environmental microbiology

Environmental Microbiology (2012) doi:10.1111/j.1462-2920.2012.02827.x

Metagenomic analysis of hadopelagic microbial assemblages thriving at the deepest part of Mediterranean Sea, Matapan-Vavilov Deep

Francesco Smedile,[1†] Enzo Messina,[1†] Violetta La Cono,[1] Olga Tsoy,[2] Luis S. Monticelli,[1] Mireno Borghini,[3] Laura Giuliano,[1,4] Peter N. Golyshin,[5] Arcady Mushegian[2,6] and Michail M. Yakimov[1*†]

[1]*Institute for Coastal Marine Environment, CNR, Spianata S.Raineri 86, 98122 Messina, Italy.*
[2]*Stowers Institute for Medical Research, 1000 E 50th St., Kansas City, MO 64110, USA.*
[3]*Institute for Marine Sciences, ISMAR-CNR, Forte S.Teresa, 19136 Pozzuolo di Lerici, La Spezia, Italy.*
[4]*Mediterranean Science Commission (CIESM),16 bd de Suisse, MC 98000, Monaco.*
[5]*School of Biological Sciences, Bangor University, ECW Bldg Deiniol Rd, Bangor, Gwynedd LL57 2UW, UK.*
[6]*Department of Microbiology, Kansas University Medical Center, 3901 Rainbow Boulevard, Kansas City, KS 66160, USA.*

Summary

The marine pelagic zone situated > 200 m below the sea level (bls) is the largest marine subsystem, comprising more than two-thirds of the oceanic volume. At the same time, it is one of the least explored ecosystems on Earth. Few large-scale environmental genomics studies have been undertaken to examine the phylogenetic diversity and functional gene repertoire of planktonic microbes present in mesopelagic and bathypelagic environments. Here, we present the description of the deep-sea microbial community thriving at > 4900 m depth in Matapan-Vavilov Deep (MVD). This canyon is the deepest site of Mediterranean Sea, with a deepest point located at approximately 5270 m, 56 km SW of city Pylos (Greece) in the Ionian Sea (36°34.00N, 21°07.44E). Comparative analysis of whole-metagenomic data revealed that unlike other deep-sea metagenomes, the prokaryotic diversity in MVD was extremely poor. The decline in the dark primary production rates, measured at 4908 m depth, was coincident with overwhelming dominance of copiotrophic *Alteromonas macleodii* 'deep-ecotype' AltDE at the expense of other prokaryotes including those potentially involved in both autotrophic and anaplerotic CO_2 fixation. We also demonstrate the occurrence in deep-sea metagenomes of several clustered regularly interspaced short palindromic repeats systems.

Received 17 April, 2012; revised 21 June, 2012; accepted 24 June, 2012. *For correspondence. E-mail michail.yakimov@iamc.cnr.it; Tel. (+39) 0906015437; Fax (+39) 090669007. †Equal contribution.

Introduction

The water masses located in aphotic pelagic zone, i.e. > 200 m bls, are characterized by the absence of solar radiation and represent the largest marine subsystem, comprising 1.3×10^9 km^3 or about two-thirds of the oceanic volume (Aristegui *et al.*, 2009). Despite their massive scale, the deep dark ocean and in particular, the bathypelagic (1000–4000 m bls) and abyssopelagic (> 4000 m bls) zones are among the least-explored ecosystems on Earth. Even less is known about the hadopelagic or trench environments, the deepest marine habitats. These unique topographic features are seafloor depressions formed at convergent margins where one lithospheric plate moves below another (Stern, 2002). In extreme cases they can reach almost 11 000 m in depth (Mariana Trench) with the hydrostatic pressure exceeding 1100 standard atmospheres (110 MPa).

Marine microbial communities, centrally involved in the fluxes of matter and energy in the oceans, are major drivers of global biogeochemical cycling (Arrigo, 2005). Compared with the euphotic layer, our understanding of the biogeochemical processes in this ocean's 'inner space' is very limited. Within pelagic deep-sea settings, microbial communities were believed to largely subsist on photosynthetically derived organic matter in the form of sinking particulate organic matter (POM), as well as the flux of semi-labile dissolved organic carbon (Reinthaler *et al.*, 2005; Hansell *et al.*, 2009). More recently, measurements of carbon assimilation within free-living mesopelagic and bathypelagic communities identified thaumarchaeal autotrophic carbon fixation as an additional significant source of carbon supporting microbial life in the deep sea (Reinthaler *et al.*, 2010; Yakimov *et al.*, 2011).

Recently, metagenomic and metatranscriptomic studies fundamentally advanced our knowledge on abundance, diversity and gene content of planktonic microbes and their communities thriving in superficial and epipelagic compartments of the sea (Venter *et al*., 2004; Rusch *et al*., 2007; Frias-Lopez *et al*., 2008; Poretsky *et al*., 2009; Shi *et al*., 2009; Feingersch *et al*., 2010). Fewer large-scale environmental metagenomic surveys have been directed to examine the phylogenetic diversity and functional diversity of microorganisms present in aphotic mesopelagic, bathypelagic and hadopelagic environments (DeLong *et al*., 2006; Martin-Cuadrado *et al*., 2007; Konstantinidis *et al*., 2009; Eloe *et al*., 2011a). Recent studies of deep-sea studies included analyses of microbial communities from depths of (i) 3000 m at the Ionian Station Km3 in Mediterranean Sea (Martin-Cuadrado *et al*., 2007), (ii) 4000 m and 4400 m at the Hawaii station ALOHA in North Pacific Subtropical Gyre (DeLong *et al*., 2006; Pham *et al*., 2008; Brown *et al*., 2009), (iii) 4121 m in North Atlantic (Sogin *et al*., 2006) and (iv) 6000 m Puerto Rico Trench (Eloe *et al*., 2011a). These studies indicated that deep-sea microbial populations exhibit lifestyles distinct from those of surface water microorganisms and is consistent with the differences observed in particle associated (mainly copiotrophic) versus free-living (mainly oligotrophic) bacteria (Lauro and Bartlett, 2008; Lauro *et al*., 2009). The number of transposable elements, as well as genes involved in phage-host arms race, such as phage integrases and clustered regularly interspaced short palindromic repeats (CRISPR) clusters generally increased with the depth. Dioxygenases, which are indicators of the potential for degrading recalcitrant compounds, also are more prevalent at higher depths, as are the chaperones that help the organism to grow at lower temperatures and higher pressure (Xu and Ma, 2007). Other groups of genes showing conspicuous depth-dependent rise are genes involved in heavy metal resistance, α-, β- and γ-subunits of nitrate reductase, and *nar* system for nitrate respiration, indicating that deep-sea microbiota is better suited for the life under microaerophilic conditions in eventual association with particulate detrital material sinking from photic zone (Ivars-Martinez *et al*., 2008; Aristegui *et al*., 2009; Konstantinidis *et al*., 2009; Eloe *et al*., 2011b). Such a particle-associated sessile lifestyle may represent a way to overcome the nutrient poorness of deep waters and to structure microbial communities at higher depths (Azam and Long, 2001; Kiørboe and Jackson, 2001).

In this study, we performed a metagenome pyrosequencing of environmental DNA extracted from hadopelagic prokaryotic community collected at 4908 m depth in Matapan-Vavilov Deep (MVD). In addition to metagenomic analyses, the total prokaryotic community composition was examined by conventional 16S rRNA gene-based clone library construction and sequencing. The MVD canyon with the maximum depth of 5267 m bsl located at the seabed of Ionian Sea approximately 56 km SW of city Pylos (Greece). This is the deepest location of the Mediterranean Sea, representing a unique warm hadopelagic environment on our planet where, despite of the depth, the water temperature is never below 13.5°C. The circulation of different water masses (Modified Atlantic Water; Levantine Intermediate Water; Eastern Mediterranean Deep Water) in this area has been proposed (Malanotte-Rizzoli *et al*., 1997). Noticeably, due to the proximity to the coastal zone, this deep station is relatively rich in POM and dissolved organic carbon, likely derived from both surface waters and continental shelves in contrast with adjacent areas (Seritti *et al*., 2003; Santinelli *et al*., 2010). The biological studies performed at this site were scarce and very little information was available about the microbial inhabitants of this environment.

Results and discussion

Environmental settings of the MVD and rates of total prokaryotic production through aphotic column

In June 2010, during the oceanographic cruise MAMBA-10, the water column at Station Matapan-Vavilov Deep was sampled from the surface to the depth of 4908 m, from where 150 l of hadolpelagic seawater were collected for metagenomic analyses. The MVD system (maximum depth 5267 m) is located *c*. 56 km SW of Pylos (Greece) at the seabed of Ionian Sea in the long narrow depression called Hellenic Trench.

At the time of sampling the temperature of the water column ranged from 22.34°C to 13.80°C and was 14.29°C at the depth of 4908 m; the mean salinity at 4908 m was 38.73 PSU. Nitrate and phosphate concentrations exhibited the common depth-increasing trend reaching at the depth of 4908 m the concentration of 4.80 ± 0.48 and 0.13 ± 0.01 μmol l^{-1} respectively. Oxygen concentration profile ranged between 5.74 and 4.26 ml l^{-1}. The oxygen minimum zone (OMZ) was found at the depth between 1150 and 1160 m (Fig. 1A). The measurements of dark primary production (DPP) and prokaryotic heterotrophic productions (PHP) were taken at six depth horizons, corresponding to the mesopelagic (200 m, 500 m and 1156 m, the last depth corresponds to the OMZ), bathypelagic (2500 m, 3500 m) and abyssopelagic (4908 m) water masses (Fig. 1B). The rate of incorporation of ^{3}H-leucine in the photic zone (25 m) was also determined. Mean PHP declined rapidly from this depth to 200 m layer from 399.0 ± 14.6 μgC m^{-3} day^{-1} to 41.5 ± 4.5 μgC m^{-3} day^{-1}, remaining further relatively constant in the mesopelagic waters (30–40 μgC m^{-3} day^{-1}), until reaching the OMZ where the maximum of the activity was detected

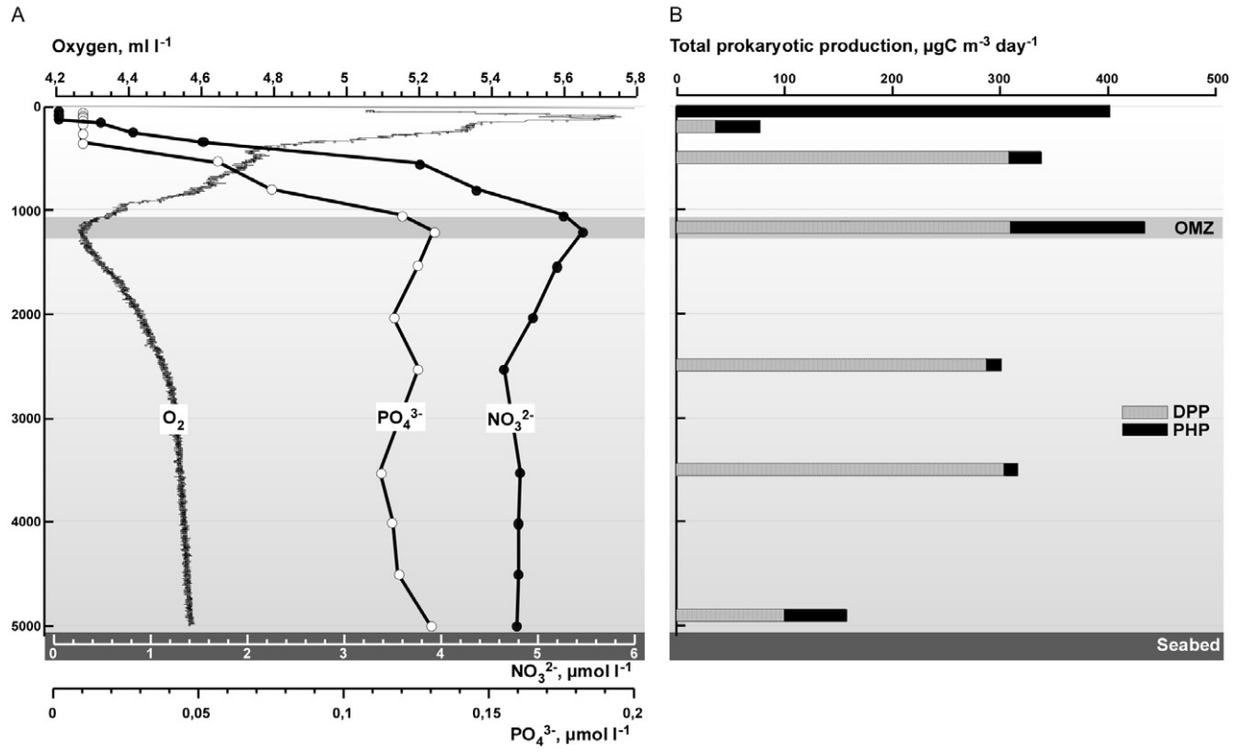

Fig. 1. Depth profiles of estimated total production rates (A) and concentration of oxygen and principal nutrients (B) along the water column at MVD. The oxygen profiling is shown by solid line and indicated as 'O_2'; nitrate and phosphate profiles are indicated by solid lines and closed and opened circles respectively.

(124.1 ± 13.5 μgC m^{-3} day^{-1}). In agreement with the extreme oligotrophic conditions in this area of Mediterranean Sea, the very low rates of heterotrophic production (12–14 μgC m^{-3} day^{-1}) were observed in bathypelagic waters below OMZ. Existence of a second peak of deep-sea PHP (57.8 ± 5.1 μgC m^{-3} day^{-1}) measured in hadopelagic waters at a depth of 4908 m could be explained by the proximity of highly productive superficial sediments (533.8 ± 5.8 μgC m^{-3} day^{-1}) and related organic carbon availability (Santinelli *et al.*, 2010). In contrast to PHP profile, the DPP in meso- and bathypelagic water did not follow negative depth-increasing trend and was found to be quite constant (mean value 303 ± 34 μgC m^{-3} day^{-1}) in samples collected from 500 m to 3500 m. These DPP estimates corroborate our previous measurements (290–370 μgC m^{-3} day^{-1}; Yakimov *et al.*, 2007), which seems to be a characteristic feature of deep water masses in the Ionian Sea. Moving downwards, the hadopelagic waters exhibited the threefold recession of the DPP values.

Total prokaryotic cell densities at this depth were 3.75 ± 0.11 × 10^4 cells ml^{-1} and comparable with total prokaryotic numbers estimated in bathypelagic water at 3500 m depth (4.8 ± 0.6 × 10^4 cells ml^{-1}). As revealed by qPCR analysis, the members of bacterial and archaeal domains of life were present at this depth at similar levels (53.8% vs. 46.2% respectively), whereas hadopelagic water was characterized by strong dominance of *Eubacteria* (96.2%). Taking into account that members of *Thaumarchaeota* Marine Group I are likely to play a pivotal role in deep-sea autotrophic activity, such a decay in archaeal abundance could explain the recession of the DPP values observed at 4908 m. However, the minor thaumarchaeal fraction still may be insufficient to support the bicarbonate fixation rates measured there. Recent studies suggested that anaplerotic pathways and chemolithoautotrophy in several *Proteobacteria* lineages that are ubiquitous in the dark oxygenated ocean can contribute to bicarbonate fixation, potentially closing the productivity gap (Alonso-Sáez *et al.*, 2010; Swan *et al.*, 2011).

Microbial diversity in MVD metagenome

To uncover the genomic composition of the MVD microbial community, the extracted environmental DNA was directly pyrosequenced and sequencing reads assembled (for details see Table 1 and Figs S1 and S2). The run yielded 1 394 500 raw sequences, reduced to 1 118 865 after trimming, with a read length of 446.7 ± 111.7 (maximum 756) nucleotides and GC content of 44.8%. We assembled sequences of this data set using the GS

Table 1. General statistics of metagenomic datasets selected for comparative study.

Metagenomic data	Mediterranean sea water (50 m) [4454267]	ALOHA euphotic zone water (75 m) [4453355]	Marmara deep-sea water (1 000 m) [4453141]	Marmara deep-sea sediment (1 300 m) [4453143]	Mediterranean sea water (3 010 m) [4454107]	ALOHA deep-sea water (4 000 m) [4454108]	Matapan deep-sea water (4 908 m) [4451947]	Puerto Rico Trench water (6000 m) [4486723]
Sequencing method	Pyrosequencing	Pyrosequencing	Pyrosequencing	Pyrosequencing	Sanger, fosmid ends	Sanger, fosmid + 454	Pyrosequencing	Pyrosequencing
Original average size (bp)	640.8	113.4	295.4	289.1	799.2	1 345.7	438.7	541.3
Original no. of fragments	1 204 425	414 323	257 798	287 785	9 005	65 738	1 394 500	620 600
Average size after trimming (bp)	179.3	108.6	178.2	175.3	200	200	446.7	261.8
No. of fragments	741 007	187 259	236 054	268 005	43 023	474 545	1 118 865	530 155
Total nucleotides (Mbp)	132.82	20.32	42.02	46.94	7.20	88.42	499.80	138.82
GC content (%)	40.70%	34.70%	45.20%	56.20%	49.90%	52.70%	44.80%	53.10%
% COG (bit score > 40)	22.43%	8.88%	19.37%	15.08%	19.63%	23.07%	50.81%	27.44%
% KEGG (*E*-value < 1e-05)	25.34%	25.52%	27.35%	15.81%	21.42%	25.72%	16.82%	27.89%
% SEED (*E*-value < 1e-05)	35.57%	35.53%	42.65%	25.12%	31.26%	39.42%	35.52%	44.65%
Effective genome size (Mbp)	2.845	7.897	3.884	7.696	4.869	3.848	4.498	3.664

COG, clusters of orthologous groups; KEGG, Kyoto Encyclopedia of Genes and Genomes.

De Novo Assembler of the Genome Sequencer FLX data analysis suite (version 2.3) with the default parameters. Almost all reads (95.33%), were recruited to form 7182 contigs with an average length of 1812.3 bp, in total 13 016 Mb of sequences, with a GC content of 44.5%. More than a half of contigs (4539 total) were large contigs with average sequence length of 2684.0 bp and maximum of 69 668 bp. The ability of the algorithm to assemble > 95% of the sequence reads indicates the low diversity of the Matapan microbial community, which was further confirmed by the analysis of recovered 16S rRNA gene sequences.

The small subunit (SSU) 16S rRNA gene signatures were profiled using the MG-RAST online server, with RDP and SILVA databases as the comparison spaces (Pruesse *et al*., 2007; Meyer *et al*., 2008; Cole *et al*., 2009). Respectively, 1434 and 1559 reads found their matches in these databases with 627 and 762 unambiguously assigned. Almost all of the 16S RNA fragments were placed within domain of *Eubacteria* with absolute dominance of sequences from *Proteobacteria*. Phylogenetic diversity profiling of 16S rRNA genes found in Matapan metagenome was similar to that obtained via conventional SSU clone library approach (Fig. 2). Five orders (*Alteromonadales*, *Enterobacteriales, Oceanospirillales*, *Pseudomonadales*, *Vibrionales*) accounted for the all Matapan gammaproteobacteria as suggested by both RDP and SILVA. At the lower taxonomic level, the greatest numbers of sequences (530 for RDP, 539 for SILVA) were classified within the clade of bathypelagic *Alteromonas macleodii* 'deep ecotype' (AltDE). This common opportunistic marine bacterium was originally isolated from the depth of 1000 m in the Eastern Mediterranean (Ivars-Martinez *et al*., 2008). Remarkably, this organism seems be ubiquitous in deep-sea compartments of Mediterranean basin; e.g. its genomic material was also abundantly recovered (up to 85% genome coverage) in the recent metagenomic study of bathypelagic plankton and bottom sediments from the Sea of Marmara (Quaiser *et al*., 2011). Relatively few sequences (less than 1% of all SSU-affiliated reads) from other proteobacterial classes (*Alpha*- and *Betaproteobacteria*) were recovered in MVD metagenome while only members of *Alphaproteobacteria* were detected in bacterial clone library. Phylum *Planctomycetes* was present by a singleton in 16S rRNA gene clone library (Fig. 2).

None of the reads related to any archaeal rRNA was detected in Matapan metagenome, in contrast with the qPCR analysis of the same environmental sample, where 3.8% or amplicons were from archaea. Cloning and sequencing of these amplicons indicate that they were almost exclusively from Group I Marine *Thaumarchaeota* (36 of 42 archaeal clones analysed), which is characteristic for pelagic archaeal communities inhabiting aphotic

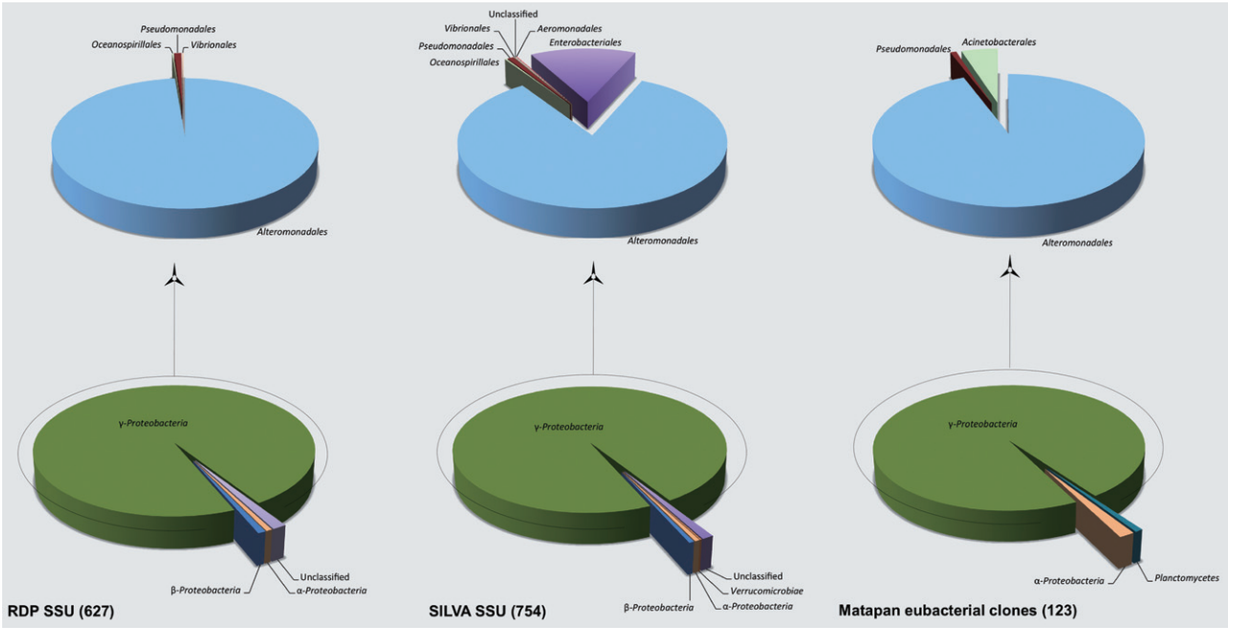

Fig. 2. Taxonomic classification of superkingdom *Bacteria* in Matapan metagenome reads using recovered SSU rDNA sequences (RDP SSU, SILVA SSU database and conventional SSU clone library). Relative proportions of different eubacterial and gammaproteobacterial taxa are shown separately in bottom and upper set of pies respectively. The taxonomic assignment was carried out using a cut-off expectation (*E*) value of 10^{-5} and alignment length of 30 bp inside the MG-RAST pipeline.

marine ecosystems (La Cono *et al*., 2009; 2010; Martin-Cuadrado *et al*., 2009; Eloe *et al*., 2011b; Yakimov *et al*., 2011). The underrepresentation of archaeal SSU-related sequences in Matapan metagenome data indicates the discrepancy between PCR-based approaches and direct pyrosequencing of environmental DNA as far as rare phylotypes are concerned, but does not affect the conclusion of the dominant role of AltDE in this habitat.

To take another view on species richness in Matapan metagenome, we compared the community composition deduced from analysis of 16S rRNA genes with the taxonomic binning of protein-encoding genes against the SEED database using MG-RAST pipeline. Genomic signatures of 928 different evolutionary neighbours were detected with this approach, whereas only 592 and 602 different species were detected using SILVA and RDP SSU rRNA genes database respectively (Table S1). The proportions of the major bacterial groups identified by taxonomic binning of protein-encoding genes were comparable to those identified through analyses of 16S rRNA genes (Fig. 3). Almost 84% of 1 131 724 Matapan genes matching with SEED taxonomy database were classified as *Gammaproteobacteria*-related genes and approximately 24.85% clustered with the genes of *Alteromonas, Pseudoalteromonas* (14.8%), *Shewanella* (12.48%), *Idiomarina* (3.17%), *Colwellia* (2.59%) and *Psychromonas* (0.2%). The order *Vibrionales* followed this category with 7.9% of all matches (Table S1). In general, the SEED analysis was characterized by much deeper resolution and by recovering of genetic signatures belonging to all kingdoms of life (Table S1). In SEED database several opportunistic bacterial pathogens were detected, along with a conspicuous number of sequences belonged to terrestrial plants (Table S1), which are likely to originate from a continental shelf of the Peloponnese peninsula located 30 nautical miles away. Although no archaeal rRNA genes were detected in MVD metagenome by pyrosequencing, 1956 protein-coding fragments (0.17% of all reads) were attributed to the members of this kingdom. A relatively low number of reads (363) were of viral origin, mainly related to *Enterobacteria* phage P2 of the *Myoviridae* family.

Comparison of hadopelagic Matapan diversity to other marine metagenomic datasets

Matapan Deep hadopelagic metagenome was compared with a set of publicly available marine metagenomic data that were processed to minimize biases related to differences in sequence length (Table 1, see *Experimental procedures*). Seven marine metagenomes were selected for comparison: ALOHA (Hawaii station in North Pacific Subtropical Gyre, depth of 4000 m) (DeLong *et al*., 2006; Konstantinidis *et al*., 2009); Puerto Rico Trench (Central Atlantic ocean depth of 6000 m) (Eloe *et al*., 2011a);

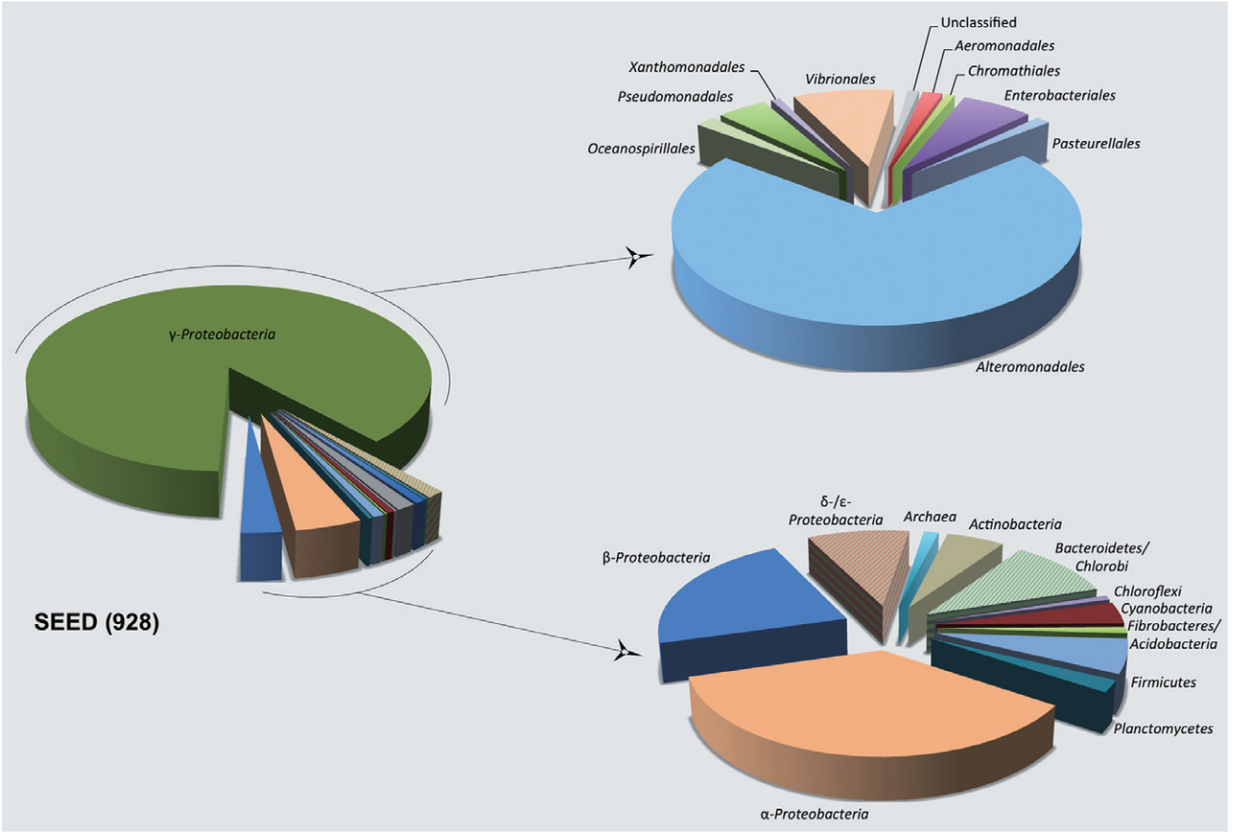

Fig. 3. Appearance in MVD metagenome of different archaeal and bacterial taxa based on the taxonomic binning of protein-encoding genes against the SEED database using MG-RAST pipeline. Relative proportions of gammaproteobacterial and non-gammaproteobacterial taxa are shown separately in bottom right and top right pies respectively. The taxonomic assignment SEED database was carried out using a cut-off expectation (*E*) value of 10^{-5} and alignment length of 30 bp inside the MG-RAST pipeline.

KM3 (Eastern Mediterranean KM3 station, depth of 3010 m) (Martin-Cuadrado *et al.*, 2007); two metagenomes obtained from 1000 m (pelagic) and 1300 m (sediment) samples of Marmara Sea (Mediterranean Sea) (Quaiser *et al.*, 2011). Two additional datasets derived from surface samples, one from Atlantic Ocean (ALOHA Station, depth of 75 m) (Frias-Lopez *et al.*, 2008) and another one from Mediterranean Sea (depth of 50 m) (Ghai *et al.*, 2010) were also considered. Eubacteria-specific protein sequences were the most abundant domains in all these metagenomes (Fig. S3A), whereas the *Archaea* fraction never exceeded 12% (Fig. S3B). Different situation was observed in superficial waters and sediments, where archaeal community was much higher, in that case by euryarchaeal organisms (Martin-Cuadrado *et al.*, 2008; Barberán *et al.*, 2011; Quaiser *et al.*, 2011). Eukaryotic DNA fragments, as well as virus DNAs, were scarce in all habitats whereas viruses were slightly more abundant in superficial seawater metagenomes (Fig. S3D) (ALOHA and Mediterranean surface water metagenomes respectively).

Predicted protein sequences of the selected metagenomes were compared with those from different databases of conserved domains and annotated molecular functions, and categorized according to their predicted functions (Tatusov *et al.*, 2003; Kanehisa *et al.*, 2004; Overbeek *et al.*, 2005). As it can be seen in Figs 4 and S4, and Tables S2A and S3, the MVD community is genetically distinct from other deep-sea metagenomes. The only outlier beside the MVD community was the surface plankton community of ALOHA Station (Pacific Ocean) which, like MVD, was characterized by extremely low biodiversity, although ALOHA was dominated by cyanobacterium *Prochlorococcus marinus*.

The more abundantly represented functional categories in MVD as where as in other deep-sea planktonic communities correspond to environmental sensing and cell-to-cell communication, namely virulence, regulation and cell signalling, motility and chemotaxis (Figs 4 and S4). Deep-sea microorganisms typically have sophisticated systems of environmental sensing and signal transduction (Galperin, 2005), and deep-sea planktonic life is most

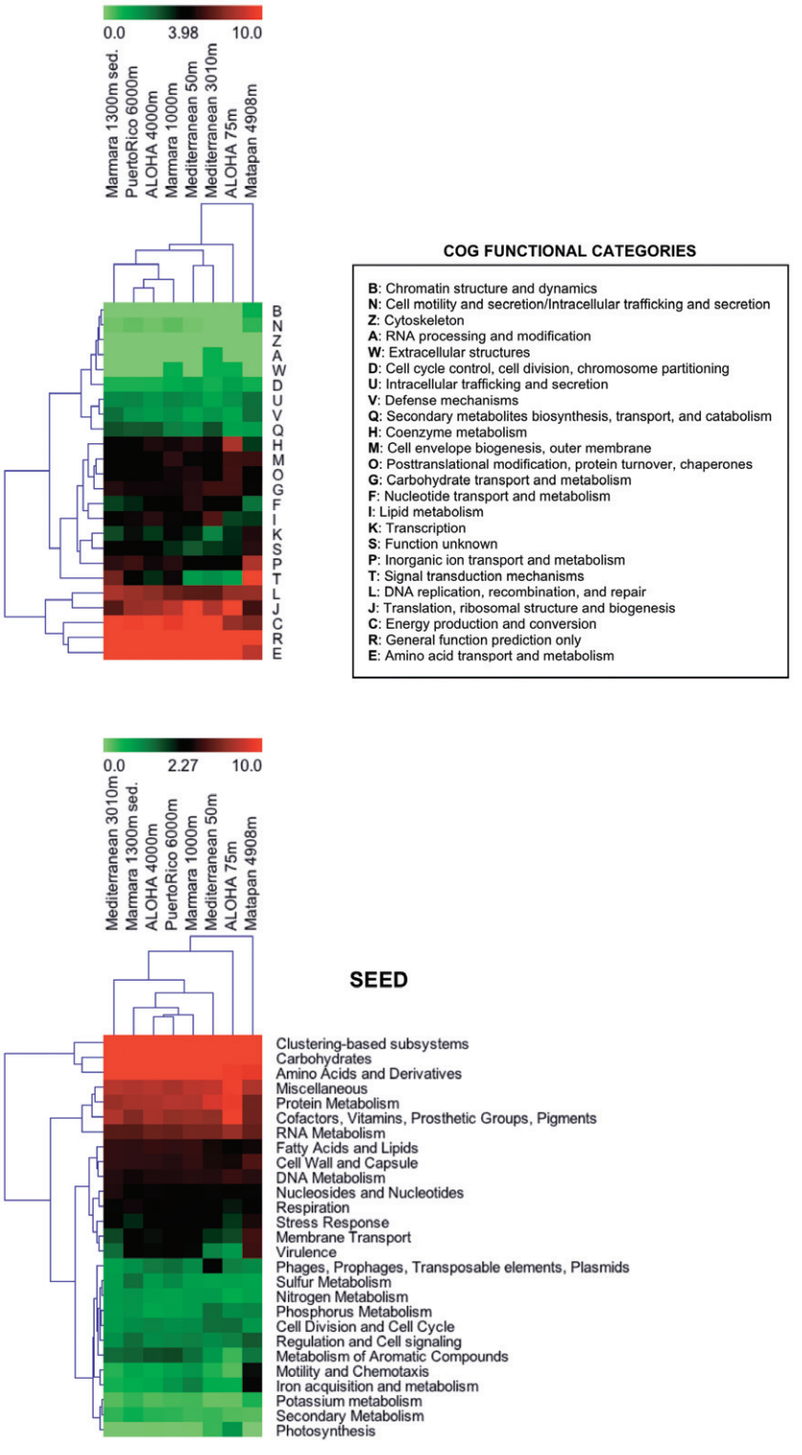

Fig. 4. Cluster analysis of several metagenomes based on matches to different COG and SEED categories expressed as percentage to the respective number of proteins identified in the metagenomes. A detailed description of all annotated sequences is reported in Tables S2 and S3.

likely to associate with particles of sinking organic matter, requiring genetic adaptations similar to those of sediment-inhabiting microbes (Martin-Cuadrado *et al*., 2007; Aristegui *et al*., 2009). As proposed more than a decade ago and confirmed by recent studies, this sessile lifestyle may be an attractive way to overcome the organic nutrient poorness which typically characterizes deep waters (Azam and Long, 2001; Kiørboe and Jackson, 2001;

Arnosti, 2002; Martin-Cuadrado *et al*., 2007; 2008; Lauro and Bartlett, 2008).

Key enzymes and pathways

We next looked for the proteins diagnostic of particular enzymatic pathways and metabolic capabilities occurred in MVD (Table 2). A partial metabolic reconstruction revealed substantial differences between characteristic metabolic features in the deep versus surface ecosystems. As has been shown before, transposases, phage integrases, plasmids, recombinases and other viral (prophage)-related hypothetical proteins [21 clusters of orthologous groups (COGs) in total, Table S2B] are ~ 10-fold more abundant in the deep sea than in the surface water samples.

A diverse set of enzymes involved in degradation of biopolymers (alpha-/beta-glucosidases, leucine aminopeptidases, arylsulfatases, alkaline phosphatases) and in uptake of phosphorus (ABC-type phosphate/phosphonate transport system) was represented by a noticeable number of matches in deep-sea metagenomes (Table 2). Many of above enzymes are required to initiate the re-mineralization of various high molecular weight organic compounds like glycoproteins, hetero- and lipopolysaccharides that are principal constituents of phytoplankton cell wall and prominent components of POM (Arnosti, 2002). Another large group of deduced protein sequences found in Matapan metagenome (711 hits) belong to subsystems and categories involved in heavy metal resistance and detoxification, such as Co/Zn/Cd resistance (efflux system components). The elevated presence of mercury resistance genes in the genome of bathypelagic isolate AltDE compared with shallow ecotype ATCC 27126 (Ivars-Martinez *et al*., 2008) is thought to help the bacterium withstand high concentration of heavy metals and trace elements adsorbed on the surface of POM.

Extremely low concentration of dissolved iron (< 10 nmol l^{-1}) is a typical signature for the photic layer of low-chlorophyll zones of open oceanic and Eastern Mediterranean Sea waters (Saager *et al*., 1993; Boyd *et al*., 2007). Remarkably, the concentration of organic ligands in these areas usually was threefold to fourfold higher, such that > 99% of dissolved iron was calculated as organically complexed (de Baar and de Jong, 2001). Produced by numerous marine bacteria to facilitate acquisition of iron, siderophores are biosynthesized in response to low iron levels (Martinez and Butler, 2007). Comparison of surface and deep-sea metagenomes shows more Fe^{3+}-siderophores transport/acquisition systems (COG1120) at higher depths (Table S2A).

Besides the virtual lack of dissolved iron, the deep Mediterranean Sea is permanently nutrient-deprived and is one of the most oligotrophic marine ecosystems with ammonium concentrations significantly below 10 nmol l^{-1} (Woodward, 1994). However, bathypelagic microbial community seems to be adapted to thrive under such extreme nutrient limitation via possessing the remarkably high affinity and sensitivity towards NH_4^+ and the metabolic capacity to use as electron donors the reduced nitrogen species, other than NH_4^+, such as urea and amino acids (Reinthaler *et al*., 2005; Yakimov *et al*., 2011). In fact, for the Matapan metagenome, the genes encoding ammonium transporters and urease subunits were frequent even when the AltDE-related gene products were excluded from consideration.

The most common alternatives to oxygen respiration found in marine pelagic prokaryotes are reduction of nitrate and sulfate (Swan *et al*., 2011). Recently it was proposed that such respiratory versatility could increase survival odds in microaerophilic/anaerobic interphase found within larger organic aggregates (Ivars-Martinez *et al*., 2008). Different subunits of nitrate reductase (*nar* system) were found in similar proportions in Matapan metagenome.

Heterotrophic bacteria can contribute to assimilation of CO_2 via carboxylation reactions performed by such enzymes as pyruvate carboxylase and phosphoenolpyruvate carboxylase (Romanenko, 1964). Several reports suggest that these anaplerotic reactions are responsible for about 1–8% of the total heterotrophic bacterioplankton biomass production in some marine ecosystems (Alonso-Sáez *et al*., 2010; Reinthaler *et al*., 2010). High proportion of pyruvate and phosphoenolpyruvate carboxylases (PEPC) (771 and 423 matches respectively) was detected in MVD hadopelagic plankton, indicating the eventual contribution of anaplerotic reactions to the total assimilation of bicarbonate at this depth.

Besides chemlithoautotrophy and anaplerotic heterotrophy, lithoheterotrophy appears to take place in MVD hadopelagic ecosystem. Similarly to ALOHA 4000 m, Puerto Rico and KM3 metagenomes, a considerable number of *cox* genes encoding different subunits of the carbon monoxide dehydrogenase (CoxL/CoxM/CoxS) was detected in MVD metagenome. Phylogenetically, more than 90% of them have deep ecotype of *A. macleodii* (AltDE) or *Alphaproteobacteria*, *Actinobacteria* and *Chloroflexi* as nearest neighbours. Initially thought to be exclusive of carboxydotrophic autotrophs, CO oxidation is being discovered in a plethora of marine bacteria and is likely involved in the lithotrophy as an alternative or supplementary energy source of heterotrophy (Martin-Cuadrado *et al*., 2009). Such energy metabolism versatility would be advantageous in highly depleted deep-sea environment much in the way that phototrophy helps heterotrophy at the surface (Martin-Cuadrado *et al*., 2007; DeLong and Béjà, 2010; Quaiser *et al*., 2011).

Table 2. Comparison of the gene abundances for key metabolic enzymes in MVD hadopelagic water masses with those from other metagenomes.

Functional key enzymes (bit score ≥ 40)	Mediterranean sea plankton (50 m)	ALOHA euphotic zone plankton (75 m)	Marmara deep-sea plankton (1 000 m)	Marmara deep-sea sediment (1 300 m)	Mediterranean sea plankton (3 010 m)	ALOHA deep-sea plankton (4 000 m)	Matapan deep-sea water (4 908 m)	Puerto Rico Trench water (6 000 m)
Archaeal ammonia monooxygenase sub. A, AmoA (pfam12942)	0	0	6	1	0	2	2	16
Leucyl aminopeptidase, PepB(COG0260)	258	24	51	0	2	77	33 (*655*)	152
ABC-type cobalamin/Fe3 ± siderophores, FepC (COG1120)	150	21	38	36	7	105	140 (*274*)	81
ABC-type phosphate/phosphonate, PhnD (COG3221)	26	0	4	5	3	23	4 (*62*)	43
Acetyl-CoA carboxylase (COG0825/COG0777/COG4799)	732	70	165	170	45	502	60 (*569*)	465
Alkaline phosphatase (COG1785)	28	0	9	5	1	12	286 (*280*)	18
Alpha-/beta-glucosidase (COG1501/COG1472)	57	0	27	24	1	61	250 (*717*)	94
Ammonia permease, AmtB (COG0004)	313	10	64	49	11	109	15 (*241*)	194
Arylsulfatase A, AslA (COG3119)	559	5	109	374	31	577	85 (*75*)	800
Carbon mon. CoxL/CoxM/CoxS (COG1529/COG1319/COG2080)	379	12	133	236	89	789	87 (*788*)	1 015
Chaperonin GroEL (COG0459)	564	27	82	69	15	344	20 (*406*)	396
Co/Zn/Cd efflux system component, CzcD (COG1230)	7	0	8	13	5	40	216 (*495*)	78
Ferredoxin subunits of nitrite reductase, NirD (COG2146)	12	0	12	5	3	25	3 (*52*)	67
Hydrogenase HypC/HypD/HypE (COG0298/COG0409/COG0309)	1	1	10	38	0	11	1 (*290*)	13
Nitrate reductase (COG1140/COG2180/COG2181)	7	0	13	31	1	12	22 (*209*)	8
Phosphoenolpyruvate carboxylase (COG2352)	178	30	41	13	3	46	6 (*765*)	56
Pyruvate/oxaloacetate carboxyltransferase (COG5016)	53	2	32	14	3	34	7 (*416*)	38
Transposase (COG5421/COG5433/COG5659)	2	0	5	10	3	4	63 (*296*)	20
Urease (COG0804/COG0832/COG0831)	168	99	51	2	0	85	71 (*516*)	139
Total hits	**3 494**	**301**	**860**	**1 095**	**223**	**2 858**	**1 371 (*7 106*)**	**3 693**
No. of fragments	**741 007**	**187 259**	**236 054**	**268 005**	**43 023**	**474 545**	**1 118 865**	**530 155**

In column eight, corresponding to MVD metagenome the numbers of reads matching to *A. macleodii* AltDE are shown separately in parentheses. Thus, the total hits in MVD metagenome means the sum of values indicated both out and within the parentheses.

CRISPR/Cas systems and viral sequences in Matapan metagenome

Prokaryotes thrive in spite of the vast number and diversity of their viruses, which partly results from the evolution of mechanisms to inactivate or silence the action of exogenous DNA. Among them, CRISPR and related Cas proteins have been revealed as a unique defence system in prokaryotes which recognizes fragment of nucleic acid of foreign origin, like phages and plasmids, degrades them using the Cas protein system (Deveau *et al*., 2010), and updates the defence system via the integration of new CRISPR spacers and passes on the modified set of CRISPR spacers to offspring generations (Barrangou *et al*., 2007).

In MVD metagenome we detected three different CRISPR systems (with associated Cas proteins), named A, B and C (Fig. 5). CRISPR A and B were very similar one to another and to a system from *A. macleodii* AltDE 'deep ecotype' (Fig. 5A) (Ivars-Martinez *et al*., 2008; Quaiser *et al*., 2011). Following the current classification, all these systems were affiliated to I-E subtype, widely spread in *Proteobacteria* (Makarova *et al*., 2011). CRISPR A contains a complete assembly of genes, namely those coding for Cas3, Cse1, Cse2, Cse4, Cas5, Cse3, phage-related protein of KilA-N superfamily, Cas1, Cas2 and the CRISPR cassette. CRISPR A and AltDE systems have not only identical stretches of interspaced repetitive elements but also virtually identical spacer content. Based on the sequence similarity, 33 potential proto-spacers are likely to be of the phage origin, including three spacers perfectly matching the known viruses (Table S4B). The CRISPR B system has the same Cas proteins except Cse2 and the same repeat region previously found only in AltDE and *Psychromonas ingrahamii* 37, but its spacer sequences do not show any similarity to database sequences. The third, partially recovered, CRISPR C system belongs to the I-F Group (Makarova *et al*., 2011). This 9182 bp long fragment contains Cas-related genes similar to those of *Shewanella baltica* (Fig. 5B). As in case with CRISPR B, CRISPR C spacers did not match any proto-spacer hits in GenBank.

We compared Matapan CRISPR with similar systems present in other marine metagenomes. Eleven frag-

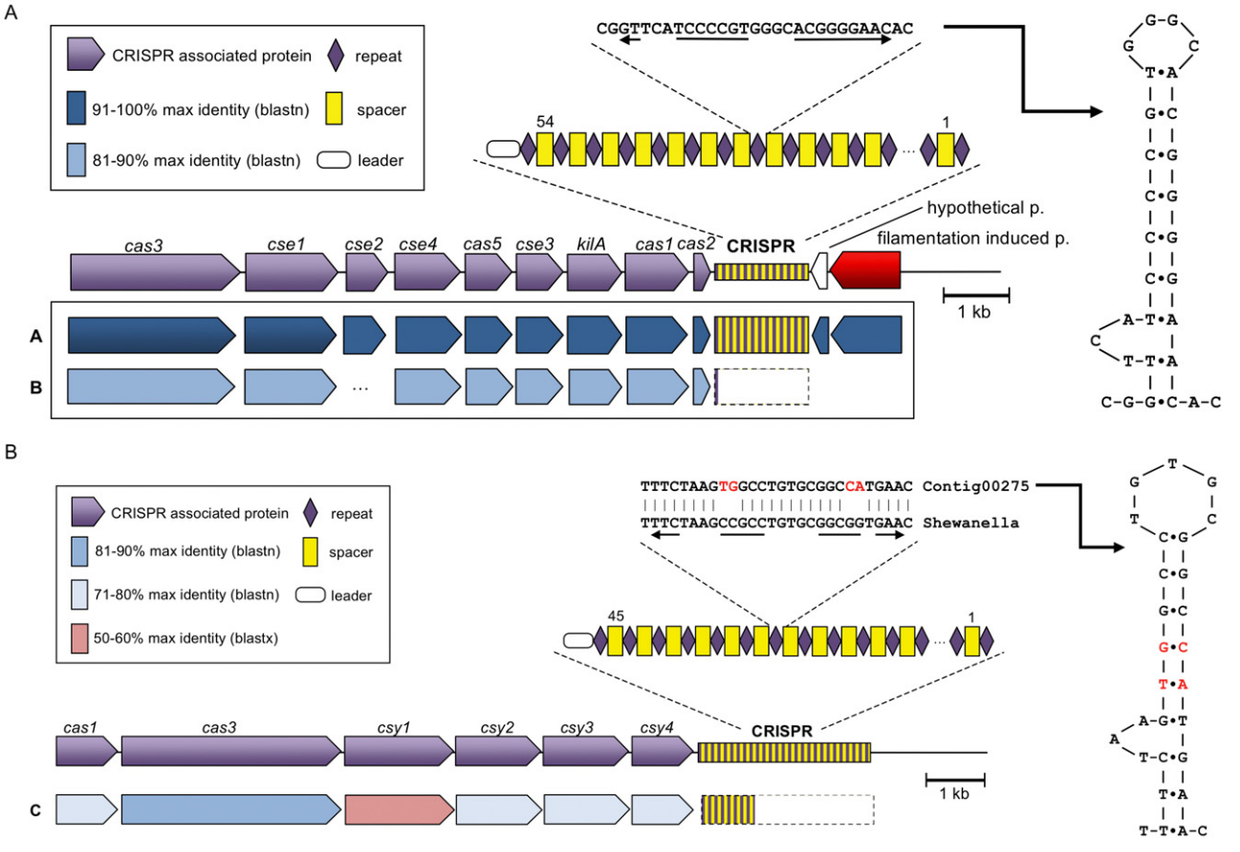

Fig. 5. Structures of CRISPR/Cas A, B (A) and C (B) systems identified in Matapan metagenome, with description of associated protein and repeat region found. See text and *Experimental procedures* for further details.

ments of CRISPR systems, that consisted of one Cas-related gene and/or some stretches of interspaced repetitive DNA and spacers, were detected in three deep-sea planktonic ecosystems, Marmara 1000 m (two CRISPR elements), ALOHA 4000 m (eight CRISPR elements) and Puerto Rico Trench 6000 m (one CRISPR element). Cas-related proteins were recovered only in two CRISPR elements found in ALOHA 4000 m metagenome, with percentage of similarity ranging from 51% to 72% for Csy4 of *Thioalkalivibrio* sp. K90mix, a sulphur-oxidizing gammaproteobacterium (Sorokin *et al*., 2008). Four fragments of CRISPR cassettes (one in Marmara and three in ALOHA metagenomes) had different repeat sequences, previously found in some members of *Proteobacteria* and *Firmicutes* (Table S4A). Remarkably, among these unaffiliated repeat sequences, the one from Marmara had only a single mismatch with those of Matapan CRISPR C cassette.

Concluding remarks

Comparative analysis of metagenomic data derived from one of the deepest habitat studied thus far confirmed the molecular signatures of the particle-associated life style, but also suggested features specific for hadopelagic MVD microbial community. MVD prokaryotic diversity was found be extremely poor, unlike any other deep-sea metagenome. A strong decline in DPP rates measured at 4908 m depth was coincident with the overwhelming dominance of copiotrophic *A. macleodii* 'deep-ecotype' AltDE at the expense of *Thaumarchaea* and other prokaryotes potentially involved in both autotrophic and anaplerotic CO_2 fixation. A marked conservation of CRISPR/Cas elements has been revealed, which does not seem to match the known viral diversity in these habitats.

Experimental procedures

Sampling

During the cruise MAMBA2010 in June 2010 the sampling was carried out from the surface to a maximum depth of 4908 m (sea bottom at 5093 m) at Station Matapan-Vavilov Deep (36°34.00N, 21°07.44E). Additionally, 150 l of hadopelagic water were collected by using a 12 l Niskin bottles, mounted on a General Oceanics conductivity temperature depth rosette sampler. Water sample were filtered through sterile Sterivex capsules (0.2 mm pore size, Millipore) using a peristaltic pump. Filters were than stored at −20°C until processing. Dissolved oxygen concentrations were determined with a SBE oxygen sensor mounted on the conductivity-temperature-depth and nutrient concentrations (i.e. phosphate and nitrate) with a nutrient autoanalyser (Bran & Luebbe Autoanalyzer II, Norderstedt, Germany).

PHP and dark ocean primary production

Net-PHP was estimated measuring the protein synthesis rate using 3H-leucine uptake (Kirchman *et al*., 1985) by the micro-method of Smith and Azam (1992). Triplicate subsamples and duplicate blanks were incubated with 20 nM of leucine (5 nM L-[4,5-^{3}H]leucine, SA = 165.2 Ci mmol^{-1} + 15 nM L-leucine), in the dark, during 90 min (≤ 100 depth samples) and 150 min (> 100 depth samples) at '*in situ*' ± 1.5°C temperatures. Heterotrophic prokaryotic carbon biomass production was calculated according to Kirchman (1993) using *in situ* determinations of leucine isotopic dilution calculated according to Pollard and Moriarty (1984). The variability between the three subsamples was determined as CV%. For the calculation of the total leucine uptake the 3H-leucine uptake and the leucine isotopic dilution were considered.

The dark ocean primary production was estimated by the incorporation of [^{14}C]bicarbonate (10 μCi ml^{-1}, Amersham), according to the protocol of Herndl and colleagues (2005). Forty millilitres for each sample in triplicate and one formaldehyde-fixed blank control were incubated in the dark for 7 days at *in situ* temperature (13°C). The incubation was terminated by the addition of formaldehyde to a final concentration of 2% (v/v), and samples were filtered through 0.1 μm polycarbonate filters (Millipore). Three washes with 10 ml of ultra-filtered (0.1 μm) seawater and vapour HCl (12 h of exposure) to eliminate unincorporated [^{14}C]bicarbonate were carried out. Filters were then placed in scintillation vials and stored in the dark at −20°C until counted. The incorporated radioactivity was measured as disintegrations per minute counts with a liquid scintillation counter (Wallac WinSpectral 1414 Liquid Scintillation Counter, PerkinElmer Life Sciences) using the internal radioisotopes library and quenching correction. The values obtained in disintegrations per minute were normalized against the values of the abiotic control and corrected for the natural DIC concentration (2 mM).

DNA extraction, PCR amplification, cloning and sequencing

In order to recover more of microbial diversity, different commercial kits for the purification of the DNA, expected to have different biases, were used. Total DNA was extracted using Qiagen RNA/DNA Mini Kit (Qiagen, Milan, Italy) and Meta-G-Nome DNA Isolation Kit (Epicentre, USA). The extraction was carried out according to the manufacturer's instructions. DNA samples were pooled together and examined by agarose gel electrophoresis. DNA concentrations were determined using the NanoDrop ND-1000 Spectrophotometer (Wilmington, DE, USA). 10 μg of DNA were used for direct 454 random whole-genome shotgun analysis. The library was processed through the breaking and enriching steps, followed by sequencing of two half plate on the Genome Sequencer FLX Titanium System (454 Life Sciences) at Macrogen (Seoul, Korea), resulting in 1 394 500 reads (average read length 438.7 bp). Archaeal and bacterial SSU rRNA genes were amplified with primers specific for each domain of life: A20F (5′-TTC CGG TTG ATC CYG CCR G-3′) and 958R (5′-YCC GGC GTT GAM TCC AAT T-3′) (Stackebrandt and Goodfellow, 1991; DeLong, 1992)

for 16S rRNA in archaea; and 27F (5′-AGA GTT TGA TCC TGG CTC AG-3′) and 1492R (5′-TAC GYT ACC TTG TTA CGA CTT-3′) (Lane, 1991) for Bacterial 16S rRNA. The PCR reaction and cloning was carried out as described elsewhere (Yakimov *et al*., 2007; 2011). Positive clones from each library were randomly selected by PCR amplification. The PCR products were further purified and sequenced at Macrogen (Seoul, Korea).

Real-time PCR experiments for quantification of total prokaryotes and *Archaea* were performed in triplicate by TaqMan® assays according to methods described elsewhere (Takai and Horikoshi, 2000). Average values of 3.6 to 4.1 16S rRNA gene copies per cell, obtained from a bacterial genome database, are often used to estimate cell numbers in natural environments, whereas estimates for *Archaea* are typically 1 to 1.5 16S rRNA gene copies per cell (Klappenbach *et al*., 2001). These values were used to estimate the MVD microbial group abundances.

For the evaluation of the total prokaryotic abundance, 20 ml of subsamples were directly filtered under vacuum (< 5 mmHg) on 25 mm diameter, 0.22 μm pore size polycarbonate black membranes (Isopore) and stained with 4,6-diamidino-2-phenylindole dihydrochloride (DAPI; Sigma) as reported by Porter and Feig (1980). Cells were visualized using an epifluorescence microscope (Axioplan; Zeiss). Prokaryotic abundance was expressed as cells per millilitre.

Sequences were deposited in GenBank with accession numbers JN563789–JN563807 (archaeal SSU rRNA genes), JN563808–JN563823 (bacterial SSU rRNA genes) and SRA043731 (metagenome sequencing dataset).

Phylogenetic analysis of SSU rRNA genes and rarefaction analyses

Sequences were checked for possible chimeric origin using Bellerophon software (Huber *et al*., 2004). For the 16S rRNA gene sequences, initial alignment of amplified sequences and close relatives identified with BLAST (Altschul *et al*., 1997) were performed using the SILVA alignment tool (Pruesse *et al*., 2007) and manually inserted in ARB (Ludwig *et al*., 2004). After alignment, the neighbour-joining algorithm and Jukes-Cantor distance matrix of ARB program package was used to generate the phylogenetic trees based on distance analysis for 16S rRNA. 1000 bootstrap resamplings were performed to estimate the reproducibility of the partitions in the tree. For statistical analyses clones of each library were separately considered to define phylotypes at 97% of similarity, using DNADIST of the Phylip package (http://evolution.genetics.washington.edu/phylip.html). PAST (PAleontological STatistics v 1.19) was used to perform Rarefaction curves. The generated curves indicated that the saturation was reached for both eubacteria and archaea libraries (data not shown).

Pyrosequencing of Matapan DNA samples and quality trimming of metagenomic sequences for comparative analyses

Seven other metagenomic datasets of marine origin were selected (surface and deep-sea waters) (Table 1) to carry out a comparative analysis between microbial population thriving in oceanic and Mediterranean superficial water, against those from the deep-sea. To minimize potential biases due to different strategies, sequencing technologies and sequence lengths, original data were traded as follows. Except for ALOHA deep-sea plankton (4000 m) and Mediterranean Sea plankton KM3 (3300 m) datasets, all 454 original data from Atlantic superficial water (ALOHA 75 m) (Frias-Lopez *et al*., 2008), Mediterranean superficial water 50 m (Ghai *et al*., 2010), Sea of Marmara (Quaiser *et al*., 2011), and the MVD metagenome were trimmed by SeqTrim software (Falgueras *et al*., 2010) using default values. For ALOHA 4000 m metagenome Sanger fosmid reads (DeLong *et al*., 2006) were joined with the random whole-genome shotgun library (Konstantinidis *et al*., 2009), both collected at 4000 m. Sequences length was then homogenized at 200 bp length using a customized version of perl script that uses bioperl libraries (split_seq.pl, see http://bioperl.org/wiki/Mailing_list). For Mediterranean KM3 metagenome, all fosmid sequences were trimmed to 200 bp long as previously described. After trimming, the MVD metagenome yielded the total number of valid reads of 987 980 with an average size of 471 bp, while the total number of cumulated base pairs in the metagenome dataset was 466 Mbp.

Estimates of effective genome sizes

We applied the *in silico* method developed by Raes and colleagues (2007) to estimate the average genome size of microbial communities using metagenomic sequence data.

Annotation and sequences analysis

Datasets were annotated using the MG-RAST server platform (Meyer *et al*., 2008). For taxonomy analysis, datasets were compared against RDP SSU (Cole *et al*., 2009), SILVA SSU (Pruesse *et al*., 2007) and SEED, using a cut-off expectation (*E*) value of 10^{-5} and alignment length of 30 bp inside the MG-RAST pipeline. Metabolic pathways analysis was conducted using SEED and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al*., 2004) subsystems with MG-RAST applying a cut-off *E*-value of 10^{-5} and alignment length of 30 bp. MG-RAST v3 annotation pipeline in some cases does not provide a single annotation for each submitted fragment of DNA. This is a consequence of genome structure, pipeline engineering, and the character of the sequence databases that MG-RAST uses for annotation. Longer reads (> 400 bp) often contain portions of two microbial genes; when the gene caller makes its prediction, the two polypeptide fragments return annotated separately. For this reason, the results in tables and figures were presented as numbers of organisms or predicted proteins/genes and not as numbers of reads. Metagenomes were also analysed using NCBI blastx (http://blast.ncbi.nlm.nih.gov/Blast.cgi) against the NCBI COG database (Tatusov *et al*., 2003): the BLAST output file was used as input file for MEGAN (Huson *et al*., 2007) applying a cut-off bit score > 40, and normalized to the number of genomic fragments for each metagenome. The hierarchical cluster analysis (Euclidean distance) and percentage of each SEED, COG and KEGG functional category shared in the eight metagenomes was determined and

shown in cross-comparative heat maps using MeV 4.7.3 software (Saeed *et al*., 2003).

Screening of functional key enzymes

We constructed a custom reference database for a number of key enzymes that are diagnostic, or strongly suggestive, for particular metabolic pathways. Many of them were extracted from the COG database (Tatusov *et al*., 2003). Only protein sequences with confirmed activity and/or strong similarity to them were included. All the metagenomes treated in this study were compared against this reference database, and the matches showing a cut-off bit score > 40 were retained. Moreover, sequences identified in Matapan metagenome were further compared against nr database in order to identify the reads matching with AltDE (see Table 2, values in parentheses).

Genome recruitment

Genome fragment recruitment analysis was performed using nucmer and promer command line applications included in MUMmer 3.0 sequence alignment software tools (Kurtz *et al*., 2004). Metagenome reads were aligned to the chosen reference genome using both nucmer and promer alignment tools under standard conditions. The output file was parsed by show-coords, included in MUMmer 3.0, applying options that knockout overlapping alignments from another reading frame (-k), display the sequence lengths (-l), display start and stop of matching region on nucleotide level (default), sort the output files according to the reference file (-r) and indicate identities and similarities of each match (default). Nucleotide regions matching the reference genome were identified, and duplicates corresponding to overlapping fragment matches were eliminated. The remaining matches correspond to unique nucleotide regions that were non-redundantly covered by the metagenome reads based on amino acid alignments.

CRISPR analysis

All CRISPR-positive reads, detected by MG-RAST pipeline, were assembled into contigs and analysed using Geneious v4.8.5 (Drummond *et al*., 2009) (http://www.geneious.com). Contigs were assembled using the 'Medium Sensitivity' method with a word length of 14, a maximum gap size of 2, maximum gaps per read of 15, and maximum mismatches of 15. Every contig was manually inspected and reassembled in larger contigs manually using MEGA5 software (Tamura *et al*., 2011). Pilercr v1.02 (Edgar, 2007) was used to identify short palindromic repeats and spacers in both assembled contigs and metagenome raw reads using default parameters. CRISPR and close relatives Cas protein were identified with BLAST (Altschul *et al*., 1997). Repeat region and spacers found in all CRISPR structures were compared each others using Cap contig assembling program (minimum base overlap 20 bp, per cent match minimum 85%) from BioEdit software (Hall, 1999). Spacer sequences for each CRISPR were analysed against env_nt BLAST using Geneious software with default parameters. In order to obtain information about gene or taxonomic affiliation, a portion of the best hit sequences (max bit score) containing the spacers of at least 100 bp were analysed against nr and nt BLAST database using blastx and blastn respectively.

Acknowledgements

We thank Captain Emanuele Gentile and all crew of RV *Urania* for their valuable professionalism and support during the cruises. This work was performed with the financial support of CNR in frame of ESF Project 09-EuroEEFGFP-044 Deep_C, with the support of the European Community in frame of Project FP7-KBBE-2009-2B-226977 (MAMBA) and of Italian Ministry of Education, Universities and Research (MIUR) in the frame of the Project FIRB 2008 EXPLODIVE RBFR08AWP6_002.

References

Alonso-Sáez, L., Galand, P.E., Casamayor, E.O., Pedrós-Alió, C., and Bertilsson, S. (2010) High bicarbonate assimilation in the dark by Arctic bacteria. *ISME J* **4:** 1581–1590.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., *et al*. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402.

Aristegui, J., Gasol, J.M., Duarte, C.M., and Herndl, G.J. (2009) Microbial oceanography of the dark ocean's pelagic realm. *Limnol Oceanogr* **54:** 1501–1529.

Arnosti, C. (2002) Microbial extracellular enzymes and their role in dissolved organic matter cycling. In *Aquatic Ecosystems: Interactivity of Dissolved Organic Matter*. Findlay, S., and Sinsabaugh, R.L. (eds). San Diego, CA, USA: Academic Press, pp. 316–342.

Arrigo, K.R. (2005) Review Marine microorganisms and global nutrient cycles. *Nature* **437:** 349–355.

Azam, F., and Long, R.A. (2001) Sea snow microcosms. *Nature* **414:** 497–498.

Barberán, A., Fernández-Guerra, A., Auguet, J.C., Galand, P.E., and Casamayor, E.O. (2011) Phylogenetic ecology of widespread uncultured clades of the Kingdom Euryarchaeota. *Mol Ecol* **20:** 1988–1996.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., *et al*. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315:** 1709–1712.

Boyd, P.W., Jickells, T., Law, C.S., Blain, S., Boyle, E.A., Buesseler, K.O., *et al*. (2007) Mesoscale iron enrichment experiments 1993–2005: synthesis and future directions. *Science* **315:** 612–617.

Brown, M.V., Philip, G.K., Bunge, J.A., Smith, M.C., Bissett, A., Lauro, F.M., *et al*. (2009) Microbial community structure in the North Pacific ocean. *ISME J* **3:** 1374–1386.

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., *et al*. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37** (Suppl. 1)**:** D141–D145.

De Baar, H.J.W., and de Jong, J.T.M. (2001) Distribution, sources and sinks of iron in seawater. In *The Biogeochemistry of Iron in Seawater*. Turner, D., and Hunter, K. (eds). IUPAC Series on Analytical and Physical Chemistry of

Environmental Systems, Vol. 7. Chichester, UK: John Wiley & Sons, pp. 123–253.
DeLong, E.F. (1992) Archaea in coastal marine environments. *Proc Natl Acad Sci USA* **89:** 5685–5689.
DeLong, E.F., and Béjà, O. (2010) The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times. *PLoS Biol* **8:** e1000359.
DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.U., *et al*. (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311:** 496–503.
Deveau, H., Garneau, J.E., and Moineau, S. (2010) CRISPR/Cas system and its role in phage-bacteria interactions. *Annu Rev Microbiol* **64:** 475–493.
Drummond, A., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., *et al*. (2009) *Geneious v4.7*. Biomatters, Auckland.
Edgar, R.C. (2007) PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* **8:** 18.
Eloe, E.A., Fadrosh, D.W., Novotny, M., Zeigler Allen, L., Kim, M., Lombardo, M.J., *et al*. (2011a) Going deeper: metagenome of a hadopelagic microbial community. *PLoS ONE* **6:** e20388.
Eloe, E.A., Shulse, C.N., Fadrosh, D.W., Williamson, S.J., Allen, E.E., and Bartlett, D.H. (2011b) Compositional differences in particle-associated and free-living microbial assemblages from an extreme deep-ocean environment. *Environ Microbiol Rep* **3:** 449–458.
Falgueras, J., Lara, A.J., Fernandez-Pozo, N., Canton, F.R., Perez-Trabado, G., and Claros, M.G. (2010) SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. *BMC Bioinformatics* **11:** 38.
Feingersch, R., Suzuki, M.T., Shmoish, M., Sharon, I., Sabehi, G., Partensky, F., *et al*. (2010) Microbial community genomics in eastern Mediterranean Sea surface waters. *ISME J* **4:** 78–87.
Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., *et al*. (2008) Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci USA* **105:** 3805–3810.
Galperin, M.Y. (2005) A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. *BMC Microbiol* **5:** 35.
Ghai, R., Martin-Cuadrado, A.B., Molto, A.G., Heredia, I.G., Cabrera, R., Martin, J., *et al*. (2010) Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing. *ISME J* **4:** 1154–1166.
Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41:** 95–98.
Hansell, D.A., Carlson, C.A., Repeta, D.J., and Schlitzer, R. (2009) Dissolved organic matter in the ocean: new insights stimulated by a controversy. *Oceanography* **22:** 52–61.
Herndl, G.J., Reinthaler, T., Teira, E., van Aken, H., Veth, C., Pernthaler, A., *et al*. (2005) Contribution of Archaea to total prokaryotic production in the deep Atlantic Ocean. *Appl Environ Microbiol* **71:** 2303–2309.
Huber, T., Faulkner, G., and Hugenholtz, P. (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20:** 2317–2319.
Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007) MEGAN analysis of metagenomic data. *Genome Res* **17:** 377–386.
Ivars-Martinez, E., Martin-Cuadrado, A.B., D'Auria, G., Mira, A., Ferriera, S., Johnson, J., *et al*. (2008) 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:** 1194–1212.
Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res* **32:** D277–D280.
Kiørboe, T., and Jackson, G.A. (2001) Marine snow, organic solute plumes, and optimal sensory behaviour of bacteria. *Linmol Oceanogr* **46:** 1309–1318.
Kirchman, D., K'nees, E., and Hodson, R. (1985) Leucine incorporation and its potential as a measure of protein synthesis by bacteria in natural aquatic systems. *Appl Environ Microbiol* **49:** 599–607.
Kirchman, D.L. (1993) Leucine incorporation as a measure of biomass production by heterotrophic bacteria. In *Handbook of Methods in Aquatic Microbial Ecology*. Vol. 58. Kemp, P.F., Sherr, B.F., Sherr, E.B., and Cole, J.J. (eds). Boca Raton, FL, USA: Lewis Publishers, pp. 509–512.
Klappenbach, J.A., Saxman, P.R., Cole, J.R., and Schmidt, T.M. (2001) rrndb: the ribosomal RNA operon copy number database. *Nucleic Acids Res* **29:** 181–184.
Konstantinidis, K.T., Braff, J., Karl, D.M., and DeLong, E.F. (2009) Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific subtropical gyre. *Appl Environ Microbiol* **75:** 5345–5355.
Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., *et al*. (2004) Versatile and open software for comparing large genomes. *Genome Biol* **5:** R12.
La Cono, V., Tamburini, C., Genovese, L., La Spada, G., Denaro, R., and Yakimov, M.M. (2009) Cultivation-independent assessment of the bathypelagic archaeal diversity of Tyrrhenian Sea: comparative study of rDNA and rRNA-derived libraries and influence of sample decompression. *Deep Sea Res Part II Top Stud Oceanogr* **56:** 768–773.
La Cono, V., Smedile, F., Ferrer, M., Golyshin, P.N., Giuliano, L., and Yakimov, M.M. (2010) Genomic signatures of fifth autotrophic carbon assimilation pathway in bathypelagic *Crenarchaeota*. *Microb Biotechnol* **3:** 595–606.
Lane, D.J. (1991) 16S/23S rRNA sequencing. In *Nucleic Acid Techniques in Bacterial Systematics*. Stackerbrandt, E., and Goodfellow, M. (eds). New York, NY, USA: Wiley, pp. 115–175.
Lauro, F.M., and Bartlett, D.H. (2008) Prokaryotic lifestyles in deep sea habitats. *Extremophiles* **12:** 15–25.
Lauro, F.M., Mc Dougald, D., Thomas, T., Williams, T.J., Egan, S., Rice, S., *et al*. (2009) The genomic basis of trophic strategy in marine bacteria. *Proc Natl Acad Sci USA* **106:** 15527–15533.
Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, *et al*. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32:** 1363–1371.
Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., *et al*. (2011) Evolution and

classification of the CRISPR-Cas systems. *Nat Rev Microbiol* **9:** 467–477.

Malanotte-Rizzoli, P., Manca, B.B., Ribera D'Alcalà, M., Theocharis, A., Bergamasco, A., Bregant, D., *et al*. (1997) A synthesis of the Ionian Sea hydrography, circulation and watermass pathways during POEM-Phase I. *Prog Oceanogr* **39:** 153–204.

Martin-Cuadrado, A.B., López-García, P., Alba, J.C., Moreira, D., Monticelli, L., Strittmatter, A., *et al*. (2007) Metagenomics of the deep Mediterranean, a warm bathypelagic habitat. *PLoS ONE* **2:** e914.

Martin-Cuadrado, A.B., Rodriguez-Valera, F., Moreira, D., Alba, J.C., Ivars-Martínez, E., Henn, M.R., *et al*. (2008) Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine archaea from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions. *ISME J* **2:** 865–886.

Martin-Cuadrado, A.B., Ghai, R., Gonzaga, A., and Rodriguez-Valera, F. (2009) CO dehydrogenase genes found in metagenomic fosmid clones from the deep Mediterranean Sea. *Appl Environ Microbiol* **75:** 7436–7444.

Martinez, J.S., and Butler, A. (2007) Marine amphiphilic siderophores: marinobactin structure, uptake, and microbial partitioning. *J Inorg Biochem* **101:** 1692–1698.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., *et al*. (2008) The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9:** 386.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., *et al*. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33:** 5691–5702.

Pham, V.D., Konstantinidis, K.T., Palden, T., and DeLong, E.F. (2008) Phylogenetic analyses of ribosomal DNA-containing bacterioplankton genome fragments from a 4000 m vertical profile in the North Pacific Subtropical Gyre. *Environ Microbiol* **10:** 2313–2330.

Pollard, P.C., and Moriarty, D.J.W. (1984) Validity of the tritiated thymidine methods for estimating bacterial growth rates: measurement of isotope dilution during DNA synthesis. *Appl Environ Microbiol* **48:** 1076–1083.

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

Porter, K.G., and Feig, Y.S. (1980) The use of DAPI for identifying and counting aquatic microflora. *Limnol Oceanogr* **25:** 943–948.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., *et al*. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35:** 7188–7196.

Quaiser, A., Zivanovic, Y., Moreira, D., and López-García, P. (2011) Comparative metagenomics of bathypelagic plankton and bottom sediment from the Sea of Marmara. *ISME J* **5:** 285–304.

Raes, J., Korbel, J.O., Lercher, M.J., von Mering, C., and Bork, P. (2007) Prediction of effective genome size in metagenomic samples. *Genome Biol* **8:** R10.

Reinthaler, T., Winter, C., and Herndl, G.J. (2005) Relationship between bacterioplankton richness, respiration, and production in the Southern North Sea. *Appl Environ Microbiol* **71:** 2260–2266.

Reinthaler, T., van Aken, H.M., and Herndl, G.J. (2010) Major contribution of autotrophy to microbial carbon cycling in the deep North Atlantic's interior. *Deep Sea Res Part II Top Stud Oceanogr* **57:** 1572–1580.

Romanenko, V.I. (1964) Heterotrophic assimilation of CO2 by bacterial flora of water. *Mikrobiologiya* **33:** 610–613.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., *et al*. (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5:** e77.

Saager, P.M., Schijf, J., and de Baar, H.J.W. (1993) Trace-metal distributions in seawater and anoxic brines in the eastern Mediterranean Sea. *Geochim Cosmochim Acta* **57:** 1419–1432.

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., *et al*. (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34:** 374–378.

Santinelli, C., Nannicini, L., and Seritti, A. (2010) DOC dynamics in the meso and bathypelagic layers of the Mediterranean Sea. *Deep Sea Res Part II Top Stud Oceanogr* **57:** 1446–1459.

Seritti, A., Manca, B.B., Santinelli, C., Murru, E., Boldrin, A., and Nannicini, L. (2003) Relationships between dissolved organic carbon (DOC) and water mass structure in the Ionian Sea (winter 1999). *J Geophys Res* **108:** 8112.

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

Smith, D.C., and Azam, F. (1992) A simple, economical method for measuring bacterial protein synthesis rates in seawater using 3H-leucine. *Mar Microbial Food Webs* **6:** 107–114.

Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., *et al*. (2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci USA* **103:** 12115–12120.

Sorokin, D.Y., van den Bosch, P.L., Abbas, B., Janssen, A.J., and Muyzer, G. (2008) Microbiological analysis of the population of extremely haloalkaliphilic sulfur-oxidizing bacteria dominating in lab-scale sulfide-removing bioreactors. *Appl Microbiol Biotechnol* **80:** 965–975.

Stackebrandt, E., and Goodfellow, M. (1991) *Nucleic Acid Techniques in Bacterial Systematics*. Chichester, UK: Wiley.

Stern, R.J. (2002) Subduction zones. *Rev Geophys* **40:** 1012.

Swan, B.K., Martinez-Garcia, M., Preston, C.M., Sczyrba, A., Woyke, T., Lamy, D., *et al*. (2011) Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* **333:** 1296–1300.

Takai, K., and Horikoshi, K. (2000) Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. *Appl Environ Microbiol* **11:** 5066–5074.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28:** 2731–2739.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., *et al*. (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4:** 41.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., *et al*. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66–74.

Woodward, E.M.S. (1994) Nanomolar ammonia concentrations in the Western Mediterranean Sea. 5th EROS 2000 Workshop, Hambourg, March 28–30.

Xu, K., and Ma, B.G. (2007) Comparative analysis of predicted gene expression among deep-sea genomes. *Gene* **397:** 136–142.

Yakimov, M.M., La Cono, V., Denaro, R., D'Auria, G., Decembrini, F., Timmis, K.N., *et al*. (2007) Primary producing prokaryotic communities of brine, interface and seawater above the halocline of deep anoxic lake L'Atalante, Eastern Mediterranean Sea. *ISME J* **1:** 743–755.

Yakimov, M.M., La Cono, V., Smedile, F., Deluca, T.H., Juárez, S., Ciordia, S., *et al*. (2011) Contribution of crenarchaeal autotrophic ammonia oxidizers to the dark primary production in Tyrrhenian deep waters (Central Mediterranean Sea). *ISME J* **5:** 945–961.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Reads length distribution obtained from Matapan metagenome after trimming.

Fig. S2. Contigs length distribution obtained from Matapan metagenome assembled with GS De Novo Assembler of the Genome Sequencer FLX data analysis suite (version 2.3) with the default parameters applied.

Fig. S3. A. Relative proportions of domains identified in different metagenomic data sets based on the taxonomic binning of protein-encoding genes against the SEED database. The total number of protein-encoding genes matches in each metagenome is indicated in parentheses. The taxonomic assignment was carried out using a cut-off expectation (*E*) value of 1e-05 and alignment length of 30 bp inside the MG-RAST pipeline.

B. Relative proportions of Archaeal orders identified in different metagenomic data sets based on the taxonomic binning of protein-encoding genes against the SEED database. Only orders having $\geq$ 0.01% match in at least one of all eight metagenomes compared were showed. The total number of protein-encoding genes matches in each metagenome is indicated in parentheses. The taxonomic assignment was carried out using a cut-off expectation (*E*) value of 1e-05 and alignment length of 30 bp inside the MG-RAST pipeline.

C. Relative proportions of eubacterial classes identified in different metagenomic data sets based on the taxonomic binning of protein-encoding genes against the SEED database. Only classes having $\geq$ 0.01% match in at least one of all 8 metagenomes compared were showed. The total number of protein-encoding genes matches in each metagenome is indicated in parentheses. The taxonomic assignment was carried out using a cut-off expectation (*E*) value of 1e-05 and alignment length of 30 bp inside the MG-RAST pipeline.

D. Relative proportions of Eukaryota phylum identified in different metagenomic data sets based on the taxonomic binning of protein-encoding genes against the SEED database. Only phylum having $\geq$ 0.01% match in at least one of all eight metagenomes compared were showed. The total number of protein-encoding genes matches in each metagenome is indicated in parentheses. The taxonomic assignment was carried out using a cut-off expectation (*E*) value of 1e-05 and alignment length of 30 bp inside the MG-RAST pipeline.

Fig. S4. Cluster analysis of several metagenomes based on matches to different KEGG categories expressed as percentage to the respective number of proteins identified in the metagenomes.

Table S1. A. Taxonomic classification results performed by MG-RAST using RDP SSU database.

B. Taxonomic classification results performed by MG-RAST using SILVA SSU database.

C. Taxonomic classification performed by MG-RAST using SEED proteins database. Results are expressed in number of proteins identified.

Table S2. A. Percentage of SEED, COG and KEGG functional categories in all metagenomes analysed.

B. Comparison of transposase proteins abundance in all metagenomes analysed.

Table S3. A. SEED functional categories classification performed by MG-RAST. Results are expressed in number of proteins identified.

B. COG functional categories classification performed by MG-RAST. Results are expressed in number of proteins identified.

C. KEGG functional categories classification performed by MG-RAST. Results are expressed in number of proteins identified.

Table S4. A. CRISPR summary table divided by metagenome projects.

B. Summary table of CRISPR in all metagenomes of this study.

C. Spacers summary table divided by metagenome projects.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.