46.1	5.1	III	IIIa	3c	CGR, Liverpool	Complete	2 471 770	2 918	6	46
Syn_BIOS-E4-1	5.1	CRD1	CRD1b	3cA	Genoscope	WGS	3 314 220	4 472	6	44
Syn_BIOS-U3-1	5.1	CRD1	CRD1c	3dA	Genoscope	Complete	2 710 834	3 458	6	44
Syn_BMK-MC-1	5.1	V	V	2	Genoscope	Complete	2 601 150	3 095	6	44
Syn_BOUM118	5.1	III	IIIa	3c	Genoscope	WGS	2 326 334	2 765	6	44

Syn_M16.1	5.1	II	IIa	3a	Genoscope	Complete	2 112 236	2 504	6	45
Syn_MEDNS5	5.1	VI	VIa	3c	Genoscope	Complete	2 284 343	2 768	6	45
Syn_MINOS11	5.3	5.3	5.3	3dB	Genoscope	Complete	2 284 343	2 665	3	42
Syn_MITS9220	5.1	CRD1	CRD1a	3dA	CGR, Liverpool	Complete	2 424 175	3 069	6	44
Syn_MVIR-18-1	5.1	I	Ib	3aA	Genoscope	Complete	2 451 974	3 054	6	44
Syn_NOUM97013	5.1	VII	VIIb	3a	Genoscope	Complete	2 552 712	2 964	6	45
Syn_PROS-7-1	5.1	VI	VIb	2	Genoscope	Complete	2 565 218	2 888	6	45
Syn_PROS-9-1	5.1	I	Ib	3dA	CGR, Liverpool	Complete	2 273 940	2 796	6	45
Syn_PROS-U-1	5.1	II	IIh	3dB	Genoscope	Complete	2 576 003	3 140	6	45
Syn_ROS8604	5.1	I	Ib	3a	Genoscope	Complete	2 876 904	3 765	6	44
Syn_RS9902	5.1	II	IIa	3c	Genoscope	Complete	2 481 217	2 973	9	46
Syn_RS9907	5.1	II	IIa	3a	Genoscope	Complete	2 581 635	2 982	9	47
Syn_RS9909	5.1	VIII	VIII	1	Genoscope	Complete	2 603 169	2 811	6	45
Syn_RS9915	5.1	III	IIIa	3dB	Genoscope	WGS	2 417 983	2 884	6	44
Syn_SYN20	5.1	I	Ib	3a	CGR, Liverpool	Complete	2 728 628	3 401	6	44
Syn_TAK9802	5.1	II	IIa	3a	Genoscope	Complete	2 190 394	2 669	6	45
Syn_WH8101	5.1	VIII	VIII	1	Genoscope	Complete	2 623 023	2 884	6	44
Syn_WH8103	5.1	III	IIIa	3bB	Genoscope	Complete	2 429 688	2 850	6	46

The taxonomic assignment at sub-cluster and clade levels derive from Dufresne et al. (2008) while the assignment at sub-clade level modified from Mazard et al. (2012). The assignment of pigment types come from Humily et al. (2013).

Detection of putative secondary metabolites or antibiotics. AntiSMASH was used by **Partner 16** to detect possible antibiotics or secondary metabolites in the 32 new genomes. Most strains did not display any result or only seemingly false predictions. However, a number of strains (including *Synechococcus* A15-24) was found to possess one gene coding for a homolog of Type III PKS coding a putative chalcone or stilbene synthase (Cyanorak cluster CK_00001781) that is definitely worth further screening.

Partner 18 (UF) has been leading the draft genome analysis of *Nodularia spumigena* sp. CCY9914, a cyanobacterium representative of the nitrogen-fixing algal summer blooms of the brackish waters of the Baltic. This has been a collaborative project involving three MaCuMBA partners (**Partner 1, NIOZ; Partner 18, UF; Partner 22, UW**) and was published mentioning support by the MaCuMBA project (Voß et al., 2013). In a collaboration with the Arizona State University and the University of Sydney, **Partner 18** was involved in generating the draft genome sequence of the filamentous cyanobacterium *Leptolyngbya* sp. Strain Heron Island J, exhibiting chromatic acclimation. Also this dataset is publicly available and the publication is mentioning support by the MaCuMBA project (Paul et al., 2014).

Genomes of newly isolated fungal strains

UBO-LUBEM (Partner 3) has sequenced the genomes of 2 newly isolated fungal strains. High quality DNA of the fungal strains ***Cadophora malorum*** (UBOCC-A-108058) and ***Cryptococcus*** sp. (UBOCC-A-208024) were extracted using the CTAB (hexadecyltrimethylammonium bromide) and Genomic-tip (Qiagen) method. Genomic DNA was used to generate shotgun and mate-pair libraries. Libraries were sequenced on Illumina HiSeq 2500 sequencer and whole genomes were assembled at the Macrogen Inc. (Seoul, South Korea).

***Cryptococcus* sp.** (Mo29): The shotgun library produced 41,481,610 reads. The mate-pair library produced 15,799,276 reads. After quality filtering, 28,246,332 shotgun reads (2,85248 Mb) and 7,870,684 mate-pair reads (732 Mb) were retained. Whole genome sequence was assembled from both libraries using ALLPATHS-LG whole-genome shotgun assembler. The assembly contained a total of 174 scaffolds with average read length of 144,951 bp. The N50 was 820 kb, and the maximum contig length was 1,690 kb. The total sequence length of the resulting draft genome was 27,528,793 bp with an overall GC content of 50.09%.

Cadophora malorum (Mo12) : The shotgun library produced 39,507,284 reads. The mate-pair library produced 15,277,956 reads. After quality filtering, 28,200,910 shotgun reads (2,848 Mb) and 8,857,004 mate-pair reads (815 Mb) were retained. Whole genome was assembled from both libraries using ALLPATHS-LG whole-genome shotgun assembler. The assembly contained a total of 164 scaffolds with average read length of 299,210 bp. The N50 was 1,408 kb, and the maximum contig length was 1,707 kb. The total sequence length of the resulting draft genome was 54,281,849 bp with an overall GC content of 47.08%.

Preliminary analysis of secondary metabolite biosynthesis gene clusters were performed using antiSMASH software (Weber et al., 2015). The genome of *Cadophora malorum* harbored 6 type I polyketide synthases, 5 non-ribosomal peptide synthetases, 2 hybrids PKS-NRPS and 2 terpene synthases genes. One the other hand, the genome of *Cryptococcus* sp. harbored one original sequence of type III polyketide synthase and also one terpene synthase gene.

Genomes of newly isolated bacterial and archaeal strains

UBO-LM2E (Partner 3) has sequenced the draft genomes of **4 newly isolated bacterial strains and 3 archaeal** strains using the Illumina MiSeq technology, with 2x 250 bp-reads (MS44). Two archaeal type strains have also been sequenced.

The draft genomes of an *Alphaproteobacterium* called ***Phaeobacter leonis* strain 306T**, and a *Gammaproteobacterium* called ***Halomonas lionensis* strain RHS90T** were obtained and published (Gaboyer et al., 2013, 2014). In addition, comparative genomics was done with the genome of the generalist species *Halomonas lionensis* and revealed a physiological potential that may explain the ecological success of the genus *Halomonas* in common and extreme environments. The estimated genome size of *Phaeobacter leonis* strain 306T totalizes 4,823,053 bp with a G+C content of 58.7%. It encodes 5578 putative genes, a single rRNA operon, 43 predicted tRNAs and has a coding density of 0.84. Among these genes, 78.4% and 52% could be classified respectively in Cluster of Orthologous Groups (COG) or in the SEED subsystems, and 25.4% were annotated as hypothetical proteins. The draft genome size of *Halomonas lionensis* strain RHS90T is 3,906,070 bp with a G+C content of 55.9%. It encodes 4,734 putative genes, a single 16S rRNA gene and has a coding density of 0.77. A total of 78.4% and 52% of all genes could be respectively classified into COG or SEED subsystems. Analysis of protein-coding genes revealed genes that might explain its capacity to cope with stressful conditions, like metal-tolerance genes or oxidative stress genes. Genes involved in the response to Carbon Starvation and to Cold Shock were also detected. A complete DNA repair system was also found, with genes involved in nucleotide excision repair, base excision repair, mismatch repair, SOS response or homologous recombination. Numerous histidine kinases or response regulators were predicted and could enable *H. lionensis* to sense external signals. Adaptation to halophily was suggested by acidic-shifted pI values of its proteome. Genes involved in osmoprotectants synthesis and uptake of ectoine, betaine or polyols osmolytes are also present in the genome and their synthesis has been demonstrated in vivo. Genes required for polyhydroxyalkanoates synthesis were also predicted and the synthesis of PHA has been experimentally demonstrated. The genome sequencing of ***Kosmotoga pacifica*** belonging to the thermotogales was finished and assembled in one contig and has been submitted for publication (Standard in Genomic Science). One additional bacterial strain ***Desulfovibrio indicus* sp. nov**. strain J2, isolated from the Southwest Indian Ridge is being sequenced.

The genomes of 5 archaeal strains belonging to the genus *Methanohalophilus*, 3 new isolates (SLHTYRO, SLHKRYOS, SLHTHETIS) coming from a Deep Anoxic Hypersaline Basin (Tyro, Kryos, Thetis) and two type strains ***Methanohalophilus portucalensis* strain FDF-1T** and ***M. halophilus* strain Z-7982T** have been also sequenced. One more archaeal strain, *Thermococcus superprofundus* strain **CDGS**, isolated from an hydrothermal vent of the Caiman Trough is being sequenced.

Partner 15 (UMIL) has sequenced the genome of ***Virgibacillus pantothenticus* 21D**. This bacterium has been isolated on 246 DSM medium from the seawater-brine interface sample of the deep hypersaline anoxic basins of the Eastern Mediterranean (De Vitis et al., 2015). The strain has the growth optimum at concentration of 6-9% NaCl.

Partner 11 (**UMH**) has sequenced the genome two novel Alphaproteobacteria strains, R1-200B4 (T) and R2-400B4, isolated from the Mediterranean Sea off the coast of Alicante, Spain. The phylogenetic analysis of the 16S rRNA gene showed that they are related to members of the Family Rhizobiaceae. The major fatty acids are those from summed feature 8 (C18:1 ω6c/C18:1 ω7c) and the C16:0. Catalase and oxidase were positive. Nitrate reduction and aesculin hydrolysis were positive. Production of β-galactosidase and urease was positive. The production of indol, arginine dehydrolase or gelatinase was negative. Growth was observed in presence of 7% NaCl. Therefore, based on the phylogenetic, chemotaxonomic and phenotypic data obtained in this study, was proposed to classify the strains isolated in a new genus named *Pseudorhizobium* gen. nov. and a new species named *Pseudorhizobium pelagicum* sp. nov. with the type strain R1-200B4T (= LMG 28314T = CECT 8629T) (Kimes et al. 2015). UMH has also sequenced and assembled into a single contig the genome of the two isolates, *Alteromonas. australica* DE170 from the South Adriatic at 1000 m depth and *A. australica* H17(T) isolated from the first metres of the Tasman Sea (16000 km away) (López-Pérez et al., 2014). In spite of the different locations and depth of the sampling sites, the two *A. australica* strains form a highly homogeneous clade with an average nucleotide identity (ANI) of 98.6%. The patterns of variation observed between the two *A. australica* strains were very similar to the ones found before when compared them with all the available genomes of strains within the genus *Alteromonas*. Among the specific metabolic features found for the *A. australica* isolates there is the degradation of xylan and production of cellulose as extracellular polymeric substance by DE170 or the potential ethanol/methanol degradation by H17(T). The UMH collection include other 24 new *Alteromonas* isolates that were sequenced, assembled and annotated, representing at least two new species. We will study this large collection of complete and fully assembled genomes in order to analyse the patterns of variations and evolution of this marine bacterium.

UMH has also used deep metagenomic sequencing (in combination with flow cytometry and FISH) to describe the two major groups of planktonic Actinobacteria from the marine habitat. The first, belonging to the new order Actinomarinales with very low GC content (33%) and the smallest free living cells described yet (cell volume ca. 0.013 µm^3). Therefore, with additional evidence of the complete 16S and the 23S genes at hand, UMH proposed the creation of the new sub-class, 'Candidatus Actinomarinidae', (order 'Ca. Actinomarinales', sub-order 'Ca. Actinomarineae', family 'Ca. Actinomarinaceae') for the taxonomic placement of this group of microbes. Metagenomic fosmids allowed a virtual genome reconstruction that also indicated very small genomes below 1 Mb. They inhabit the deep chlorophyll maximum and have a new kind of rhodopsin indicating a photoheterotrophic lifestyle (Ghai et al., 2013). The second group is the first that represent marine microorganisms belonging to the order Acidimicrobiales and only the second group of planktonic marine *Actinobacteria* to be described at a genomic level. These microbes inhabit the deeper photic zone and have streamlined genomes (Mizuno et al., 2015). UMH has described four nearly complete genomes of these marine *Actinobacteria* providing insights into the lifestyle of this diverse and important group of microbes. A novel rhodopsin clade, acidirhodopsins, related to freshwater actinorhodopsins was found in these organisms. In the same way, applying deep metagenomics sequencing to the microbial community of a freshwater reservoir, **UMH** was able to circumvent the traditional bottleneck and reconstruct by *de novo* assembly some distinct streamlined actinobacterial genomes. **UMH** has recovered and described seven genomes from the most abundant, free-living, low-GC Actinobacteria of the microbial community of a freshwater reservoir, some of which are related to the marine Acidimicrobiales genomes (Ghai et al., 2014). From a metagenomic library constructed from the biomass recovered at the DCM in the Mediterranean, we sequenced about 7000 fosmid clones and we have analyzed the Euryarchaeota group IIB genome fragments present in these fosmids. Euryarchaeota group IIB contigs assembled from different DCM metagenomes obtained from the same location at different times were also included in this study. With all these data, UMH has published Martin-Cuadrado et al. (2014), with the first genomic information about a new class of abundant low GC marine Euryarchaeota group IIB, which has been named Thalassoarchaea.

Based on their contribution to the Deliverables D6.4 and D6.5, Partner 22 **RIBO** has provided bioinformatics infrastructure and expertise in genome data analysis related to Deliverable D6.8 by supporting other MaCuMBA partners (within and also outside of WP6). In particular, close collaboration has been carried out with partners UvA and NIOZ as outlined below, also including visits at partner Ribocon for training and joined data mining in the context of both projects.

UvA project (contact: Gerard Muyzer, Anne-Catherine Ahn) In the context of genetic diversity and biogeography of haloalkaliphilic sulphur-oxidizing bacteria belonging to the genus Thioalkalivibrio, 77 genomes of isolates from global Soda lakes were compared based on the level of average nucleotide identities (ANI) and tetra nucleotide signatures (TETRA) for species differentiation. Soda lakes are the only habitats on Earth which provide a stable extreme alkaline condition (pH 9.5–11) due to the high buffering capacity of sodium carbonate. The genome comparison methods used are also implemented by the JSpeciesWS web server, which has been set up by partner RIBO within the MaCuMBA project but however does not allow for calculating such large matrices online. As a result of the study, all genomes compared show high identity on the 16S rRNA gene level but can be very good separated based on whole genome approaches. Additional visualization of the data, e.g. based on genome trees, was done for further evaluation.

Metagenomic analyses

A metagenomic dataset was analyzed by Partner 18 (UF), using a sample from a decaying *Trichodesmium* bloom taken during the VAHINE project led by Sophie Bonnet (IRD/MIO Noumea/Marseille) in the South West Pacific (New Caledonia) in Feb 2013. The metagenome was sequenced from three pooled samples that in parallel were processed to generate 3 separate metatranscriptome datasets (see deliverable 6.9). Sequencing was performed by BGI Tech Solutions (China) on an Illumina 1.5 platform. After quality filtering, the shotgun library produced 59,546,716 reads, corresponding to 59.546 Mb. Metagenomic assemblies were obtained using Ray Meta assemblers and the gene prediction and annotation was performed using the Prodigal (Hyatt et al., 2012) and Rapid Annotations Using Subsystems Technology (Aziz et al., 2008) platforms.

The assembly contained a total of 309,886 contigs with an average length of 296.5 bp, the N50 was 305 bp and the maximum contig length was 265.071 bp. The metagenome was dominated by three types of sequences. As expected, a high percentage of sequences closely matching the *Trichodesmium* reference genome. Secondly, a high number of sequences matching *Alteromonas*, a copiotroph bacterium, probably benefitting from the high nutrient concentrations released from the decaying *Trichodesmium* bloom. The third type of sequences occurred in high copy numbers that were assembled into the largest contig of 265 kb, remotely matching database sequences of giant phages.

Partner 11 (**UMH**) has performed several trips to collect samples from pelagic Mediterranean seawater. Metagenomic samples were collected at three different depths of the water column including the deep chlorophyll maximum (DCM), one during the summer (September) and another in winter (February). DCM is a zone of maximal photosynthetic activity, generally located toward the base of the photic zone in lakes and oceans. In the tropical waters, this is a permanent feature, but in the Mediterranean and other temperate waters, the DCM is a seasonal phenomenon. For each depth three size fractions (5-20 microns, 0.1 to 5 and 0.1 to 100 kDa) that retain respectively eukaryotic and particle attached prokaryotic microbes, free-living prokaryotes and viruses were used. DNA sequencing was performed using Illumina Hi-seq 2000 obtaining 10-30 Gb of sequencing data from each data set. The assembled data for each metagenome provided between 15,000-118,000 contigs > 1000 bp. We used the phylogenetic classification of both annotated contigs > 10 and rRNA sequences (>~100 bp) (Figure 1) from the raw metagenomic reads to identify the dominant microorganisms in each metagenome. At the level of major phylogenetic clades the community structure was remarkably well conserved, Alphaproteobacteria dominated followed by Gammaproteobacteria and Cyanobacteria. One major exception was the low proportion of cyanobacteria in the July 2012 sample what might derive from a slight deviation from the real maximum, i.e. the sample was taken slightly above the DCM.

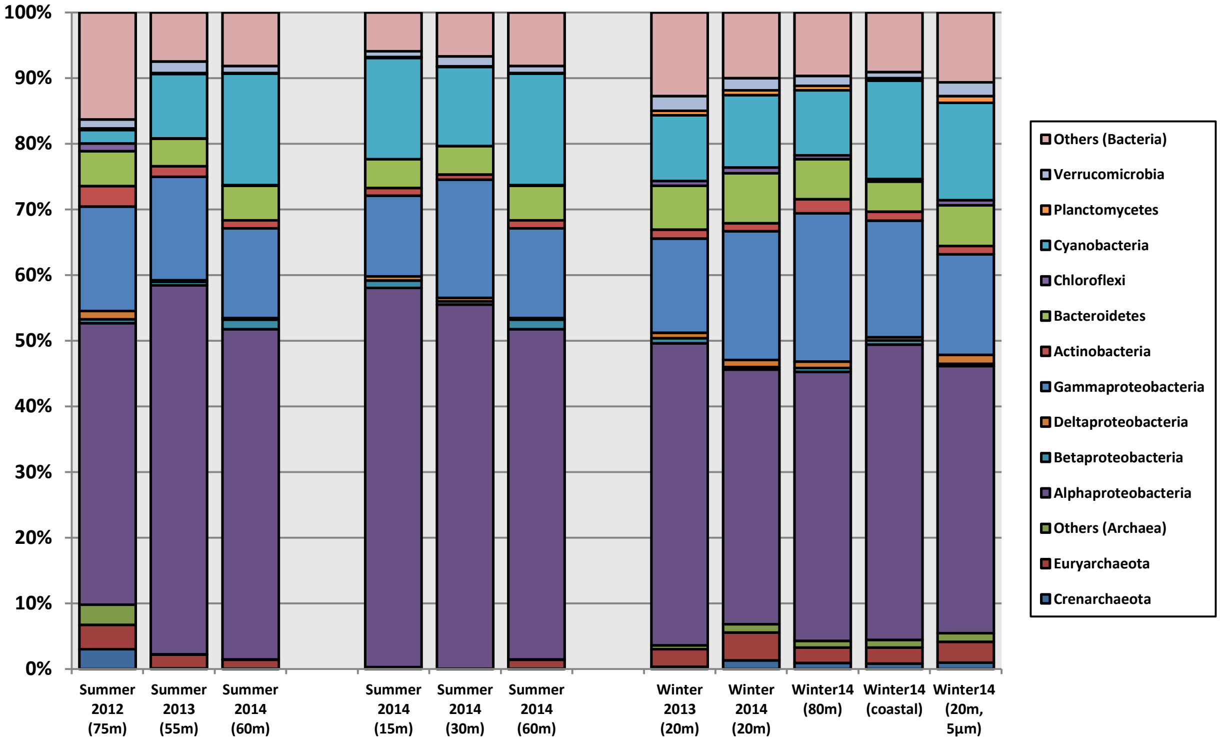

Figure 1. Phylogenetic classification based on annotated rRNA reads

References (*papers acknowledging MaCumBa)

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. (2008) The RAST Server: Rapid annotations using subsystems technology. **BMC Genomics** 9: 75.

Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, Paulsen IT, Tandeau de Marsac N, Wincker P, Dossat C, Ferriera S, Johnson J, Post AF, Hess WR and Partensky F. Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. **Genome Biology** 9(5): R90.

*De Vitis V, Guidi B, Contente ML, Granato T, Conti P, Molinari F, Crotti E, Mapelli F, Borin S, Daffonchio D, Romano D. Marine Microorganisms as Source of Stereoselective Esterases and Ketoreductases: Kinetic Resolution of a Prostaglandin Intermediate. **Mar Biotechnol** 2:144-52. doi: 10.1007/s10126-014-9602-z.

*Farrant GK, Hoebeke M, Partensky F, Andres G, Garczarek L. WiseScaffolder: an algorithm for the semi-automatic scaffolding of Next Generation Sequencing data. **BMC Bioinformatics** (In revision).

Gaboyer F, Tindall BJ, Ciobanu MC, Duthoit F, Le Romancer M, Alain K (2013). *Phaeobacter leonis* sp. nov., an alphaproteobacterium from Mediterranean Sea sediments. **Int J Syst Evol Microbiol.** 63(Pt 9):3301-6.

*Gaboyer F, Vandenabeele-Trambouze O, Cao J, Ciobanu MC, Jebbar M, Le Romancer M, Alain K (2014) Physiological features of *Halomonas lionensis* sp. nov., a novel bacterium isolated from a Mediterranean Sea sediment. **Res Microbiol.** 165(7):490-500.

*Ghai R, Mizuno CM, Picazo A, Camacho A, Rodriguez-Valera F. (2013). Metagenomics uncovers a new group of low GC and ultra-small marine Actinobacteria. **Sci Rep** 3: 2471.

*Ghai R, Mizuno CM, Picazo A, Camacho A, Rodriguez-Valera F. (2014). Key roles for freshwater Actinobacteria revealed by deep metagenomic sequencing. Molecular Ecology doi:doi: 10.1111/mec.12985.

*Humily F, Partensky F, Six C, Farrant GK, Ratin M, Marie D and Garczarek L (2013). A gene island with two possible configurations is involved in chromatic acclimation in marine *Synechococcus*. **PLoS One** 8: e84459.

Hyatt D, Locascio PF, Hauser LJ, Uberbacher EC (2012) Gene and translation initiation site prediction in metagenomic sequences. **Bioinformatics** 28(17):2223-2230.

*Kimes NE, López-Pérez M, Flores-Félix JD, Ramírez-Bahena MH, Igual JM, Peix A, Rodriguez-Valera F, Velázquez E. (2015). Pseudorhizobium pelagicum gen. nov., sp. nov. isolated from a pelagic Mediterranean zone. **Syst Appl Microbiol** 38(5):293-9. doi: 10.1016/j.syapm.2015.05.003.

*López-Pérez M, Gonzaga A, Ivanova EP, Rodriguez-Valera F (2014). Genomes of *Alteromonas australica*, a world apart. **BMC Genomics**. 18;15:483. doi: 10.1186/1471-2164-15-483.

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

*Martin-Cuadrado AB, Garcia-Heredia I, Moltó AG, López-Úbeda R, Kimes NE, López-García P, Moreira D, Rodriguez-Valera F. (2015) A new class of marine Euryarchaeota group II from the mediterranean deep chlorophyll maximum. **ISME J** 7:1619-34. doi: 10.1038/ismej.2014.249.

*Mizuno CM, Rodriguez-Valera F, Ghai R. (2015) Genomes of planktonic Acidimicrobiales: widening horizons for marine Actinobacteria by metagenomics. **MBio.** 10;6(1). doi: 10.1128/mBio.02083-14

*Paul R., Jinkerson R.E., Buss K., Steel J., Mohr R., Hess W.R., Chen M., Fromme P. (2014) Draft genome sequence of the filamentous cyanobacteria *Leptolyngbya Heron Island* strain J exhibiting chromatic acclimation. **Genome Announcements** *2(1).* e01166-13

Six C, Thomas JC, Garczarek L, Ostrowski M, Dufresne A, Blot N, Scanlan DJ and Partensky F (2007). Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. **Genome Biol** 8: R259.

*Voß B., Bolhuis H., Fewer D., Kopf M., Möke F., Haas F., El-Shehawy R., Hayes P., Bergman B., Sivonen K., Dittmann E., Scanlan D.J., Hagemann M., Stal L.J., Hess W.R. (2013) Insights into the physiology and ecology of the brackish-water-adapted cyanobacterium *Nodularia spumigena* sp. CCY9414 based on a genome-transcriptome analysis. **PLoS ONE**, *8(3):* e60224.1- e60224.22.

List of reviewers

Issue	Date	Implemented by
v.1	Aug-4-2015	CNRS-Roscoff
v.2	Aug-20-2015	UF

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.*