Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters

Mylène Hugoni, Najwa Taib, Didier Debroas, Isabelle Domaizon, Isabelle Jouan-Dufournel, Gisèle Bronner, Ian Salter, Hélène Agogué, Isabelle Mary, Pierre E. Galand

Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters

Mylène Hugoni[a,b,1], Najwa Taib[a,b,1], Didier Debroas[a,b], Isabelle Domaizon[c], Isabelle Jouan Dufournel[a,b], Gisèle Bronner[a,b], Ian Salter[d,e], Hélène Agogué[f], Isabelle Mary[a,b,2], and Pierre E. Galand[d,g]

[a]Laboratoire "Microorganismes: Génome et Environnement," Clermont University, Université Blaise Pascal, F-63000 Clermont-Ferrand, France; [b]Laboratoire Microorganismes, Génome et Environnement, Centre National de la Recherche Scientifique (CNRS), Unité Mixte de Recherche (UMR) 6023, F-63171 Aubière, France; [c]Institut National de la Recherche Agronomique, UMR 42 Centre Alpin de Recherche sur les Réseaux Trophiques et Ecosystèmes Limniques, F-74200 Thonon les Bains, France; [d]Université Pierre et Marie Curie–Paris 6, UMR 8222, Laboratoire d'Ecogéochimie des Environments Benthiques (LECOB), UMR 7621, Laboratoire d'Océanographie Microbienne (LOMIC), Observatoire Océanologique, F-66650 Banyuls-sur-Mer, France; [e]CNRS, UMR 7621, LOMIC, Observatoire Océanologique, F-66650 Banyuls-sur-Mer, France; [f]Littoral, Environnement et Sociétés, UMR 7266, CNRS, University of La Rochelle, 17000 La Rochelle, France; and [g]CNRS, UMR 8222, LECOB, Observatoire Océanologique, F-66650 Banyuls-sur-Mer, France

Edited by David M. Karl, University of Hawaii, Honolulu, HI, and approved March 1, 2013 (received for review September 28, 2012)

Marine Archaea are important players among microbial plankton and significantly contribute to biogeochemical cycles, but details regarding their community structure and long-term seasonal activity and dynamics remain largely unexplored. In this study, we monitored the interannual archaeal community composition of abundant and rare biospheres in northwestern Mediterranean Sea surface waters by pyrosequencing 16S rDNA and rRNA. A detailed analysis of the rare biosphere structure showed that the rare archaeal community was composed of three distinct fractions. One contained the rare Archaea that became abundant at different times within the same ecosystem; these cells were typically not dormant, and we hypothesize that they represent a local seed bank that is specific and essential for ecosystem functioning through cycling seasonal environmental conditions. The second fraction contained cells that were uncommon in public databases and not active, consisting of aliens to the studied ecosystem and representing a nonlocal seed bank of potential colonizers. The third fraction contained Archaea that were always rare but actively growing; their affiliation and seasonal dynamics were similar to the abundant microbes and could not be considered a seed bank. We also showed that the major archaeal groups, Thaumarchaeota marine group I and Euryarchaeota group II.B in winter and Euryarchaeota group II.A in summer, contained different ecotypes with varying activities. Our findings suggest that archaeal diversity could be associated with distinct metabolisms or life strategies, and that the rare archaeal biosphere is composed of a complex assortment of organisms with distinct histories that affect their potential for growth.

ong-term dynamic | dormancy | taxonomic diversity | microbial observatory | Somlit

The seasonal dynamics of marine microorganisms have traditionally been studied at the DNA level (1, 2), but recent studies have shown the importance of differentiating the active communities from the total communities (3–5). One method to explore an aspect of activity (i.e., the growth rate for specific taxa) is to investigate microbial communities with both 16S rRNA and 16S rDNA (6-8). The use of the 16S rRNA-to-rDNA sequence ratio as an index of microbial growth has revealed a generally positive correlation between abundance and activity in coastal surface bacterial communities (4, 9). However, abundant microbes are not always the most active (3), even though they contribute greatly to ecosystem functioning. An important finding is that growth can be detected among low-abundance taxa, also known as the rare biosphere (4, 7), which was first defined with the development of new sequencing technologies, allowing a deep coverage of the diversity of natural communities (10). Rare taxa have been hypothesized to consist of dormant microorganisms (or a seed bank) that could potentially be resuscitated under different environmental conditions (11). However, the discoveries that the rare biosphere had a biogeography (12), and that a significant portion of the rare community was active (4, 7), with growth rates that decreased as abundance increased (4), suggest that the rare biosphere is not solely a dormant seed bank (13). A rare biosphere has been detected within the domain Archaea (12), and although we have begun to gain insights into the dominant archaeal phylotypes, the community structure of the rare Archaea remains largely uncharacterized.

Marine planktonic Archaea have been recently recognized as main drivers of the aerobic ammonia oxidation in many aquatic ecosystems, suggesting an important role in the nitrogen cycle (14–16). They have traditionally been described as spanning three major groups: Thaumarchaeota marine group (MG) I, which is more abundant in meso- and bathypelagic waters (17–19), Euryarchaeota MGII, which is more abundant in surface waters, and Euryarchaeota MGIII, which is restricted to deeper waters (20, 21). The diversity of Archaea is, however, much more complex; for instance, MGI appears to have distinct clusters segregated according to depth and location (22). A recent metagenomic characterization of MGI from north Atlantic coastal surface waters also suggested the presence of at least two dominant environmental populations that are divergent from each other (23). The presence of at least two clusters was also demonstrated in the Mediterranean Sea (24) and corresponded to groups previously detected in different oceanic provinces (20). Whether this taxonomic diversity corresponds to distinct ecotypes, i.e., groups of microorganisms playing distinct ecological roles and belonging to genetically cohesive and irreversibly separate evolutionary lineages (25), is not known because the relationship among archaeal activity, environmental conditions, and sequence abundance has never been studied. Moreover, the ecological control of archaeal diversity patterns over long time scales remains poorly understood (24).

By monitoring surface archaeal communities in monthly intervals during a 3.5-y period at the Banyuls-sur-Mer Bay Microbial Observatory, a site representative of the coastal northwest Mediterranean Sea, we aimed to describe the structure of the rare archaeal

Author contributions: I.D., H.A., I.M., and P.E.G. designed research; M.H. and N.T. performed research; M.H., N.T., D.D., I.S., and I.M. contributed new reagents/analytic tools; M.H., N.T., I.J.D., G.B., and P.E.G. analyzed data; M.H., N.T., I.M., and P.E.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The pyrosequencing data reported in this paper has been deposited in the Dryad database, http://datadryad.org (doi no. 10.5061/dryad.q5903).

[1]M.H. and N.T. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: isabelle.mary@univ-bpclermont.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1216863110/-/DCSupplemental.

biosphere by testing whether it is composed of a seed bank of dormant cells or represents microorganisms with high growth rates. By targeting both 16S rRNA and rDNA, we also verified whether different archaeal clusters represent distinct ecotypes and assessed the seasonal activity dynamics of marine Archaea.

Results

Rare and Abundant Phylotypes. We used pyrosequencing to follow changes in the community structure, relative sequence abundance, and potential activity of Archaea over time. A total of 351 operational taxonomic units (OTUs) were retrieved in the 16S rDNA dataset, representing a total of 65,833 sequences. Seventeen OTUs were abundant (>1% and occurred in more than one sample) and contained 97% of all of the sequences. The 16S rRNA dataset was composed of 52,181 sequences consisting of 348 OTUs. Rarefaction curves for both 16S rDNA and 16S rRNA indicated that, in most cases, the sequencing depth captured the diversity present in the natural archaeal community (Fig. S1).

Only two OTUs were always abundant (OTU 2 affiliated with MGI, and OTU 13 affiliated with MGII.B), whereas the remaining 15 abundant OTUs were rare in some samples (≤0.2% of the sequences in a sample). Typically, OTUs abundant in winter became rare in summer and vice versa. All the abundant OTUs were active when they were abundant (Fig. 1*A*), and the plot of the 16S rDNA against 16S rRNA OTU frequencies had an intercept at zero and showed a high correlation between 16S rRNA and 16S rDNA (Kendall nonparametric $\tau = 0.7$; $P < 0.001$; $n = 224$). However, 16S rRNA and 16S rDNA were more poorly correlated when the abundant OTUs became rare ($\tau = 0.3$; $P < 0.001$; $n = 72$), with some OTUs showing high activity (16S rRNA/rDNA ratio >1) whereas others had low or no activity (16S rRNA/rDNA ratio <1; Fig. 1*B*).

All OTUs were compared with the entire SILVA database to ascertain if they were globally common (i.e., a high similarity to reference sequences) or uncommon (i.e., a low similarity). The abundant DNA OTUs were common, with an average 98% sequence similarity to the public database sequences (Fig. 2*A*). The always rare DNA also contained a group of common OTUs (96% similarity), but, notably, half the OTUs were uncommon, with only 84% identity to the public reference sequences (Fig. 2*B*). The abundant and always-rare RNA OTUs were common (98% and 96% sequence identity, respectively; Fig. 2) for the active fraction of the community, and the absence of uncommon OTUs in the rare 16S rRNA fraction indicates that the uncommon OTUs were never active. The low similarity to the SILVA database displayed by the uncommon rare OTUs (84% identity) suggests that they originated from undersampled ecosystems, not well covered by the public database. The closest relatives to the uncommon rare OTUs belonged to the Euryarchaeota Deep Hydrothermal Vent Euryarchaeotic Group 6 (DHVEG-6), pMC1, and South African Gold Mine Euryarchaeotic Group-1 (SAGMEG-1) clusters, which are frequently detected in deep marine sediments (26). In contrast, the rare but common OTUs were identified as MGI and MGII.

Archaeal Community Structure, Dynamics, and Activity. The OTUs abundance followed a log-series distribution for the 16S rDNA and rRNA datasets (Fig. S2 *A* and *B*), with most OTUs included in the first octaves (i.e., species characterized by a low number of reads). The abundant OTUs belonged mostly to MGI and MGII.A and MGII.B (Fig. S2*C*) but also to MGIII. In the active fraction (the 16S rRNA dataset), the major taxonomic groups (MGI, MGII.A, and MGII.B) represented ∼93% of the reads (Fig. S2*C*). Interestingly, some abundant OTUs in the 16S rDNA dataset were less represented in the 16S rRNA dataset, for example, OTUs 2 and 9 affiliated with MGI, suggesting a weak activity. In contrast, some abundant OTUs were also very active, as shown by a greater relative abundance of 16S rRNA, such as OTU 28 affiliated with MGI (Fig. S2*C*).

The MGI sequences followed a seasonal pattern and were more abundant during winter (Fig. 3*A*). The MGI 16S rRNA dynamics showed the same trend as that for 16S rDNA, suggesting metabolically active communities. Our analysis showed that the MGI OTUs fell into four different clusters: A, B, C, and D (Fig. S3*A*). Most of the OTUs were affiliated with MGI.B, followed by MGI.A, which is closely related to *Nitrosopumilus maritimus*. These two clusters comprised all the abundant MGI OTUs. Interestingly, MGI.A and MGI.B sequences exhibited alternative patterns of 16S rDNA and rRNA representation. The MGI.A sequences outnumbered the MGI.B sequences in the 16S rDNA dataset (approximately two times more), whereas the opposite was observed for the 16S rRNA dataset (Fig. 4). This result suggested that MGI.B was much more active than MGI.A, even though it was not the most abundant in the ecosystem. The rare OTUs belonged to MGI.C, which is affiliated with sequences retrieved from deep waters (20) and distantly related to *Cenarchaeum symbiosum* and to MGI.D. This cluster was distinct from the others (89–92% similarity) and emerged earlier in the phylogeny. MGI.C was active when present, whereas MGI.D was not always active when present.

The MGII.A sequence abundance showed marked differences between seasons, with the highest relative abundance during the summer period (from May to October) and the lowest during the winter months (Fig. 3*B*). The 16S rRNA dataset revealed a similar seasonal pattern of activity. In contrast, MGII.B dominated in abundance and activity during winter, with the highest relative abundance in February and recurrent peaks each year (Fig. 3*C*). MGII.A was more active than MGII.B, consistent with its higher relative abundance (Fig. 3 *B* and *C*). Euryarchaeota MGII.A was

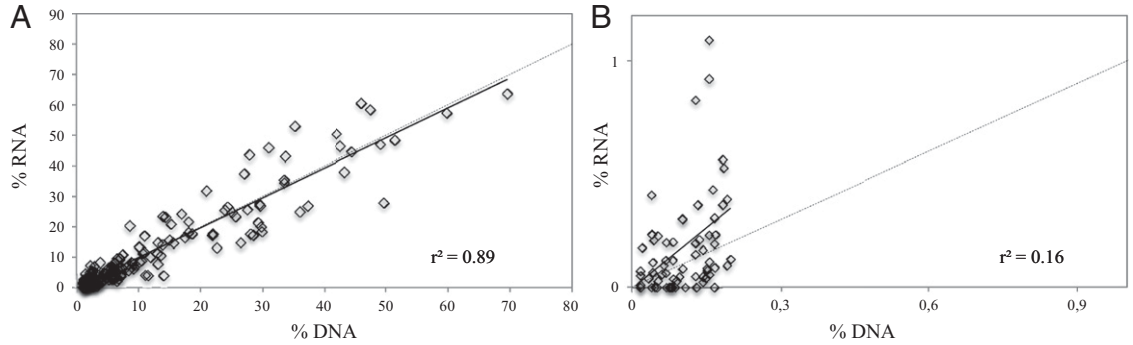

Fig. 1. 16S rDNA against 16S rRNA OTUs frequencies for abundant OTUs when they are abundant (*A*) and when they become rare (*B*). The RNA and DNA frequencies are plotted against each other for all abundant OTUs and all time points. The black line represents the regression, and the dotted line is the 1:1 line.

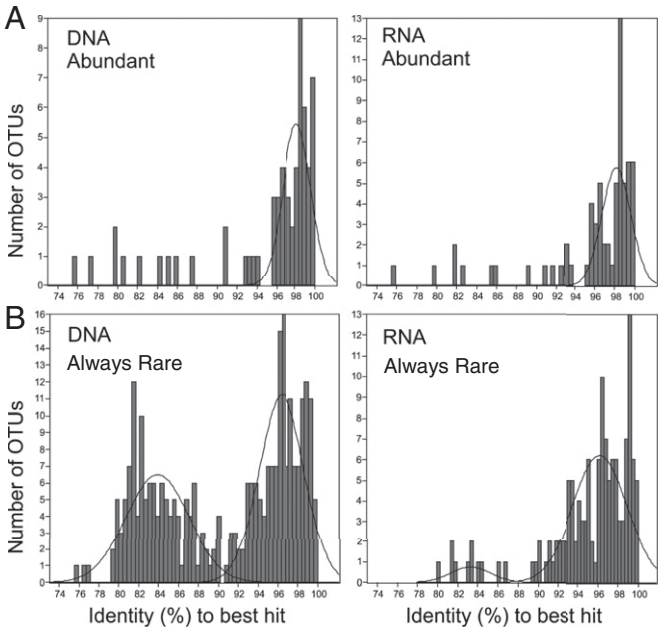

Fig. 2. Distribution of the percent identity from a comparison between public database sequences (SILVA) and abundant 16S rDNA and rRNA sequences (*A*) and always-rare 16S rDNA and rRNA sequences (*B*). The data are fitted to a model of normal distributions (black lines) that identifies groups of OTUs as common (i.e., a high percentage identity) or uncommon (i.e., a low percentage identity).

separated into two previously described (20) main subclusters (M and K; Fig. S4), and most of the sequences belonged to subcluster M, which was also the most active (Fig. S5*A*). The subcluster M activity pattern was different from that of subcluster K (Fig. S5). Most MGII.B sequences and activity were affiliated with the WHARN subcluster (Figs. S5*B* and S6) that corresponds to phylotypes II-CC, which are widely distributed in surface waters of various oceanic provinces (20). Other Euryarchaeota were affiliated with the MGIII and the RC-V cluster and with methanogenic lineages (Fig. S6).

Less abundant groups, including OTUs affiliated with MGIII, were also present and active during winter but were also detected in July 2008 and 2009, together with reduced activity. The Miscellaneous Euryarchaeotic Group (MEG) and DHVEG-6 did not present seasonal patterns of relative abundance and activity.

The canonical correspondence analysis plot (*SI Materials and Methods*) showed a clear difference between the activity of the two MGII clusters (Fig. S7): MGII.A appeared as a summer community associated mainly with temperature, whereas the activity of MGII.B was related to such winter features as nitrite, nitrate, and oxygen. These winter features also characterized the activity of MGI overall, and there were fewer differences between the different MGI clusters when considering the parameters followed in the present study. Contrary to MGII, MGI clusters were discriminated according the second axis, which was positively correlated with phosphate (Fig. S7).

Discussion

Our long-term study of archaeal dynamics and activity in surface Mediterranean waters showed that rare Archaea were heterogeneous in their pattern of seasonal activity and phylogenetic affiliation. We propose that the rare archaeal biosphere could be divided into three different fractions classified as follows: the local seed bank, the nonlocal seed bank (or the alien colonizers), and the active-but-always-rare fraction.

The local seed bank represented Archaea that were rare but became abundant at certain times. When abundant, their 16S rDNA and 16S rRNA sequences were closely correlated, indicating that these OTUs were also active. Scatter plots of 16S rRNA vs. rDNA yielded an intercept at zero, suggesting that growth rates were constant as abundance varied (4). However, when these OTUs became rare, their 16S rDNA and rRNA sequences were poorly correlated, which, according to a described model (4), indicates increasing or decreasing growth rates as abundance decreases. Such variable activity suggests changing growth rates, possibly reflecting differences in the metabolic state of the cells as they cycle between abundant and rare fractions. Contrasting activity levels among rare microbes have been reported recently for Bacteria in a coastal system (4) and in lakes (7). Within the context of our seasonal study, the observations could indicate that these rare microorganisms are able to react to seasonal fluctuations of environmental conditions. Moreover, these Archaea could not be considered as being typically dormant cells because some of them lacked dormancy stages (i.e., were always active) and others had only short ones. We propose that this local seed bank maintains sufficient metabolic diversity to react to fluctuating environmental conditions.

The second fraction contained rare Archaea that were uncommon and always inactive in the northwestern Mediterranean Sea. They were aliens to the studied pelagic ecosystem, and their low similarity to database sequences indicates that they may originate from undersampled ecosystems, such as deep marine sediments. This nonlocal seed bank may be dispersed by such episodic events as river flooding, strong storms, or even atmospheric deposition. It is possible that these microorganisms may never grow in the water column as a result of a requirement for very different conditions to those found in the pelagic environment. This fraction of the rare archaeal biosphere could be on its way to extinction (13); alternatively, it may have the ability to

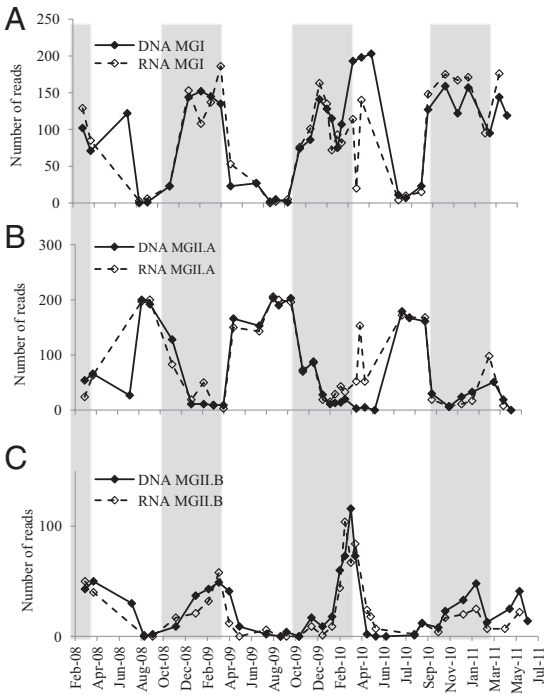

Fig. 3. Seasonal dynamics of the most abundant taxonomic groups in both the 16S rDNA and rRNA sequence datasets: (*A*) MGI, (*B*) MGII.A, and (*C*) MGII.B. Winter months are shown in gray; summer months are shown in white.

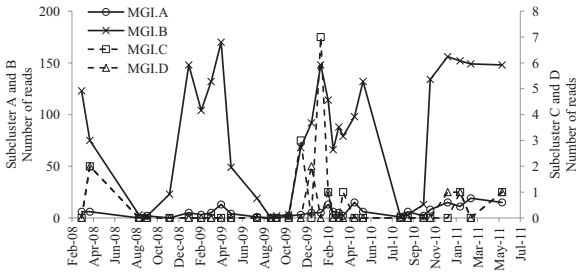

Fig. 4. Seasonal dynamics of active (16S rRNA dataset) MGI clusters: MGI.A, MGI.B, MGI.C, and MGI.D. Winter months are shown in gray; summer months are shown in white.

survive in a dormant stage representing a pool of potential colonizers of very different ecosystems.

Finally, the third fraction of rare Archaea was represented by cells that were always rare but actively growing, and their phylogenetic affiliation and seasonal dynamics were similar to those of the abundant Archaea. Although their activity suggests that these cells are not dormant, the fact that they never made the transition from rare to abundant indicates that they are not operating as a seed bank under the range of environmental conditions encountered during this study. Always rare but actively growing populations may be maintained at a low abundance because they are less efficient than others when competing for resources, or they may have a different life strategy that promotes rarity to minimize predation (11). Alternatively, these groups may be more susceptible to specific viral attack (27).

Although MGI was overall more abundant and active in winter, we found differences in activity between the clusters. Four MGI clusters were identified through our analysis: MGI.A (or I-α), MGI.B (or I-β), and MGI.C (I-γ) were first described in a large-scale study based on terminal restriction fragment length polymorphism (T-RFLP) (20), but only MGI.A and MGI.B were reported in a previous study conducted in the Mediterranean Sea (24). The similar delineation of MGI.A and MGI.B by cloning/sequencing (24) and pyrosequencing confirmed that shorter pyrosequencing reads may be as informative as near full-length sequences, and thus suitable for building phylogenies (28, 29). Interestingly, the MGI.A sequences were more abundant within the 16S rDNA fraction, whereas MGI.B was more abundant within the 16S rRNA fraction, suggesting that the most abundant MGI cluster (i.e., MGI.A) was not the most active. The rarest clusters, MGI.C and MGI.D, also presented different patterns of activity: MGI.C was active when present, yet MGI.D was not always active when present. Both of these rare clusters were active in winter, showing similar responses to such environmental conditions as oxygen, nitrite, nitrate, or Chlorophyll *a* (Chl*a*) content. However, there was a mismatch between their peaks of maximum activity. The variable activity levels for the different MGI clusters could suggest the presence of MGI clusters adapted to different niches. The clusters' recurrent seasonal activity patterns concomitant with reproducible environmental conditions may indicate an ecological specialization and predictable population–environment linkage, comforting the idea of separate ecotypes, as previously shown for Vibrionaceae populations (30). Although we could not precisely define niches in the present study, we hypothesize that the MGI clusters represent ecotypes that correspond to distinct metabolisms, as previously demonstrated for Bacteria (4). In fact, closely related SAR11 phylotypes were recently shown to have very different growth rates (4), which could correspond to differences in metabolism, such as those observed between the SAR11 ecotypes for phosphorus acquisition or glucose utilization (31, 32). Different metabolisms could thus allow the MGI ecotypes to occupy a variety of niches and explain a global ecological success that is similar to that of SAR11 (33).

The MGI seasonal dynamics, illustrated by a higher relative abundance during winter, correlated with the seasonality of such surface water environmental parameters as nitrite and nitrate concentrations. MGI is thought to play an important role in the marine nitrogen cycle by oxidizing ammonia to nitrite (34, 35). In our survey, most of the 16S rDNA sequences belonged to MGI.A, which is closely related to the cultivated planktonic marine Archaea *N. maritimus* (36), an autotrophic ammonia-oxidizer that produces nitrite, suggesting the potential for ammonia oxidation of the MGI found in these waters. This hypothesis is supported by the MGI winter peaks that coincided with an increase in the nitrite and nitrate concentrations, followed by a decrease in ammonia. However, because phytoplankton can also release nitrite (37), the exact origin of winter nitrite in this ecosystem cannot be conclusively determined. Moreover, MGI.C was distantly related to *C. symbiosum*, which is able to use urea as energy and carbon source (38). As archaeal genes for urea utilization have been detected in the marine environment (39–41), we can also speculate that urea could be used as an alternative source of nitrogen and carbon by the MGI clusters that are not closely related to *N. maritimus*, which does not possess genes for urea utilization (42). The decrease in MGI abundance and activity was coincident with an increase in Chl*a*, as previously shown in marine and freshwater ecosystems (43–45). This finding suggests two hypotheses: a limitation of MGI abundance by organic material excreted by phototrophic primary production (36) or a competition with phytoplankton for ammonium that may be unfavorable to MGI (46).

Euryarchaeota were present year-round, but the opposite seasonal dynamics of MGII.A and MGII.B suggests the presence of different ecotypes. The predominance of MGII.A in summer corresponds to a season of generally low abundance of MGII and Archaea in the northwestern Mediterranean Sea (24). The winter peak of MGII.B corresponds to the highest archaeal and MGI abundances (24), suggesting separate niches for the two MGII clusters in surface waters. Their abundance and activity dynamics might be affected by competition for resources with other organisms and could reflect the development of different strategies to improve their metabolic potential. For instance, genes encoding proteorhodopsin have been found in members of the Euryarchaeota MGII.A in the surface waters of the North Pacific (47), and recent studies showed a single copy of the proteorhodopsin (*pop*) gene in the reconstruction of a coastal MGII.A genome (48). In our study, the MGII.A sequences were only 90% similar to the proteorhodopsin clade from a previous work (47), but closer to the sequence reported in another (48) (96% similarity). We therefore hypothesize that phototrophic metabolism could be present in the Mediterranean Sea: MGII.A could use light as an energy source, explaining the summer peaks of abundance and activity, as irradiance is more important in this season. In contrast, the MGII.B distribution in the rRNA datasets correlated with nitrogen compounds, as with MGI.

The results based on the 16S rRNA/rDNA ratios could have been affected by the 16S rDNA copy number per genome. However, to our knowledge, all available complete genomes of mesophilic Archaea, including representatives from Euryarchaeota and Thaumarchaeota showed only one copy of 16S rDNA: thus, we assume that this is also the case in the natural communities. To assess how well our sequence data could represent the community abundance, we compared the pyrosequencing quantification to the metagenomic data obtained in September 2010 from the same sampling station through the Global Ocean Sampling project. The two methods showed a strong dominance of Euryarchaeota (95% and 88% for the pyrosequencing and metagenomic data, respectively) vs. Thaumarchaeota, suggesting no major primer bias at the phylum level in our approach. A

significant correlation has also been found between the abundance estimated by quantitative methods and pyrosequencing for Bacteria (4, 6, 9) and by sequencing approaches for MGI in the northwestern Mediterranean Sea (24). We therefore hypothesize that the relative sequence abundance measured in this study was comparable to the cell abundance dynamics.

In summary, this study clearly showed that the rare biosphere could not solely be characterized as a seed bank of dormant cells; rather, it is a complex association of indigenous and itinerant cell types with contrasted origins and fate that contribute to microbial interaction networks and metabolic processes in the environment. Our phylogenetic affiliation suggested that the diversity found within the environmental clusters of Archaea may correspond to different activity levels or growth rates, thus possibly illustrating different metabolism and life strategies. Our results show that we need to rethink our view of how abundant and rare microbes contribute to ecosystem processes.

Materials and Methods

Sampling and Environmental Parameters. Surface seawater (3 m) was collected monthly from March 2008 to June 2011 (40 samples) by using a 10-L Niskin bottle at the Service d'Observation du Laboratoire Arago station (42°31′N, 03°11′E) in the Bay of Banyuls-sur-Mer in France. The water was kept in 10-L high density polyethylene carboys in the dark until being processed in the laboratory (within 1.5 h). A subsample of 5 L was prefiltered through 3-μm pore-size polycarbonate filters (Millipore), and the microbial biomass was collected on 0.22-μm pore-size GV Sterivex cartridges (Millipore) and stored at −80 °C until nucleic acid extraction. The physicochemical parameters (Fig. S8) were provided by the Service d'Observation en Milieu Littoral (www.domino.u-bordeaux.fr/somlit_national).

The water sample used for the metagenomic analysis was collected at 3 m depth on 28 September 2010 as part of the J. Craig Venter Institute European Sampling Expedition following a protocol previously published (49). Annotation of the metagenomic data were performed through the J. Craig Venter Institute metagenomics analysis pipeline (San Diego) (50).

Nucleic Acid Extraction and Pyrosequencing. The nucleic acid extraction method was modified from Lami et al. (8) by using a combination of mechanic and enzymatic cell lysis applied directly to Sterivex cartridges, followed by extraction by using the AllPrep DNA/RNA kit (Qiagen). The RNA samples were tested for the presence of contaminating genomic DNA by PCR and then reverse-transcribed with random primers using the SuperScript III Reverse Transcriptase kit (Invitrogen). The amplification of the V3–V5 region of the 16S rRNA gene was performed by Research and Testing Laboratory (Lubbock, TX) with universal archaeal primers Arch349F (CCC TAC GGG GTG CAS CAG) and Arch806R (GGA CTA CVS GGG TAT CTA AT) (51), followed by pyrosequencing by using a Roche 454 GS-FLX system with titanium chemistry.

Bioinformatic Analysis and Statistics. The pyrosequencing data produced from the 80 samples (16S rDNA and 16S rRNA) represented 477,589 raw sequences. All sequences were checked against the following quality criteria: (*i*) no Ns; (*ii*) quality score ≥27 according to PANGEA trimming (52); (*iii*) a minimum sequence length of 200 bp; (*iv*) no sequencing error in the forward primer; and (*v*) no chimeras [checked with UCHIME (53)]. The quality filtering step eliminated ∼15% of all sequences (1.6% were chimeras). The remaining reads were clustered using USEARCH (54) at a 97% similarity threshold (55). For the taxonomic affiliation, we constructed a dedicated archaeal database based on the SSURef 108 database of the SILVA project (56) and added annotated reference sequences from the Mediterranean Sea (24). The process was automated by PANAM (http://code.google.com/p/panam-phylogenetic-annotation/downloads/list) that constructs phylogenetic trees for taxonomic annotation (57) as detailed in *SI Materials and Methods*. After that step, all sequences affiliated to Bacteria were removed from the data set, leaving a total of 65,833 archaeal sequences for the 16S rDNA dataset and 52,181 sequences for the 16S rRNA dataset (Table S1). Phylogenetic trees containing only the main taxonomic groups detected by PANAM (MGI Thaumarchaeota, MGII.A and MGII.B Euryarchaeota), and environmental OTUs affiliated with those groups, are included as Figs. S3, S4, and S6.

For the analysis of the seasonal dynamics, the 16S rDNA and 16S rRNA samples were randomly resampled down to 208 sequences by using DaisyChopper (www.genomics.ceh.ac.uk/GeneSwytch/). We chose to resample down to a relatively low number of sequences to retain the largest possible number of samples; a total of 12 samples were discarded because of a low number of sequences (< 208). However, for the analysis of the rare biosphere, a deeper sequencing effort was needed to define the rare Archaea, and only samples with >488 sequences were retained (55 samples). To verify if the different sampling cutoff could bias our analysis, we compared the seasonal dynamics based on 208 sequences per samples to that based on 488 sequences. The two results were similar for the major groups, as, for example, for MGI (Fig. S9). We also compared the number and identity of the abundant OTUs found for each cutoff. The entire 16S rDNA dataset, the one resampled at 208 sequences, and the one resampled at 488 sequences, showed 17, 18, and 21 abundant OTUs (> 1%), respectively (19, 22, and 21 for the 16S rRNA sequences). Notably, the abundant OTUs were always the same in the different datasets.

Defining Abundant and Rare Phylotypes. OTUs were considered abundant when they comprised more than 1% of the sequences (11) and were present in more than one sample. In contrast, rare OTUs were defined as OTUs representing ≤0.2% of the sequences in a sample (present once in a sample of 488 sequences). This definition is well within the 0.1% to 1% range commonly considered (58), and is more strict than the 1% threshold used recently (4). OTUs were defined as always rare when they were rare in all the samples.

Representative sequences from all OTUs were compared with reference sequences from the entire SILVA database (56) using BlastN (59) to identify the percentage similarity between the queried sequences and their top hits. To assess the commonness of the sequences, the distribution of their percentage identity was plotted and fitted to normal distributions by using a maximum-likelihood method implemented in the mixture analysis of the PAST program (60). The method allowed us to define sequences as common (96–98% identity to database sequences) or uncommon (83% identity in average).

ACKNOWLEDGMENTS. We thank Cyrielle Tricoire, Eric Maria, and the captain and crew of the *Nereis II* for sample collection; and the people involved in the long-term series of hydrobiogeochemical data collected within the Service d'Observation en Milieu Littoral network (SOMLIT). This work was supported by a PhD fellowship from the French Ministère de l'Enseignement Supérieur et de la Recherche (to M.H.), a PhD fellowship from the French Conseil Régional d'Auvergne (to N.T.), and a CNRS Program Ecosphère Continentale et Côtière (EC2CO, 2010–2012). The work of P.E.G is supported by the Agence Nationale de la Recherche (ANR) project MICADO (ANR-11JSV7-003-01). J. Craig Venter Institute (JCVI) Global Ocean sampling, sequencing, and sequence analyses were funded by grants from the Beyster fund of the San Diego Foundation and the Life Technologies Foundation (to JCVI).

1. Alonso-Saez L, et al. (2007) Seasonality in bacterial diversity in north-west Mediterranean coastal waters: Assessment through clone libraries, fingerprinting and FISH. *FEMS Microbiol Ecol* 60:98–112.
2. Mary I, et al. (2006) Seasonal dynamics of bacterioplankton community structure at a coastal station in the western English Channel. *Aquat Microb Ecol* 42: 119–126.
3. Campbell BJ, Kirchman DL (2012) Bacterial diversity, community structure and potential growth rates along an estuarine salinity gradient. *ISME J* 7:210–220.
4. Campbell BJ, Yu L, Heidelberg JF, Kirchman DL (2011) Activity of abundant and rare bacteria in a coastal ocean. *Proc Natl Acad Sci USA* 108:12776–12781.
5. Lennon JT, Jones SE (2011) Microbial seed banks: The ecological and evolutionary implications of dormancy. *Nat Rev Microbiol* 9:119–130.
6. Campbell BJ, Yu L, Straza TRA, Kirchman DL (2009) Temporal changes in bacterial rRNA and rRNA genes in Delaware (USA) coastal waters. *Aquat Microb Ecol* 57: 123–135.
7. Jones SE, Lennon JT (2010) Dormancy contributes to the maintenance of microbial diversity. *Proc Natl Acad Sci USA* 107:5881–5886.
8. Lami R, Ghiglione JF, Desdevises JF, West NJ, Lebaron P (2009) Annual patterns of presence and activity of marine bacteria monitored by 16S rDNA-16SrRNA fingerprints in the coastal NW Mediterranean Sea. *Aquat Microb Ecol* 54:199–210.
9. Gaidos E, Rusch A, Ilardo M (2011) Ribosomal tag pyrosequencing of DNA and RNA from benthic coral reef microbiota: Community spatial structure, rare members and nitrogen-cycling guilds. *Environ Microbiol* 13:1138–1152.
10. Sogin ML, et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere" *Proc Natl Acad Sci USA* 103:12115–12120.
11. Pedros-Alio C (2006) Marine microbial diversity: Can it be determined? *Trends Microbiol* 14:257–263.
12. Galand PE, Casamayor EO, Kirchman DL, Lovejoy C (2009) Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci USA* 106:22427–22432.
13. Pedros-Alio C (2012) The rare bacterial biosphere. *Annu Rev Mar Sci* 4:449–466.

14. Francis CA, Beman JM, Kuypers MM (2007) New processes and players in the nitrogen cycle: The microbial ecology of anaerobic and archaeal ammonia oxidation. *ISME J* 1:19–27.
15. Hugoni M, et al. (2013) Dynamics of ammonia-oxidizing *Archaea* and *Bacteria* in contrasted freshwater ecosystems. *Res Microbiol*, 10.1016/j.resmic.2013.01.004.
16. Karner MB, DeLong EF, Karl DM (2001) Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* 409:507–510.
17. Herndl GJ, et al. (2005) Contribution of *Archaea* to total prokaryotic production in the deep Atlantic Ocean. *Appl Environ Microbiol* 71:2303–2309.
18. Teira E, et al. (2008) Linkages between bacterioplankton community composition, heterotrophic carbon cycling and environmental conditions in a highly dynamic coastal ecosystem. *Environ Microbiol* 10:906–917.
19. Varela MM, van Aken HM, Sintes E, Herndl GJ (2008) Latitudinal trends of *Crenarchaeota* and *Bacteria* in the meso- and bathypelagic water masses of the Eastern North Atlantic. *Environ Microbiol* 10:110–124.
20. Massana R, DeLong EF, Pedros-Alio C (2000) A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. *Appl Environ Microbiol* 66:1777–1787.
21. Galand PE, Casamayor EO, Kirchman DL, Lovejoy C (2009) Unique archaeal assemblages in the Arctic Ocean unveiled by massively parallel tag sequencing. *ISME J* 3: 860–869.
22. Garcia-Martinez J, Rodriguez-Valera F (2000) Microdiversity of uncultured marine prokaryotes: The SAR11 cluster and the marine *Archaea* of Group I. *Environ Microbiol* 9:935–948.
23. Tully BJ, Nelson WC, Heidelberg JF (2012) Metagenomic analysis of a complex marine planktonic thaumarchaeal community from the Gulf of Maine. *Environ Microbiol* 14: 254–267.
24. Galand PE, Gutierrez-Provecho C, Massana R, Gasol J, Casamayor EO (2010) Interannual recurrence of archaeal assemblages in the coastal NW Mediterranean Sea. *Limnol Oceanogr* 55:2117–2125.
25. Koeppel A, et al. (2008) Identifying the fundamental units of bacterial diversity: A paradigm shift to incorporate ecology into bacterial systematics. *Proc Natl Acad Sci USA* 105:2504–2509.
26. Teske A, Sorensen KB (2008) Uncultured *archaea* in deep marine subsurface sediments: Have we caught them all? *ISME J* 2:3–18.
27. Bouvier T, del Giorgio PA (2007) Key role of selective viral-induced mortality in determining marine bacterial community composition. *Environ Microbiol* 9:287–297.
28. Jeraldo P, Chia N, Goldenfeld N (2011) On the suitability of short reads of 16S rRNA for phylogeny-based analyses in environmental surveys. *Environ Microbiol* 13: 3000–3009.
29. Ragan-Kelley B, et al. (2012) Collaborative cloud-enabled tools allow rapid, reproducible biological insights. *ISME J* 7:461–464.
30. Szabo G, et al. (2013) Reproducibility of Vibrionaceae population structure in coastal bacterioplankton. *ISME J* 7:509–519.
31. Coleman ML, Chisholm SW (2010) Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc Natl Acad Sci USA* 107:18634–18639.
32. Schwalbach MS, Tripp HJ, Steindler L, Smith DP, Giovannoni SJ (2010) The presence of the glycolysis operon in SAR11 genomes is positively correlated with ocean productivity. *Environ Microbiol* 12:490–500.
33. Brown MV, et al. (2012) Global biogeography of SAR11 marine bacteria. *Mol Syst Biol* 8:595.
34. Francis CA, Roberts KJ, Beman JM, Santoro AE, Oakley BB (2005) Ubiquity and diversity of ammonia-oxidizing *archaea* in water columns and sediments of the ocean. *Proc Natl Acad Sci USA* 102:14683–14688.
35. Wuchter C, et al. (2006) Archaeal nitrification in the ocean. *Proc Natl Acad Sci USA* 103:12317–12322.
36. Konneke M, et al. (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 437:543–546.
37. Lomas MW, Lipschultz F (2006) Forming the primary nitrite maximum: Nitrifiers or phytoplankton? *Limnol Oceanogr* 51:2453–2467.
38. Hallam SJ, et al. (2006) Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc Natl Acad Sci USA* 103:18296–18301.
39. Baker BJ, Lesniewski RA, Dick GJ (2012) Genome-enabled transcriptomics reveals archaeal populations that drive nitrification in a deep-sea hydrothermal plume. *ISME J* 6:2269–2279.
40. Konstantinidis KT, Braff J, Karl DM, DeLong EF (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.
41. Yakimov MM, 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.
42. Walker CB, et al. (2010) *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc Natl Acad Sci USA* 107:8818–8823.
43. Auguet JC, Casamayor EO (2008) A hotspot for cold *crenarchaeota* in the neuston of high mountain lakes. *Environ Microbiol* 10:1080–1086.
44. Herfort L, et al. (2007) Variations in spatial and temporal distribution of *Archaea* in the North Sea in relation to environmental variables. *FEMS Microbiol Ecol* 62:242–257.
45. Robidart JC, et al. (2012) Seasonal Synechococcus and Thaumarchaeal population dynamics examined with high resolution with remote in situ instrumentation. *ISME J* 6:513–523.
46. Martens-Habbena W, Berube PM, Urakawa H, de la Torre JR, Stahl DA (2009) Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature* 461:976–979.
47. Frigaard NU, Martinez A, Mincer TJ, DeLong EF (2006) Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and *Archaea*. *Nature* 439:847–850.
48. Iverson V, et al. (2012) Untangling genomes from metagenomes: Revealing an uncultured class of marine *Euryarchaeota*. *Science* 335:587–590.
49. Rusch DB, et al. (2007) The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol* 5:e77.
50. Tanenbaum DM, et al. (2010) The JCVI standard operating procedure for annotating prokaryotic metagenomic shotgun sequencing data. *Stand Genomic Sci* 2:229–237.
51. Takai K, Horikoshi K (2000) Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. *Appl Environ Microbiol* 66:5066–5072.
52. Giongo A, et al. (2010) PANGEA: Pipeline for analysis of next generation amplicons. *ISME J* 4:852–861.
53. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194–2200.
54. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461.
55. Kim M, Morrison M, Yu Z (2011) Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *J Microbiol Methods* 84: 81–87.
56. Pruesse E, 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.
57. Taib N, Mangot JF, Domaizon I, Bronner G, Debroas D (2013) Phylogenetic affiliation of SSU rRNA genes generated by massively parallel sequencing: New insights into the freshwater protist diversity. *PLoS ONE*, 10.137/journal.pone.0058950.
58. Fuhrman JA (2009) Microbial community structure and its functional implications. *Nature* 459:193–199.
59. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
60. Hammer Ø, Harper DAT, Ryan PD (2001) PAST: Paleontological Statistics Software Package for Education and Data Analysis. Available at http://folk.uio.no/ohammer/past/. Accessed May 20, 2009.