Spatial patterns of bacterial richness and evenness in the NW Mediterranean Sea explored by pyrosequencing of the 16S rRNA

Thomas Pommier[1,2], P. R. Neal[3], Josep M. Gasol[1], Montserrat Coll[1], Silvia G. Acinas[1], Carlos Pedrós-Alió[1,*]

[1]Departament de Biologia Marina i Oceanografia, Institut de Ciències del Mar, CSIC, 08003 Barcelona, Spain
[2]Ecologie Microbienne, UMR5557, USC1193 CNRS, INRA, Université Lyon I, 69622 Villeurbanne, France
[3]Marine Biological Laboratory, Woods Hole, 02543 Massachusetts, USA

ABSTRACT: Due to analytical limitations, patterns of richness and evenness of microbes are scarce in the current literature. The newest and powerful pyrosequencing technology may solve this issue by sampling thousands of sequences from the same community. We conducted a study of diversity along a horizontal transect (ca. 120 km) and a depth profile (surface to bottom at ca. 2000 m) in the northwestern Mediterranean Sea, using this technology on the V6 region of the 16S rDNA gene and analyzed patterns of richness and evenness of marine free-living bacterial communities. A total of 201 605 tag sequences were obtained from the 10 samples considered and clustered according to their similarity in 1200 operational taxonomic units (OTUs) per sample on average. We found a parallel decrease in richness and evenness from coast to offshore and from bottom to surface. We also observed a predominance of a few OTUs in each sample, while ca. 50% of all OTUs were found as singletons, which indicated that the community structures differed dramatically between sites despite the relative proximity and the physical connectivity between the samples. Despite these differences, using the 300 most abundant OTUs only was sufficient to obtain the same clustering of samples as with the complete dataset. Finally, both richness and evenness were negatively correlated with bacterial biomass and heterotrophic production.

KEY WORDS: Bacterial diversity · Pyrosequencing · Blanes Bay Microbial Observatory

INTRODUCTION

Diversity is considered one of the key properties of ecosystems, necessary to analyze both structure and dynamics of communities (MacArthur & Wilson 1967, Margalef 1972, Whittaker 1972, May 1988, Hubbell 2001, Magurran 2004). This property includes 2 components: richness, that counts how many different species can be found in the community, and evenness, that considers how the individuals in the community are distributed into the different species. Although a large body of literature concerning these components of diversity exists for animals and plants, an understanding of microbial diversity is still lacking. Traditionally, the study of microbial diversity has been hampered by the difficulties in iden-

tifying individuals. For a long time morphology was the only criterion upon which to assign individuals to species. This was clearly useless for bacteria and very difficult for protists. For example, one particularly intensive study analyzed the diversity of 'phytoplankton' (everything seen under an inverted microscope) in the western Mediterranean Sea. Margalef (1995) determined around 20 000 individual phytoplankton cells belonging to 257 taxa in 260 samples. The 3 most abundant taxa in the samples, however, were 'flagellates 1', 'flagellates 2', and 'flagellates 3'. This raises the question of whether the diversity indices would have the same values and whether rank abundance curves would have the same shape if these dominant taxa could have been separated into all their component species.

Molecular approaches seemed to be the solution to this problem. Application of cloning and sequencing of PCR amplified 16S rDNA in the last 2 decades has resulted in thousands of different sequences from the oceans alone (Hagström et al. 2002). Of course, these sequences are not species per se and operational taxonomic units (OTUs) of sequences showing arbitrary levels of similarity have been used instead. But, paradoxically, most molecular studies have analyzed a very limited number of individuals (assuming one clone represents one individual). Most libraries consisted of only a few hundred clones. Even the most careful studies barely reached around 1000 clones per sample (Acinas et al. 2004, Pommier et al. 2007). This is one order of magnitude less than that in the study by Margalef (1995) and clearly insufficient to draw robust conclusions about microbial diversity. Thus, for example, estimates of the total number of bacterial OTUs in the ocean have raised much debate, ranging from a few thousand when based on extensive cloning and sequencing (Hagström et al. 2002, Hong et al. 2006), to hundreds of millions by theoretical approaches (Dykhuizen 1998, Curtis et al. 2002). Furthermore, if a significant number of species were present in low concentrations, the number of clones needed to retrieve them would be enormous (Pedrós-Alió 2006, 2007). If the theoretical estimates of richness mentioned above were true, cloning would be obviously inadequate to analyze microbial diversity.

More powerful sequencing technologies have recently become available and have been used to study richness of marine microbial samples. One such approach consists of obtaining the metagenome of a sample, and breaking it into small fragments that are then cloned and sequenced (Venter et al. 2004). Even though the total number of sequences obtained is enormous, only a small proportion corresponds to 16S rDNA genes. The increase in taxonomically relevant sequences obtained is, therefore, small compared to conventional libraries. Hence, for example, only 4125 rRNA sequences were retrieved during the Global Ocean Sampling (GOS) study (Rusch et al. 2007) which included 811 different OTUs when clustered at a 97 % simliarity level (for the whole study). Large conventional libraries, in turn, produced for each sample between 240 and 500 OTUs at the 97 % similarity level (Acinas et al. 2004, Pommier et al. 2007). The metagenomic approaches have the advantage that other marker genes can be used for determination of richness. Thus, Venter et al. (2004) could compare the rRNA data to the information provided by single copy genes such as *RecA*, *RpoB* and others. In their study, 600 different *RecA* genes were retrieved compared to about 450 different rRNA genes. However, this kind of analysis is not efficient if the purpose of the study is to analyze diversity and not metagenomics.

Sogin et al. (2006) introduced a different approach based on an initial PCR step to obtain only 16S rDNA gene sequences followed by pyrosequencing. This technique provides hundreds of thousands of reads, increasing the number of OTUs by at least one order of magnitude. The approach has been applied to marine samples a handful of times (Sogin et al. 2006, Huber et al. 2007, Galand et al. 2009, Gilbert et al. 2009, Kirchman et al. 2010), to the microbiota of humans (Claesson et al. 2009), termites (Engelbrektson et al. 2010), soils (Youssef & Elshahed 2009), and lakes (Xing et al. 2009). The approach still requires careful evaluation of possible biases and potential errors (Huse et al. 2007, 2008, 2010, Quince et al. 2009, Reeder & Knight 2009, Kunin et al. 2010), as well as studies in more ecosystems, but it has already started a whole new era in the study of microbial diversity. The approach by Sogin et al (2006) offers, for the first time, the opportunity to address questions about microbial diversity combining both a large number of individuals and their identity.

Pedrós-Alió (2006, 2007) suggested that rank abundance curves for bacteria (and likely also for protists) would show a decaying exponential shape. One implication of these distributions involved a long 'tail' of low abundant species. This could constitute a 'seed bank' of species that, should environmental conditions change, could grow and become abundant. Furthermore, conventional sequencing techniques would be only capable of retrieving the abundant taxa, while pure culture techniques would retrieve only a few of the taxa present, although they could be either abundant or rare. This long tail of low abundant taxa was coined 'the rare biosphere' (Sogin et al. 2006) and the pyrosequencing approach proposed in the latter paper provided an avenue to explore it. Here we analyze pyrosequencing data from a cruise in the NW Mediterranean Sea where we sampled a horizontal transect of about 120 km from coast to offshore and a vertical profile at the offshore station from the surface to the bottom (2000 m). The purpose was to test the effects of a significantly increased number of individuals analyzed on the rank abundance curves and diversity indices of marine bacterial assemblages, independently of taxonomy.

MATERIALS AND METHODS

Study area and sampling. We sampled a coast to offshore transect in the NW Mediterranean Sea in September 2007 (Fig. 1). The Blanes Bay Microbial Observatory (BBMO, www.icm.csic.es/bio/projects/icmicrobis/bbmo/), close to the shore, has been the site of a large variety of studies describing the microbial component of the marine plankton. Blanes Bay presents relatively oligotrophic conditions just 1 km from the
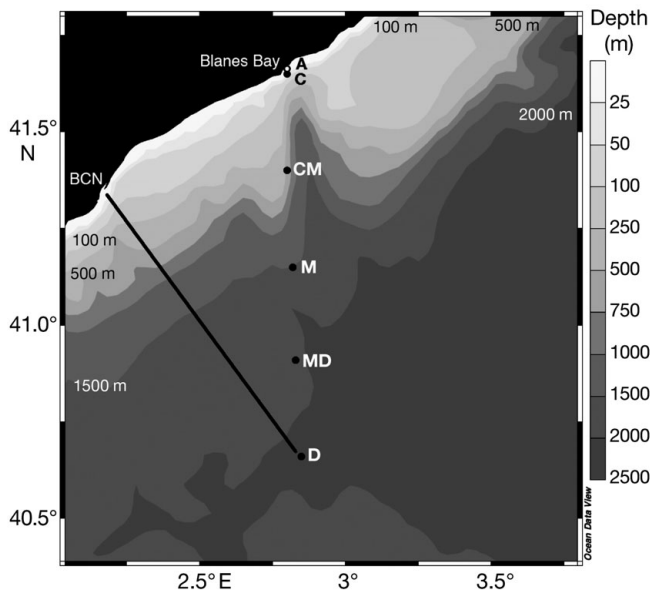
Fig. 1. Bathymetric map of the NW Mediterranean Sea showing the Blanes Bay Microbial Observatory (Stn A), and the stations sampled (A, C, CM, M, MD and D). The black line from Barcelona (BCN) to Stn D shows the transect sampled in past studies (see 'Materials and methods' and 'Discussion'). The Blanes submarine canyon can be seen to the right of Stn CM

shore, with moderate allochthonous contributions and certain oceanic marine water intrusions through the submarine canyon off Blanes (Font et al. 1988). We chose the BBMO station as the most coastal site of a transect offshore. The transect was in an area that has been repeatedly studied from different points of view: physical oceanography (Font et al. 1988, Salat 1996), chemistry (Doval et al. 2001), phytoplankton (Estrada & Salat 1989, Estrada et al. 1993, 1999), heterotrophic bacteria (Gasol et al. 1998, Pedrós-Alió et al. 1999), and viruses (Guixa-Boixereu et al. 1999). The transect studied here starts slightly north of the previously studied transect to accommodate the BBMO station and ends at the same open sea station, Stn D (Fig. 1). But both transects cross the same water masses and fronts.

Stn A was sampled at about 1 km offshore (41° 40′ N, 2° 48′ E) on September 20, 2007, and water was immediately filtered through a 200 µm mesh net. Seawater was kept in 25 l acid rinsed polycarbonate carboys and transported under dim light and cold conditions to the laboratory (less than 2 h). The rest of the transect was sampled on board the BO 'García del Cid' during the Modivus cruise (September 20 to 23, 2007). Samples were collected

with Niskin bottles mounted on a rosette with a CTD. Five stations were occupied and vertical profiles sampled at each station. Water was prefiltered through a 200 µm mesh and immediately processed on board. The samples chosen for pyrosequencing in the present study are shown in Table 1, together with date, position, depth, and chlorophyll *a* (chl *a*), determined as previously described (Pedrós-Alió et al. 1999). Bacterial abundance and size were determined by flow cytometry, and bacterial biomass was calculated from those 2 variables. Bacterial heterotrophic production was determined by leucine incorporation (T. Lefort & J. M. Gasol unpubl.).

Collection of community DNA. To collect microbial biomass, between 5 and 15 l of seawater were filtered through a 5 µm pore size Durapore filter (Millipore) and a 0.2 µm Sterivex filter (Durapore, Millipore) in succession using a peristaltic pump. The 0.2 µm Sterivex unit was filled with 1.8 ml of lysis buffer (40 mM EDTA, 50Mm Tris-HCl, 0.75 M sucrose) and stored at −70°C. DNA was extracted by a standard protocol using phenol/chloroform (details in Schauer et al. 2003).

Sequencing and noise removal. The V6 region of the 16S rRNA was amplified with bacterial universal primers and the amplicons were pyrosequenced with a 454 Life Sciences GS-FLX sequencer at the Josephine Paul Bay Center, Marine Biological Laboratory in Woods Hole, MA, USA. The protocols have been described in detail in Sogin et al. (2006), Huber et al. (2007), Huse et al. (2007), and on the Visualization and Analysis of Microbial Population Structures (VAMPS) webpage (http://vamps.mbl.edu/). For each read from the sequencer, the primer bases were trimmed from the beginning and the end, and low-quality sequences were removed in accordance with Huse et al. (2007).

Table 1. Sampling date, location, depth, and chlorophyll *a* (chl *a*) for the samples used for pyrosequencing. See Fig. 1 for station locations. Stn D samples are numbered according to the respective sampling depth. ND: not determined. Note that Stn D appears twice

Datasets/ sample	Sampling date (d/mo/yr)	Latitude, longitude	Sampling depth (m)	Chl *a* (µg l⁻¹)
Coastal to offshore surface transect				
A	20/09/07	41° 40′ 0.0″ N, 2° 48′ 0.0″ E	0.5	0.12
C	20/09/07	41° 39′ 5.6″ N, 2° 48′ 1.3″ E	5	0.09
CM	21/09/07	41° 24′ 5.9″ N, 2° 48′ 4.9″ E	5	0.07
M	21/09/07	41° 9′ 1.3″ N, 2° 49′ 3.3″ E	5	0.08
MD	22/09/07	40° 54′ 52.8″ N, 2° 50′ 43.8″ E	4	0.10
D5	23/09/07	40° 39′ 4.7″ N, 2° 51′ 1.6″ E	5	0.08
Depth profile				
D5	23/09/07	40° 39′ 4.7″ N, 2° 51′ 1.6″ E	5	0.08
D25	23/09/07	40° 39′ 4.7″ N, 2° 51′ 1.6″ E	25	0.09
D65	23/09/07	40° 39′ 4.7″ N, 2° 51′ 1.6″ E	65	0.41
D500	22/09/07	40° 39′ 20.4″ N, 2° 51′ 28″ E	500	ND
D2000	22/09/07	40° 39′ 20.4″ N, 2° 51′ 28″ E	2000	ND

Sequences were flagged as low quality (1) when they were <50 nucleotides in length, (2) when the start of the sequence did not exactly match a primer sequence, (3) when the sequences contained ambiguous nucleotide assignments with one or more Ns (unknown nucleotides), or (4) if the first 5 nucleotides of a tag sequence did not correspond to the expected 5 nucleotides run key (used to sort the pyrosequencing reads). In addition, the tags were subjected to the preclustering method mentioned in Huse et al. (2010). All sequences obtained for the present study are available at http://vamps.mbl.edu/ and have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under the accession number SRP001214.

Tag clustering and normalisation of sample sizes. The datasets of interest were defined as follows: the coastal to offshore transect included samples from Stns A, C, CM, M, MD, and D5; the depth profile included the samples D5, D25, D65, D500, and D2000, with the number indicating the depth (m). For each dataset, random normalization of sample sizes was carried out using the Daisy-Chopper tool (Gilbert et al. 2009), based on the smallest sample (i.e. sample A with 12 628 sequences for the coastal to offshore transect and sample D25 with 13 435 sequences for the depth profile). Each subsampled dataset was then clustered in OTUs of decreasing genetic distance using MOTHUR v.10.0 (Schloss et al. 2009) according the furthest neighbor-clustering algorithm. To build the clusters in MOTHUR, we used the aligned version (April 2010) of SILVA (Pruesse et al. 2007). The resulting alignment was automatically checked by MOTHUR by calculating the number of potentially misaligned bases using the secondary structure map available from the greengenes database (DeSantis et al. 2006). The output at each level for each distance level was then parsed to produce occurrence tables of each OTU in each sample at the different clustering levels. We focused our analyses on the unique and 0.03 levels of clustering to assess potential sequencing and/or PCR errors that may have occurred during the amplification and pyrosequencing steps.

Analysis of sequences and comparisons of communities. Rarefaction curves were calculated using the online rarefaction calculator software (www2.biology.ualberta.ca/jbrzusto/rarefact.php). Based on a Bray-Curtis distance matrix between all communities, hierarchical clusters for each dataset were performed according to the complete linkage method, with increasing number of OTUs from the 0.03 levels of clustering, using R (R Development Core Team 2008). We first clustered the 30 most abundant OTUs, then the 300 most abundant OTUs and finally the 3000 most abundant OTUs. Pairwise simple Mantel tests on the distance matrices between the samples were carried

out using the software zt (Bonnet & Van de Peer 2002) using 9999 permutations.

Calculation of richness, evenness and diversity. Richness (S) was computed as the total number of OTUs (at the 0.03 level) in each sample. The Shannon index (H') was used to determine diversity:

$$H' = -\Sigma\, p_i \ln(p_i) \tag{1}$$

where $p_i = N_i/N$, the number of individuals of species i divided by the total number of individuals in the sample (N). Finally, evenness was computed with the Pielou index:

$$J' = H'/H_{max}, \tag{2}$$

where H' is the Shannon index and H_{max} is the maximal possible Shannon index if all the species were equally abundant

$$H_{max} = -\Sigma\,(1/S)\ln(1/S) = \ln S \tag{3}$$

where S is the total number of OTUs (richness).

RESULTS

Sequencing and clustering

A total of 201 605 tag sequences of sufficient quality were obtained along the transect and the depth profile (Table 2). On average, about 20 160 tag sequences were obtained per sample. We normalized to the smallest sample size for the transect (sample A with 12 682 tags) and the profile (sample D25 with 13 435 tags) data sets separately. Preclustering the data into unique OTUs (in accordance with Huse et al. 2010) resulted in a 42% reduction in the number of OTUs (Table 2, columns 3 and 4). The proportion of singletons in this data set was 52% (Table 2, column 5). We further clustered the sequences at the 0.03 difference level (Table 2, column 6), which caused a modest 12% reduction in OTU numbers and in the proportion of singletons to 46% (Table 2, column 7). Despite the successive clustering steps, the number of OTUs remained within the same order of magnitude (overall reduction in OTUs of 50%, Table 2, column 8). The most numerous OTUs were represented by close to 2000 tags, and the average number of tags per OTU was approximately 15 (Table 2, column 9).
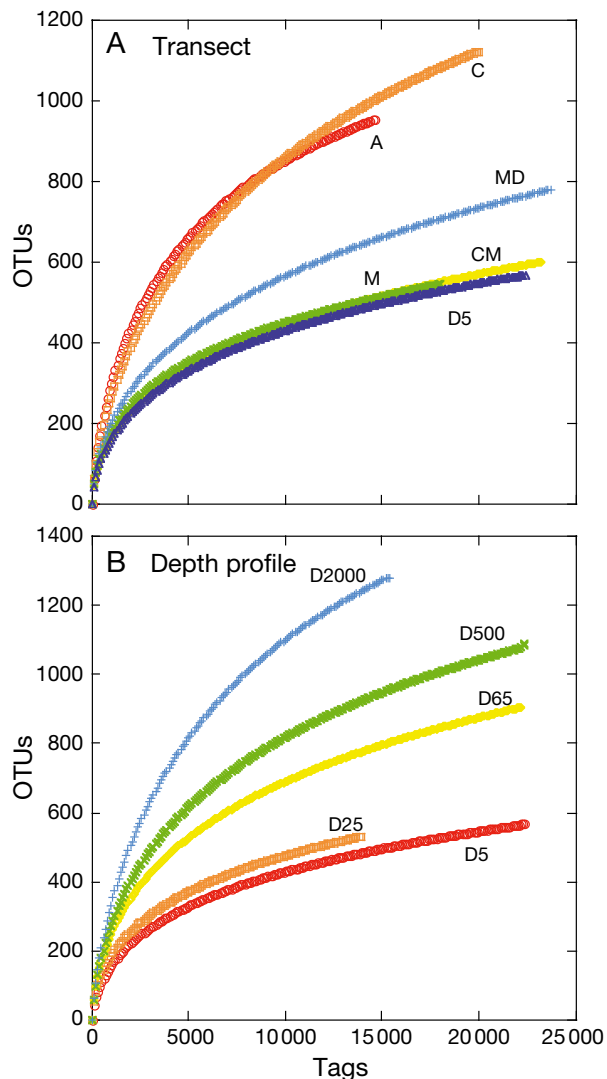
Rarefaction curves

Rarefaction curves were computed for the transect (Fig. 2A) and the vertical profile samples (Fig. 2B). These curves allow ranking of the samples in terms of OTU richness if the same number of tags had been obtained from all of them. In our datasets, the lowest

Table 2. Summary of total sequences, total operational taxonomic units (OTUs) and singletons in the 10 samples analyzed in the coastal to offshore transect and in the depth profile. Columns are numbered 1 to 10 to facilitate referencing. S_{Chao} is an estimator of the OTU richness in the sample

Station	1 Depth (m)	2 Tags	3 OTU (unique)[a]	4 OTU (unique)[b]	5 Singletons (%)	6 OTU (0.03)	7 Singletons (%)	8 Overall reduction	9 Tags per OTU	10 S_{Chao} index
A	0.5	14 775	2719	1808	57.1	1534	49.0	56.4	8.2	2659.0
C	5	20 118	3201	1823	65.3	1577	57.8	49.3	8.0	3193.4
CM	5	23 325	1452	654	47.1	599	41.6	41.3	21.1	980.2
M	5	18 067	1313	686	43.4	640	40.6	48.7	19.7	1010.0
MD	4	23 994	1820	880	49.1	799	44.8	43.9	15.8	1349.9
D	5	22 581	1349	679	50.2	632	46.7	46.8	20.0	1289.1
D	25	13 961	1139	777	43.5	677	39.3	59.4	19.8	1082.1
D	65	23 864	2020	1018	48.0	900	43.9	44.6	14.9	1613.9
D	500	24 690	2226	1215	48.2	1073	43.0	48.2	12.5	1637.0
D	2000	16 230	3383	2469	63.9	2065	55.8	61.0	6.5	4156.4
Total		201 605								
Average		20 161	2062	1201	51.6	1049.6	46.2	50.0	14.7	1897.1

[a]Unique OTUs without the preclustering step (see 'Materials and methods — Sequencing and noise removal'); [b]unique OTUs after the preclustering step (see Materials and methods — Sequencing and noise removal')

Fig. 2. (A) Rarefaction curves for surface samples along the transect. (B) Rarefaction curves for samples along the vertical profile. 0.03 difference level clusters were used. See Table 1 for details of samples

number of tags for a sample was around 13 000. The curves did not reach an asymptote. Thus, we estimated the total number of OTUs in our sequence collections by calculating the S_{Chao} index (Table 2, column 10). The average value was close to 1900 estimated OTUs per sample. The 2 coastal samples were clearly richer than the open sea samples along the transect, and richness increased with depth along the profile.

Patterns of richness and evenness

Richness and evenness were computed and were plotted against the value of the Shannon diversity index (Fig. 3A). Evenness and richness increased in parallel. They were strongly correlated ($r^2 = 0.876$, $p < 0.0001$). The evenness and richness indexes calculated for sample D5 in the transect and the vertical profile were not exactly equal as a result of different normalization factors. A group of 3 samples with very high richness and evenness included the 2 coastal samples from the transect and the deepest sample from the profile. Two samples at intermediate depths in the profile (65 and 500 m) showed intermediate values, and the remaining samples from the surface open sea and shallow depths of the profile formed a group of relatively low evenness and richness. Richness and evenness were plotted again in Fig. 3B to better show how they changed along both environmental gradients. Both variables were
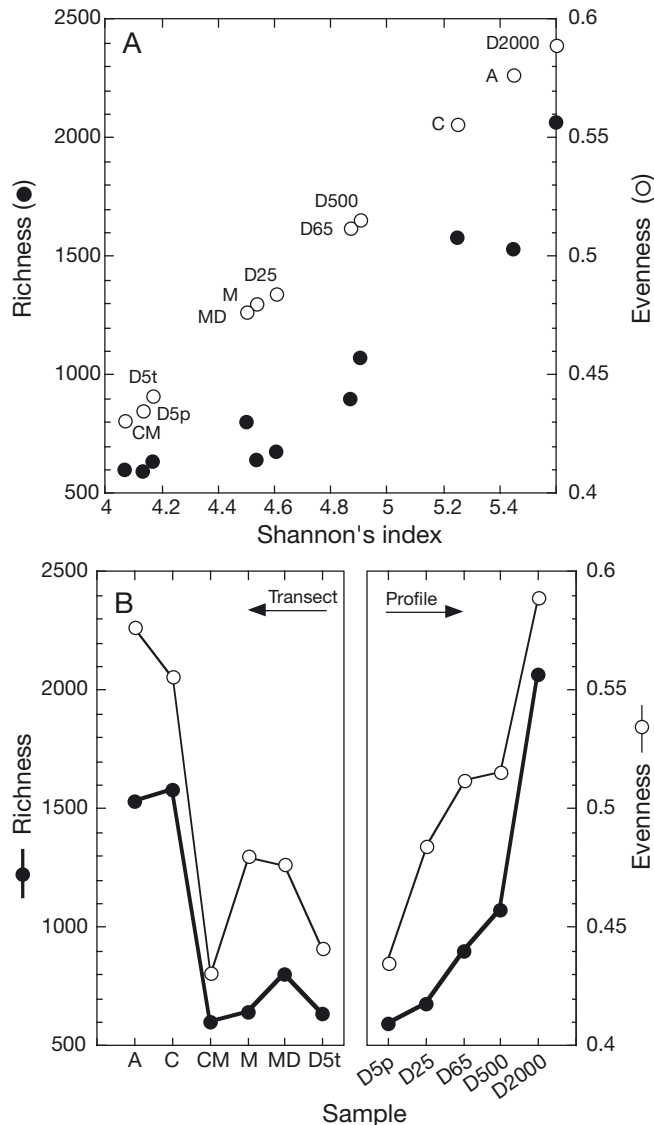
Fig. 3. (A) Richness and evenness plotted against Shannon's diversity index. (B) Richness and evenness for our transect and depth profile. Changes with distance to shore (Stns A to D5t) or with depth (Stns D5p to D2000) are shown. Sample D5 appears twice. The slight difference in values is due to the different number of tags for normalization of the transect (D5t) and the vertical profile (D5p). See Table 1 for details of stations

high in the 2 coastal stations and sharply decreased in the open sea stations. Stns M and MD showed slightly higher values than Stns CM and D. In the vertical profile, both variables increased with depth, although evenness increased at shallower depths than richness.

Rank-abundance curves

Rank-abundance curves for the transect and vertical profile samples were plotted for the 2 sets of samples (Fig. 4A,B). For clarity, the rank-abundance curves for all stations are also shown in log-log scales in Fig. 4C,D. The data set exhibited a very strong dominance by a few OTUs. The slope of the rank-abundance curves changed from one sample to the next. This change was progressive and significant both in the transect and in the depth profile. In the transect, the open sea sample (D5) showed more dominance and less number of OTUs, while the coastal sample (A) showed the opposite pattern with low dominance and higher richness. Similarly, in the vertical profile, the deepest sample (D2000) showed less dominance and more OTUs than the surface samples. This is in accordance with the values of evenness and richness calculated in the previous section.

Abundant and rare OTUs

The richness of OTUs at the 0.03 difference level that were found in different numbers of samples in the 2 datasets (i.e. coastal to offshore transect and depth profile) were plotted (Fig. 5). Remarkably, about 4000 OTUs (ca. 70%) were found in only 1 sample, in each of the 2 datasets, while only 250 (4%, transect) or 72 (1%, profile) OTUs were found in all the samples within each dataset.

Comparison among samples

A Non-Metric Multidimensional Scaling (NMDS) plot of the samples analysed based on the pairwise shared OTUs was constructed (Fig. 6). In the transect, the 2 coastal stations formed a cluster, while the open sea stations from the mixed layer formed a separate one. In the case of the vertical profile, the 2 samples from the surface photic layer clustered together, while samples from the other layers were different from each other.

To determine the number of abundant OTUs needed to group the different samples, we carried out a cluster analyses for the 2 datasets considering an increasing number of abundant OTUs (Fig. 7). In the case of the transect, the 30 most abundant OTUs were sufficient to cluster the samples in the same way as with 3000. In the case of the vertical profile, the 300 most abundant OTUs were necessary for the same outcome. When Mantel tests were carried out, all distance matrices turned out to be significantly correlated (p < 0.001).

Relationships with biomass and production

We did not find any relationship between our diversity indices and chl a. Concentrations along the transect were uniform, around 0.1 µg l^{-1}, but diversity ranged
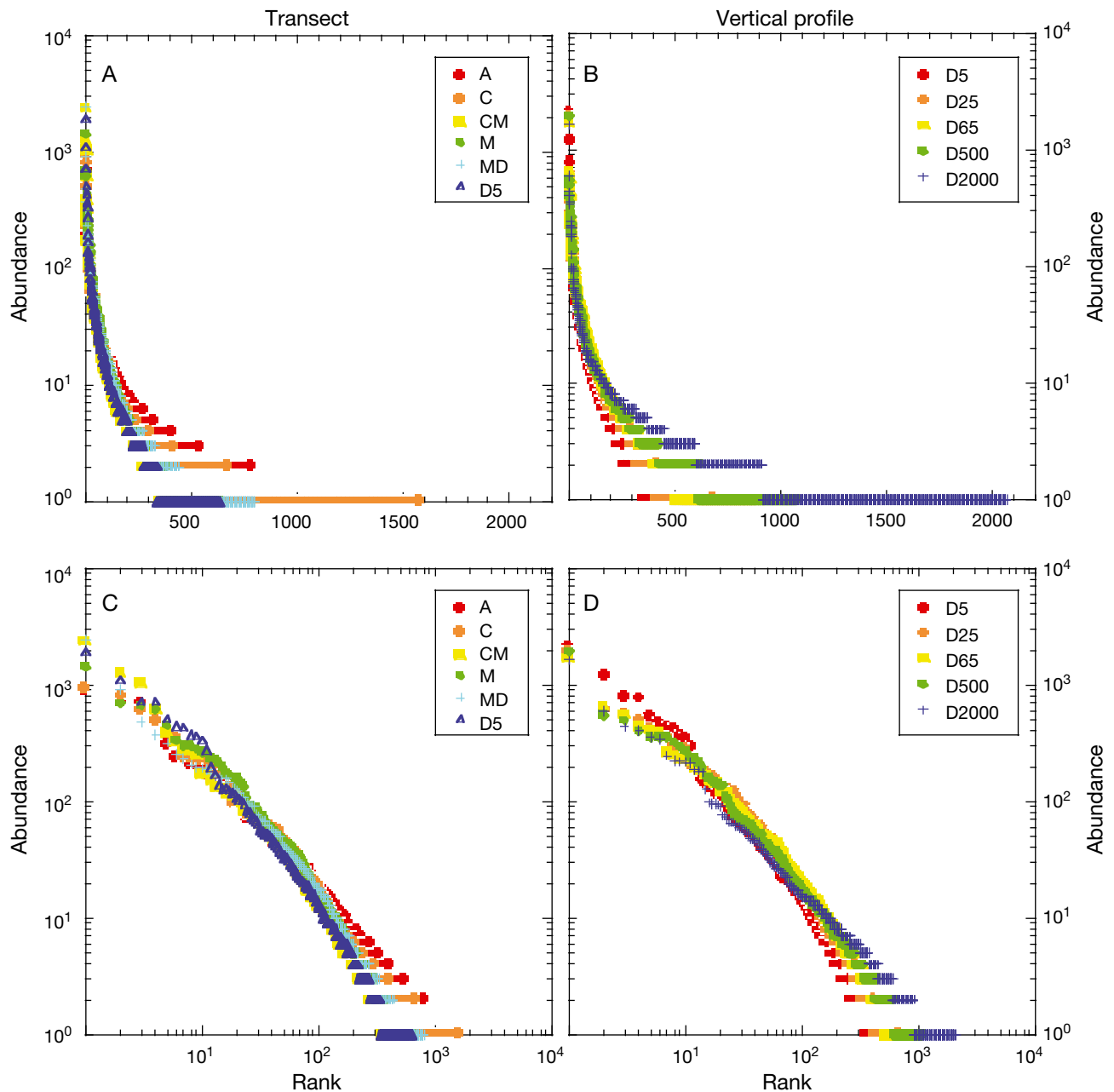
Fig. 4. Rank-abundance curves for the transect (A,C) and the vertical profile (B,D) samples. Panels A and B present the usual semilog plot, while plots C and D show log-log plots of the same data. Operational taxonomic units clustered at the 0.03 level were used. See Table 1 for details of samples

from some of the highest values at the coast to the lowest at Stn D. The vertical profile had similar values of chl *a* at the 2 surface samples, the highest concentration at the deep chlorophyll maximum (sample D65) and was zero at the 2 deepest samples. However, diversity increased from the surface to the bottom. In summary, there was no relationship between diversity and chl *a*.

The relationships of richness and evenness with bacterial biomass and heterotrophic production were also analyzed (Fig. 8). The 2 richest samples (C and D2000)

had higher production to biomass (P/B) ratios, but there was no clear relationship between richness and the P/B ratio for the other samples (Fig. 8A). Both richness and evennes were inversely correlated with bacterial biomass (Fig. 8B). Evenness showed a very good negative correlation with heterotrophic production ($r^2 = 0.781$, $p = 0.0036$ with 1 point excluded). Likewise, richness generally decreased with production: the relationship was not as good, and there was 1 clear outlier: sample C (Fig. 8C).
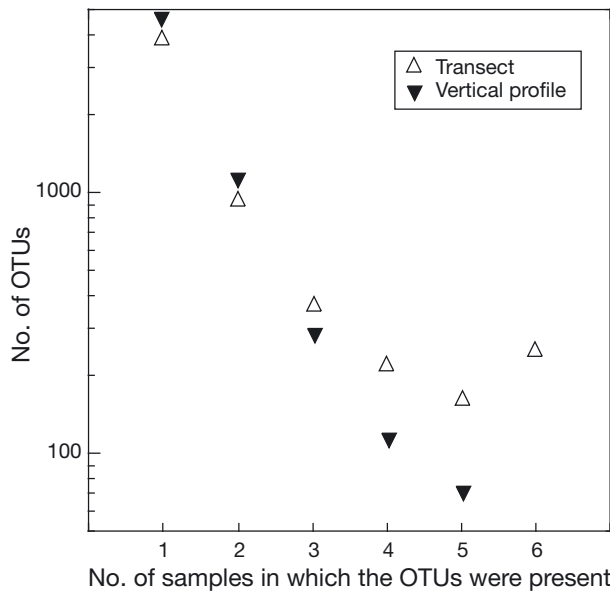
Fig. 5. Number of operational taxonomic units (OTUs) (at the 0.03 difference level) that were retrieved from 1, 2, or up to all the samples in the transect (open symbols) or the vertical profile (filled symbols)

Fig. 6. NMDS plot showing similarities among samples with Bray-Curtis distances

DISCUSSION

Diversity of microorganisms, including phytoplankton (Margalef 1995, Estrada 2008), ciliates (Dolan & Marrasé 1995, Dolan et al. 1999), bacteria (Acinas et al. 1997, 1999), and archaea (Galand et al. 2010), has been particularly well studied in the area of the pre-
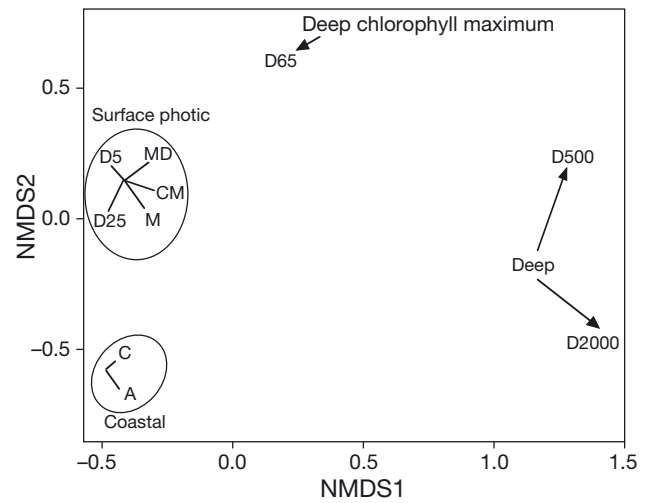
sent work (shown in Fig. 1). Thus, an objective of this study was to analyze whether the observation of more individuals confirmed, rejected, or somehow expanded what was known from conventional diversity studies in the Blanes Bay. The pyrosequencing technology offers the possibility to analyze at least one order of magnitude more individuals from a microbial community than is possible with conventional cloning and sequencing. Moreover, the number of individuals (sequences) analyzed is also much larger than what is usually analyzed in typical studies of diversity.
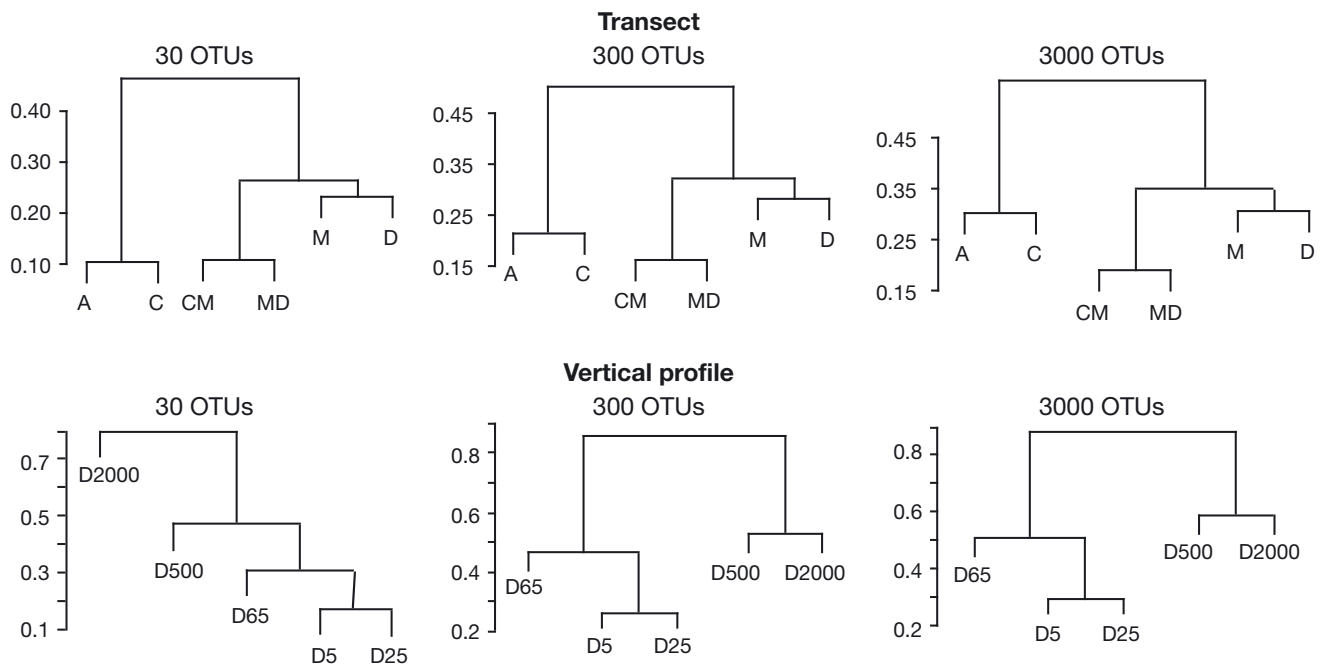

Fig. 7. Cluster analysis using increasing number of most abundant OTUs along the coastal to offshore transect (upper panels), and along the vertical profile (lower panels). *Y*-axis indicates height of clustering
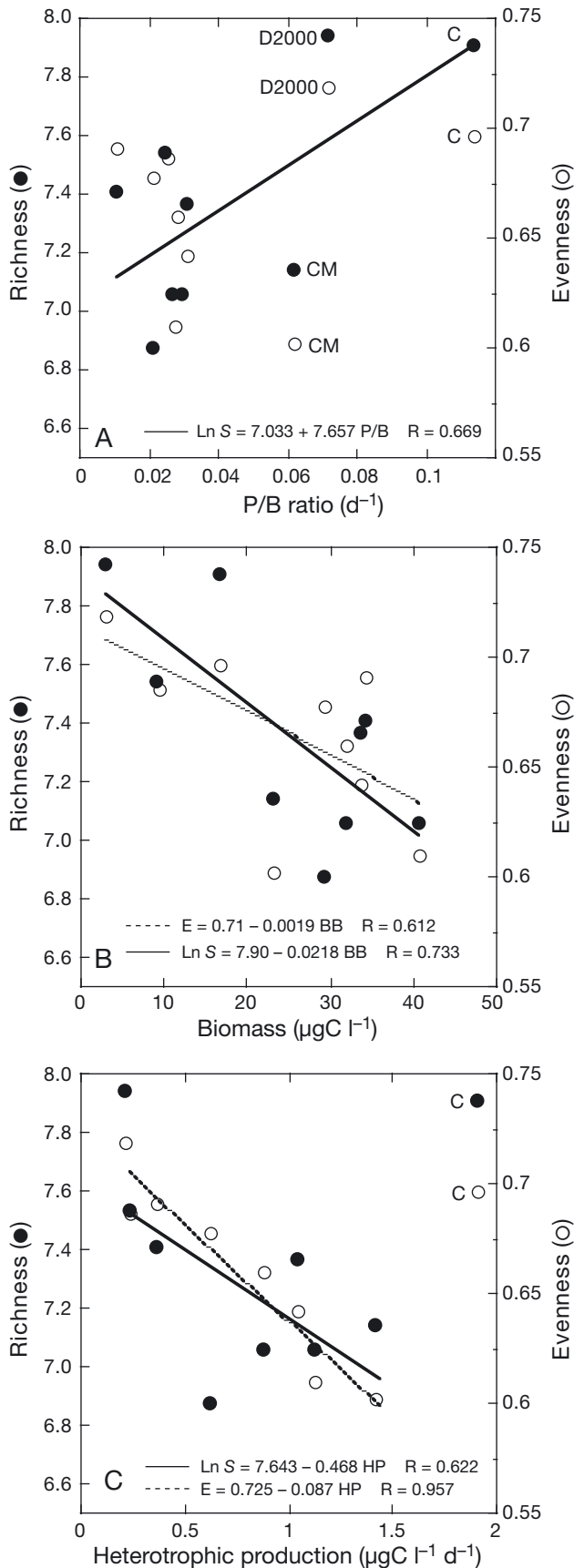
Fig. 8. Relationship between diversity indicators (richness and evenness) and (A) community production/biomass (P/B) ratio, (B) biomass and (C) heterotrophic production. Stations referred to in the text are indicated

Spatial patterns of diversity

Both the NMDS and the cluster analysis showed that surface samples formed 2 independent clusters: coastal and open sea. The 2 coastal stations had very high values of diversity. Both stations were over the continental shelf and were separated from the open sea stations by the shelf front. On the other side of the front, the open sea samples grouped together. They all had lower values of diversity. The D25 sample also clustered with this group. This was expected since it was within the mixed layer. The D65 sample, on the other hand, was in the deep chlorophyll maximum and, accordingly, did not join the surface cluster. However, it was more similar to the surface than to the deep samples. Finally, the deep samples clustered separately. The D2000 sample, in particular, had the highest diversity of all the samples. Both the NMDS plot and the dendrograms showed that differences among surface samples were smaller than among samples differing in depth. The observed clustering of samples is what could be expected from previous studies of microbial diversity. For example, Díez et al. (2004) found that diversity of picoeukaryotes changed more with depth than along surface transects in the Southern Ocean. They also observed that samples taken within the same water masses clustered together. Thus, similar assemblages could be found for 100s of kilometers as long as no fronts were crossed, but communities changed abruptly when an oceanographic front was crossed. The differences between bacterial communities on both sides of an oceanic front have also been observed in other marine regions (Pinhassi et al. 2003, Hewson et al. 2006). Along the 'standard transect' in the NW Mediterranean Sea, Acinas et al. (1997, 1999) found that their coastal station and Stn D were clearly distinct, and that the biggest differences were found between surface samples and deep samples. Similar results were found in an area of the NW Mediterranean Sea north of Blanes (Ghiglione et al. 2005). These patterns were the same as those that we found for bacteria in the present study: larger differences with depth than along the surface, and changes in composition when oceanographic fronts were crossed.

Díez et al. (2004) carried out their study with denaturing gradient gel electrophoresis (DGGE) and, therefore, with a very limited number of OTUs (around 30 per sample). The same was true for the study of Acinas et al. (1997, 1999) and the others mentioned above. Yet, the main conclusions were similar. In the present

study, we analyzed increasing numbers of OTUs, from 30 to 3000, to see how this would affect the clustering of samples. In the case of the coastal to offshore transect, the 30 most abundant OTUs were sufficient to obtain the same clustering as with 3000 OTUs. In the case of the vertical profile, the 300 most abundant OTUs were required for the same outcome. In addition, when a Mantel test was carried out, even the distance matrices giving the different clustering of samples were significantly correlated. This explains why the fingerprinting approaches were sufficient to cluster the samples appropriately despite the fact that they could not retrieve most of the richness in the samples. There is one caveat, however, since we do not know whether the same clustering would have been obtained with for example, 30 000 OTUs. This will have to wait for more intensive sequencing projects.

When assessing diversity using conventional cloning and sequencing, the number of OTUs is usually higher than with fingerprinting techniques. Kemp & Aller (2004) combined data from clone libraries of similar ecosystems and computed rarefaction curves (see their Fig. 7). The total number of clones examined ranged from 60 to 300 approximately. None of the curves reached a plateau. Calculation of the Chao 1 estimator of richness provided values under 300 OTUs in all cases. Similar numbers were found for clone libraries from the BBMO (Alonso-Sáez et al. 2007). Therefore, one could expect that examining around 13 000 individual sequences (as done in the present study) would allow reaching a plateau in the rarefaction curves at around 300 OTUs. Rarefaction curves, however, were not asymptotic. Further, the calculated Chao 1 indices had values ranging between 1000 and 4000 OTUs (one order of magnitude higher than those of conventional clone libraries for equivalent 97% similarity levels of clustering, Table 2, column 10). This indicates that compared to pyrosequencing, conventional clone libraries are often too small for good estimations of richness of bacterial assemblages. In fact, the shape of the rarefaction curves indicates that the actual richness was even higher than what was retrieved in the present effort. So that even estimates from sequencing 13 000 individuals are far from revealing the real richness of bacterial assemblages.

Evenness and richness indices

The samples with higher richness were also those with the highest evenness, although the exact relationship showed some variation along the coastal to offshore transect and the vertical profile. Different explanations have to be invoked to explain finding the highest diversity values in the coastal and deep samples, respectively. One notion is that stable ecosystems, such as tropical rainforests or coral reefs, tend to have higher diversity. The deep sample fits this description. On the other hand, higher diversity can be found at ecotones and this is what could be expected for the coastal stations, where land influence would interact with deep waters from the Blanes canyon potentially generating more ecological niches.

When standard diversity indices were computed, the values found were higher than those previously reported in the literature (Fig. 3A). We can compare them to those computed for tintinnid ciliates (Dolan & Marrasé 1995, Dolan et al. 1999) and phytoplankton (Estrada 2008) along a transect in the same area as that covered by the present study. The rank abundance curves of all these microorganisms have the same shape as those found here for bacteria. In the latter study, the deep chlorophyll maximum community was more species rich than the surface community. This is also what we found in the present study: the D65 sample from the Deep Chlorophyll Maximum (DCM) was richer than the D5 and D25 surface mixed layer samples. The values of the Shannon index found, however, were much lower than those found here. They ranged from less than 1.5 at the surface to 3.5 at the DCM, while ours were all above 5. Interestingly, the only study displaying higher diversity indices with values of the Shannon index over 7 was reported for marine viral communities using 1061 sequences from a viral shotgun genomic library (Breitbart et al. 2002).

Rank-abundance curves

Another point of interest is whether the shape of rank-abundance curves is an artifact of a low number of individuals sampled. While around 2000 individual tags represented the most abundant OTUs in our samples, most OTUs were doubletons or singletons (46%), which exceeded 1000 OTUs among the 10 samples. However, even if the OTUs represented by 1, 2, or even 3 tags were eliminated from the analysis to avoid potential artifacts, the shapes of the rank abundance curves would not change dramatically. The abundance of OTUs spanned 3 orders of magnitude. After examination of about 19 000 tags, the curves obtained did not resemble any of the established models (Fig. 4). Neither the broken-stick nor the normal distribution models fit our data. The geometric series would fit the first part of the curves, but it would miss the long tails completely. This lack of fit indicates that at least some of the assumptions on which the established models are based are not valid for bacteria.

The rank abundances curves presented here were very similar to those found for phages (Breitbart et al.

2002). These authors could fit a power-law function to their data. Furthermore, this pattern was confirmed afterwards in a new type of power law model modified from the version of Lotka-Volterra predator–prey dynamics to explain how phage–host interactions keep OTUs cycling through time and ranks (Hoffmann et al. 2007). It is intriguing that the same model should fit bacterial data, despite the fact that the relationship of the latter to their food source is completely different from that between phages and their hosts.

Diversity–productivity relationship

There has been considerable discussion about whether there are relationships between diversity and productivity (Rosenzweig 1995, Mittelbach et al. 2001, Drake et al. 2008). The relationship has been found to be linear and positive, linear and negative, U-shaped or hump-shaped depending on the organisms and systems studied. In such studies, chl *a* is often used as an estimator of productivity. We did not find any relationship between our diversity indices and chl *a* and, thus, presumably there was no clear relationship between diversity and primary production. In a study of several mesocosms with a gradient of productivity, Horner-Devine et al. (2003) did find a hump-shaped relationship between algal richness and chl *a* but no relationship between bacterial richness and chl *a*. However, these authors found different (and significant) relationships between the richness of certain bacterial groups and chl *a*. In our study, samples along the vertical profile were not directly comparable because the dark samples belonged to a different ecosystem than the surface samples and, therefore, there was no reason to expect a relationship. The transect samples, despite being comparable, did not show a relationship. A much larger sample set will have to be analyzed with the same methods in order to clarify this issue.

It has been proposed that the most diverse communities would have the lowest production/biomass (P/B) ratios (Margalef 1968). A mature forest, for example, holds a lot of biomass in the form of wood that does not contribute to production. A prairie, on the contrary, has very high production with little biomass. In this example, the diversity of plants in a mature forest would be larger than that of a prairie.

We compared our values of diversity with those of bacterial biomass and heterotrophic production determined in the same samples. The results did not show the expected relationship. Rather, the samples with highest diversity values also had the largest P/B ratios. In fact, this was due to the 2 most diverse samples: samples A and D2000. The other samples did not show any clear relationship to P/B. We decided to look at the

relationships with biomass and heterotrophic production separately. Both richness and evenness showed negative relationships with both biomass and production. The only exception was the coastal surface sample that had the highest production together with a very high diversity. Likely, the plant model is not appropriate for bacteria or microorganisms since, in the case of plants, most of the biomass is non productive. In bacterial assemblages, increasing production results in increased biomass (predation apart). Even though many of the bacteria may not be active at any particular time, bacteria do not accumulate large amounts of unproductive biomass, such as wood. Our rank abundance curves showed that higher production correlated with more dominance and, therefore, with lower richness and evenness.

CONCLUSIONS

This study of a marine area by the powerful pyrosequencing technology confirms that bacterial richness is very large, that rank-abundance curves show a very long tail of rare taxa, and that the signature composition of bacterial assemblages can be retrieved from the most abundant taxa, thus making comparison with less powerful techniques possible.

LITERATURE CITED

Acinas SG, Rodríguez-Valera F, Pedrós-Alió C (1997) Spatial and temporal variation in marine bacterioplankton diversity as shown by RFLP fingerprinting of PCR amplified 16S rDNA. FEMS Microbiol Ecol 24:27–40

Acinas SG, Antón J, Rodríguez-Valera F (1999) Diversity of free-living and attached bacteria in offshore western Mediterranean waters as depicted by analysis of genes encoding 16S rRNA. Appl Environ Microbiol 65:514–522

Acinas SG, Klepac-Ceraj V, Hunt DE, Pharino C, Ceraj I, Distel DL, Polz MF (2004) Fine-scale phylogenetic architecture of a complex bacterial community. Nature 430:551–554

Alonso-Sáez L, Balagué V, Sa EL, Sánchez O and others (2007) Seasonality in bacterial diversity in north-western Mediterranean coastal waters: assessment through clone

libraries, fingerprinting and FISH. FEMS Microbiol Ecol 60:98–112

Bonnet E, Van de Peer Y (2002) zt: A software tool for simple and partial Mantel tests. J Statistical Software 7:1–12

Breitbart M, Salamon P, Andresen B, Mahaffy JM and others (2002) Genomic analysis of uncultured marine viral communities. Proc Natl Acad Sci USA 99:14250–14255

Claesson MJ, O'Sullivan O, Wang Q, Nikkila J and others (2009) Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. PLoS ONE 4:e6669

Curtis TP, Sloan WT, Scannell JW (2002) Estimating prokaryotic diversity and its limits. Proc Natl Acad Sci USA 99: 10494–10499

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M and others (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072

Díez B, Massana R, Estrada M, Pedrós-Alió C (2004) Distribution of eukaryotic picoplankton assemblages across hydrographic fronts in the Southern Ocean, studied by denaturing gradient gel electrophoresis. Limnol Oceanogr 49:1022–1034

Dolan JR, Marrasé C (1995) Planktonic ciliate distribution relative to a deep chlorophyll maximum: Catalan Sea, NW Mediterranean, June 1993. Deep-Sea Res I 42: 1965–1987

Dolan JR, Vidussi F, Claustre H (1999) Planktonic ciliates in the Mediterranean Sea: longitudinal trends. Deep-Sea Res I 46:2025–2039

Doval MD, Álvarez-Salgado XA, Gasol JM, Lorenzo LM, Mirón I, Figueiras FG, Pedrós-Alió C (2001) Dissolved and suspended organic carbon in the Atlantic sector of the Southern Ocean. Stock dynamics in upper ocean waters. Mar Ecol Prog Ser 223:27–31

Drake JM, Cleland EE, Horner-Devine MC, Fleishman E and others (2008) Do non-native plant species affect the shape of productivity-diversity relationships? Am Midl Nat 159: 55–66

Dykhuizen DE (1998) Santa Rosalia revisited: Why are there so many species of bacteria? Antonie van Leeuwenhoek 73:25–33

Engelbrektson A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, Ochman H, Hugenholtz P (2010) Experimental factors affecting PCR-based estimates of microbial species richness and evenness. ISME J 4:642–647

Estrada M (2008) Explorations on phytoplankton diversity. An appreciation of Ramon Margalef's contributions. In: Valladares F, Camacho A, Elosegui A, Gracia C, Estrada M, Senar JC, Gili JM (eds) Unity in diversity—reflections on ecology after the legacy of Ramon Margalef. Fundación BBVA, Bilbao, p 191–222

Estrada M, Salat J (1989) Phytoplankton assemblages of deep and surface water layers in a Mediterranean frontal zone. Sci Mar 53:203–214

Estrada M, Marrasé C, Latasa M, Berdalet E, Delgado M, Riera T (1993) Variability of deep chlorophyll maximum characteristics in the Northwestern Mediterranean. Mar Ecol Prog Ser 92:289–300

Estrada M, Varela RA, Salat J, Cruzado A, Arias E (1999) Spatio-temporal variability of the winter phytoplankton distribution across the Catalan and North Balearic fronts (NW Mediterranean). J Plankton Res 21:1–20

Font J, Salat J, Tintoré J (1988) Permanent features of the circulation in the Catalan Sea. Oceanol Acta. Océanographie Pélagique Méditerranénne, 9: SP 51–57

Galand PE, Casamayor EO, Kirchman DL, Potvin M, Lovejoy C (2009) Unique archaeal assemblages in the Arctic Ocean unveiled by massively parallel tag sequencing. ISME J 7:860–869

Galand PE, Gutiérrez-Provecho C, Massana R, Gasol JM, Casamayor EO (2010) Interannual recurrence of archaeal assemblages in the coastal NW Mediterranean Sea (Blanes Bay Microbial Observatory). Limnol Oceanogr 55:2117–2125

Gasol JM, Doval MD, Pinhassi J, Calderón-Paz I, Guixa-Boixereu N, Vaqué D, Pedrós-Alió C (1998) Diel variations in bacterial heterotrophic activity and growth in the northwestern Mediterranean Sea. Mar Ecol Prog Ser 164:107–124

Ghiglione JF, Larcher M, Lebaron P (2005) Spatial and temporal scales of variation in bacterioplankton community structure in the NW Mediterranean Sea. Aquat Microb Ecol 40:229–240

Gilbert JA, Field D, Swift P, Newbold L and others (2009) The seasonal structure of microbial communities in the Western English Channel. Environ Microbiol 11:3132–3139

Guixa-Boixereu N, Vaqué D, Gasol JM, Pedrós-Alió C (1999) Distribution of viruses and their potential effect on bacterioplankton in an oligotrophic marine system. Aquat Microb Ecol 19:205–213

Hagström A, Pommier T, Rohwer F, Simu K, Stolte W, Svensson D, Zweifel UL (2002) Use of 16S ribosomal DNA for delineation of marine bacterioplankton species. Appl Environ Microbiol 68:3628–3633

Hewson I, Steele JA, Capone DG, Fuhrman JA (2006) Temporal and spatial scales of variation in bacterioplankton assemblages of oligotrophic surface waters. Mar Ecol Prog Ser 311:67–77

Hoffmann KH, Rodríguez-Brito B, Breitbart M, Bangor D and others (2007) Power law rank-abundance models for marine phage communities. FEMS Microbiol Lett 273: 224–228

Hong SH, Bunge J, Jeon SO, Epstein SS (2006) Predicting microbial species richness. Proc Natl Acad Sci USA 103: 117–122

Horner-Devine MC, Leibold MA, Smith VH, Bohannan BJM (2003) Bacterial diversity patterns along a gradient of primary productivity. Ecol Lett 6:613–622

Hubbell SP (2001) The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, NJ

Huber JA, Mark Welch D, Morrison HG, Huse SM, Neal PR, Butterfield DA, Sogin ML (2007) Microbial population structures in the deep marine biosphere. Science 318: 97–100

Huse SM, Huber JA, Morrison HG, Sogin ML, Mark Welch D (2007) Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol 8:R143 doi:10.1186/gb-2007-8-7-r143

Huse SM, Dethlefsen L, Huber JA, Mark Welch D, Relman DA, Sogin ML (2008) Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS Genet 4:e1000255

Huse SM, Dethlefsen L, Huber JA, Mark Welch D, Relman DA, Sogin ML (2010) Ironing the wrinkles in the rare biosphere. Environ Microbiol 12:1889–1898

Kemp PF, Aller YJ (2004) Bacterial diversity in aquatic and other environments: what 16S rDNA libraries can tell us. FEMS Microbiol Ecol 47:161–177

Kirchman DL, Cottrell MT, Lovejoy C (2010) The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. Environ Microbiol 12:1132–1143

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P (2010) Wrinkles in the rare biosphere: pyrosequencing errors lead to artificial inflation of diversity estimates. Environ Microbiol 12:118–123

MacArthur RH, Wilson EO (1967) The theory of island biogeography. Princeton University Press, Princeton, NJ

Magurran AE (2004) Measuring biological diversity. Blackwell Publishing, Malden, MA

Margalef R (1968) Perspectives in ecological theory. University of Chicago Press, Chicago, IL

Margalef R (1972) Homage to Evelyn Hutchinson, or why is there an upper limit to diversity. In: Deevey ES (ed) Growth by intussusception. Ecological essays in honor of G. E. Hutchinson. Trans Connecticut Acad Arts & Sci New Haven 44:211–235

Margalef R (1995) Fitoplancton del NW del Mediterráneo (Mar Catalán) en junio de 1993, y factores que condicionan su producción y distribución. Memorias de la Real Academia de Ciencias y Artes de Barcelona. 55:1–56

May RM (1988) How many species are there on Earth? Science 241:1441–1449

Mittelbach GG, Steiner CF, Scheiner SM, Gross KL and others (2001) What is the observed relationship between species richness and productivity? Ecology 82:2381–2396

Pedrós-Alió C (2006) Marine microbial diversity: Can it be determined? Trends Microbiol 14:257–263

Pedrós-Alió C (2007) Dipping into the rare biosphere. Science 315:192–193

Pedrós-Alió C, Calderón-Paz JI, Guixa-Boixereu N, Estrada M, Gasol JM (1999) Bacterioplankton and phytoplankton biomass and production during summer stratification in the northwestern Mediterranean Sea. Deep Sea Res I 46:985–1019

Pinhassi J, Winding A, Binnerup SJ, Zweifel UL, Riemann B, Hagström A (2003) Spatial variability in bacterioplankton community composition at the Skagerrak-Kattegat front. Mar Ecol Prog Ser 255:1–13

Pommier T, Canbäck L, Riemann L, Boström KH and others (2007) Global patterns of diversity and community structure in marine bacterioplankton. Mol Ecol 16:867–880

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

Quince C, Lanzen A, Curtis TP, Davenport RJ and others (2009) Noise and the accurate determination of microbial diversity from 454 pyrosequencing data. Nat Methods 6:639–641

R Development Core Team (2008) A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna (www.r-project.org)

Reeder J, Knight R (2009) The 'rare biosphere': a reality check. Nat Methods 6:636–637

Rosenzweig ML (1995) Species diversity in space and time. Cambridge University Press, Cambridge

Rusch DB, Halpern AL, Sutton G, Heidelberg KB and others (2007) The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biol 5:e77

Salat J (1996) Review of hydrographic environmental factors that may influence anchovy habitats in northwestern Mediterranean. Sci Mar 60(Suppl. 2):21–32

Schauer M, Balagué V, Pedrós-Alió C, Massana R (2003) Seasonal changes in the taxonomic composition of bacterioplankton in a coastal oligotrophic system. Aquat Microb Ecol 31:163–174

Schloss PD, Westcott SL, Ryabin T, Hall JR and others (2009) Introducing mothur: open source, platform-independent, community supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541

Sogin ML, Morrison HG, Huber JA, Welch DBM and others (2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. Proc Natl Acad Sci USA 103:12115–12120

Venter JC, Remington K, Heidelberg JF, Halpern AL and others (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304:66–74

Whittaker RH (1972) Evolution and measurement of species diversity. Taxon 21:213–251

Xing P, Hahn MW, Wu QL (2009) Low taxon richness of bacterioplankton in high-altitude lakes of the eastern tibetan plateau, with a predominance of bacteroidetes and *synechococcus* spp. Appl Environ Microbiol 75:7017–7025

Youssef NH, Elshahed MS (2009) Diversity rankings among bacterial lineages in soil. ISME J 3:305–313