

High local and global diversity of *Flavobacteria* in marine plankton

Cecilia Alonso,^{1,2*} Falk Warnecke,^{2,3} Rudolf Amann² and Jakob Pernthaler¹

¹Limnological Station Kilchberg, University of Zürich, Switzerland.

²Max Planck Institute for Marine Microbiology, Bremen, Germany.

³DOE Joint Genome Institute, Walnut Creek, CA, USA.

Summary

Members of the phylum *Bacteroidetes* are among the most abundant microbes in coastal marine waters, but it is unclear to which extent the diversity within this phylum is covered by currently available 16S rRNA gene sequence information. We, thus, obtained a comprehensive collection of sequence types affiliated with *Bacteroidetes* in coastal North Sea surface waters and we compared this local diversity with the available sequences of marine planktonic and other aquatic *Bacteroidetes*. Approximately 15% of > 600 clones from two libraries (August 2000, June 2001) were related to *Bacteroidetes*, specifically to the *Flavobacteria*. Local diversity appeared to be almost exhaustively sampled. However, the diversity of the two libraries virtually did not overlap, indicating a pronounced temporal variability of the planktonic *Flavobacteria* assemblage. The majority of sequence types represented novel phylogenetic lineages, adding 6–7% to the currently known genera and species of *Bacteroidetes* in marine waters. Different diversity estimators suggested that so far only approximately half of the global diversity of planktonic marine *Bacteroidetes* has been described. The data set moreover indicated that cultivation-independent techniques and isolation approaches have recovered almost equally sized and virtually non-overlapping fractions of the currently known diversity within this phylum. Interestingly, only 15% of genera of *Bacteroidetes* from various aquatic environments appear to occur in more than one habitat type.

Introduction

Bacteroidetes, also known as CFB (*Cytophaga-Flavobacteria-Bacteroides*) thrive in a variety of marine environments including coastal and offshore waters (Eilers *et al.*, 2001; Zubkov *et al.*, 2002; Kirchman *et al.*, 2003; DeLong *et al.*, 2006), sediments (Llobet-Brossa *et al.*, 1998; Musat *et al.*, 2006), hydrothermal vents (Kormas *et al.*, 2006) and polar regions (Brinkmeyer *et al.*, 2003; Abell and Bowman, 2005a). *Bacteroidetes* have been found both free living and attached to organic aggregates (DeLong *et al.*, 1993; Eilers *et al.*, 2001; Abell and Bowman, 2005b), and they can be associated with marine phytoplankton or animals (Webster *et al.*, 2001; Grossart *et al.*, 2005).

Bacteroidetes often constitute the most abundant group of bacteria in coastal pelagic habitats, as, e.g. revealed by fluorescence *in situ* hybridization (FISH) with rRNA-targeted probes and microscopic counts (Cottrell and Kirchman, 2000a; Eilers *et al.*, 2001). Various lines of evidence suggest that some members of this phylum might play an important part in the degradation of complex and polymeric organic matter (for a review see Kirchman, 2002). Moreover, *Bacteroidetes*, and in particular *Flavobacteria*, have been found in high abundances during natural and induced phytoplankton blooms and as primary colonizers of marine phytoplankton, suggesting a potential role as consumers of algae-derived metabolites (Simon *et al.*, 1999; Riemann *et al.*, 2000; O'Sullivan *et al.*, 2004; Pinhassi *et al.*, 2004; Grossart *et al.*, 2005).

Presently there is no comprehensive appreciation of the overall diversity of marine *Bacteroidetes*. Members of this phylum are often underrepresented in polymerase chain reaction (PCR)-generated libraries of bacterial 16S rRNA genes retrieved from marine habitats (Cottrell and Kirchman, 2000a; Eilers *et al.*, 2000). Thus, a considerable effort is required to obtain an inventory of *Bacteroidetes* sequence types that adequately covers their diversity at a particular sampling site. Such exhaustive collections of local diversity would be an essential prerequisite to, e.g. address issues related to microbial biogeography (Pommier *et al.*, 2005) or the underexplored 'rare biosphere' (Sogin *et al.*, 2006). In addition, Hagström and colleagues have observed that a large fraction of culturable representatives from this lineage showed little similarity to previously sequenced bacteria (Hagström *et al.*,

Received 1 September, 2006; accepted 17 December, 2006.
*For correspondence. E-mail calonso@limnol.unizh.ch; Tel. (+41) 1 716 1210; Fax (+41) 1 716 1225.

2000). So far, it is an unresolved issue whether the numerous isolates from this lineage are indeed adding to total diversity, or if they fall into those phylogenetic lineages (i.e. genera and species) that are retrieved by the direct analysis of bacterial 16S rRNA genes from marine habitats.

The aim of this study was to obtain a comprehensive collection of 16S rRNA gene sequences affiliated with *Bacteroidetes* from coastal North Sea bacterioplankton, and to analyse their phylogeny and contribution to global diversity in the context of all available almost complete sequences from this phylum. Using freely available bioinformatics tools, we attempted to: (i) assess which fraction of the total diversity of pelagic marine *Bacteroidetes* is presently known at various levels of sequence similarity; (ii) resolve whether either cultivation or cultivation-independent approaches alone could provide an unbiased view on the total diversity of marine *Bacteroidetes*; and (iii) analyse which lineages of *Bacteroidetes* from marine waters are also found in other aquatic habitats.

Results

Local diversity of planktonic marine Bacteroidetes in North Sea waters

Six hundred and thirty clones were screened in two clone libraries produced from samples of coastal North Sea surface waters (August 2000, June 2001). Ninety-seven clones carried *Bacteroidetes*-related inserts. Thirty enzymatic restriction patterns were discriminated and 55 complete sequences were obtained. Two sequences were discarded as chimeric. Thirty-five operational taxonomic units (OTUs) were distinguished at a similarity level of 99% (subsequently referred to as unique sequence types, see section 'Diversity analysis' in *Experimental procedures*). Twenty-five of these sequences were without neighbour with at least 99% similarity in the NCBI nucleotide database (Table 1). The coverage index of the June library was 74% and that of the August library was 91%. There was little overlap in the diversity of the sequence types from the two clone libraries, and only one unique sequence type was recovered in both months (Table 1). Even at the lowest analysed level of similarity (> 95%, OTUs subsequently referred to as genera) the two libraries shared only two of 26 groups.

All obtained sequences were affiliated to the class *Flavobacteria*, mainly to the family *Flavobacteriaceae* (Table 1, Fig. 1). With three exceptions the most closely related phylotypes of our sequences were other uncultured bacteria (Table 1). In the majority of cases these closest neighbours were recovered from marine or other saline aquatic environments (Table 1). Half of the monophyletic groups that contained our sequences (NS4, NS5,

NS8, NS10, NS11, NS12) were exclusively constituted of clone sequences of marine or estuarine pelagic habitats (Fig. 1). Most of the other groups were also formed by *Flavobacteria* from aquatic habitats (for which no more precise description was available) (NS3a, NS6, NS9). Group NS7 included sequences from marine, freshwater and sediment origin, and representatives from the AGG58 cluster (O'Sullivan *et al.*, 2004). Group NS2 and group NS3b included isolates from marine sponges (*Winodgraskyella*) or dinoflagellates (*Aequorivita*) respectively. Finally, group NS1 contained sequences of isolates and uncultured bacteria from very diverse origin: from dental caries to hydrothermal vents, also including isolates of the genus *Chryseobacterium*.

Most of the sequences that were retrieved in this study have not been previously reported from marine environments (Table 2). In addition, the majority of our sequences represented novel groups at the three analysed levels of sequence similarity, adding approximately 5–7% to the known number of OTUs with similarities > 95%, > 97% and > 99% from marine pelagic environments (Table 2).

Taxonomic classification and phylogeny of planktonic marine Bacteroidetes

Altogether 798 almost complete sequences originating from the Ribosomal Database Project (RDP) database and from this study were unambiguously identified as planktonic marine *Bacteroidetes*. We performed taxonomic classification using the RDP II 'Hierarchy Browser' and 'Classifier' functions. Seventy per cent of the sequences belonged to the class *Flavobacteria*, mainly to the family *Flavobacteriaceae* (Fig. 2). Twenty-nine of the 43 known genera from this class harboured representatives of planktonic marine *Bacteroidetes*. Twenty of these genera were primarily formed by sequences from pelagic marine habitats (e.g. *Croceibacter*, *Polaribacter*, *Psychroserpens*, *Zobellia*). The *Sphingobacteria* formed the second most abundant class of planktonic marine *Bacteroidetes* (20% of the sequences) (Fig. 2). Only few genera within the *Sphingobacteria* – predominantly from the family *Flexibacteraceae* – harboured sequences of marine origin (e.g. *Cytophaga*, *Aquimarina*, *Saprospira*). The remaining sequences belonged to the class *Bacteroidetes* and typically could not be assigned to any described genus, with the exception of *Anaerophaga* (Fig. 2).

Diversity of cultured and uncultured marine Bacteroidetes

More than half of the genera of planktonic marine *Bacteroidetes* were defined by a single sequence (singletons), and three quarters by less than three sequences

Table 1. Unique 16S rRNA gene sequences types (sequence similarity \geq 99%) recovered during this study, their taxonomic affiliation according to the RDP classifier system, and their closest neighbours in the GenBank database.

| Accession numbers | Taxonomic affiliation | Month | Origin | Closest neighbour | |
|--|--------------------------|-----------------|------------------------------|--------------------------|--|
| | | | | % of sequence similarity | Clone name and accession number |
| AM279163 AM279166 AM279167 AM279173 AM279177 AM279206 AM279180 | <i>Flavobacteriaceae</i> | August | North Sea | 99% | ZD0403 AJ400347 |
| AM279178 | <i>Flavobacteriaceae</i> | August | Atlantic Ocean | 99% | Arctic96B-24 AF354616 |
| AM279179 | <i>Flavobacteriaceae</i> | August | Pacific coast | 99% | ZA2515c AF382110 |
| AM279165 AM279169 AM279171 AM279182 AM279204 | <i>Flavobacteriaceae</i> | August/ June | Pacific coast | 94% | SPOTSAPR01_5m210 DQ009083 SPOTSAPR01_5m235 DQ009115 |
| AM279188 | <i>Flavobacteriaceae</i> | August | Antarctic sponge | 98% | K126 AY321388 |
| AM279188 | <i>Flavobacteriaceae</i> | June | Temperate river | 99% | PRD18D09 AY948029 |
| AM279161 | <i>Flavobacteriaceae</i> | August | Antarctic sponge | 98% | K126 AY321388 |
| AM279197 | <i>Flavobacteriaceae</i> | August | Antarctic sponge | 98% | K126 AY321388 |
| AM279172 AM279183 AM279199 | <i>Flavobacteriaceae</i> | August | Deep sea octacoral | 98% | ctg_CGCOA79 DQ395552 |
| AM279211 | <i>Cryomorphaceae</i> | August | Pacific coast | 99% | SPOTSAPR01_5m143 DQ009091 |
| AM279164 AM279186 AM279195 AM279213 AM279181 | <i>Flavobacteriaceae</i> | August | Arctic Ocean | 94% | Arctic97A-14 AF355051 |
| AM279164 AM279186 AM279195 AM279213 AM279181 | <i>Flavobacteriaceae</i> | August | Pacific coast | 99% | SPOTSAPR01_5m210 DQ009083 |
| AM279195 AM279213 AM279181 | <i>Cryomorphaceae</i> | June | Uranium-contaminated aquifer | 91% | 1013-28-CG47 AY532584 |
| AM279176 AM279200 AM279189 AM279194 AM279209 | <i>Flavobacteriaceae</i> | August | Deep sea octacoral | 98% | ctg_CGCOA79 DQ395552 |
| AM279176 AM279200 AM279189 AM279194 AM279209 | <i>Flavobacteriaceae</i> | August | Deep sea octacoral | 99% | ctg_CGOF307 DQ395894 |
| AM279189 AM279194 AM279209 | <i>Cryomorphaceae</i> | June | Mono Lake | 94% | ML617.5J-5 AF507867 |
| AM279170 AM279191 AM279193 | <i>Flavobacteriaceae</i> | June | Arctic Ocean | 98% | Arctic97A-14 AF355051 |
| AM279193 | <i>Flavobacteriaceae</i> | June | <i>Gymnodinium catenatum</i> | 95% | Bacterium DG945 AY258123 |
| AM279187 AM279174 AM279162 AM279201 | <i>Flavobacteriaceae</i> | August | Pacific coast | 96% | SPOTSAPR01_5m55 DQ009099 |
| AM279176 AM279200 AM279189 AM279194 AM279209 | <i>Flavobacteriaceae</i> | August | Deep sea hydrothermal vent | 98% | CH4_1_BAC_16SrRNA_9N_EPR AY672523 |
| AM279192 | <i>Flavobacteriaceae</i> | June | Delaware estuary | 99% | 1D10 AY274838 |
| AM279190 | <i>Cryomorphaceae</i> | June | Uranium-contaminated aquifer | 92% | 1013-28-CG47 AY532584 |
| AM279175 AM279210 AM279184 | <i>Flavobacteriaceae</i> | June | <i>Gymnodinium catenatum</i> | 95% | Bacterium DG945 AY258123 |
| AM279184 | <i>Flavobacteriaceae</i> | August | Pacific coast | 94% | SPOTSAPR01_5m235 DQ009115 |
| AM279185 | <i>Flavobacteriaceae</i> | August | Atlantic Ocean | 91% | ZA3907c AF382130 |
| AM279196 | <i>Flexibacteriaceae</i> | June | Atlantic coast | 94% | PB1.7 DQ071085 |

Table 1. cont.

| Accession numbers | Taxonomic affiliation | Month | Origin | Closest neighbour | |
|-------------------|--------------------------|--------|----------------------|--------------------------|---|
| | | | | % of sequence similarity | Clone name and accession number |
| AM279205 | <i>Flavobacteriaceae</i> | August | Pacific coast | 98% | SPOTSAPR01_5m232 DQ009081 |
| AM279198 | <i>Flavobacteriaceae</i> | August | Pacific coast | 99% | SPOTSAPR01_5m123 DQ009101 |
| AM279207 | <i>Flavobacteriaceae</i> | August | Arctic | 95% | <i>Winogradskyella thalassocola</i> AY771731 |
| AM279203 | <i>Flavobacteriaceae</i> | August | Lake Kauhako, Hawaii | 93% | K2-30-6 AY344418 |
| AM279208 | <i>Flavobacteriaceae</i> | June | Pacific coast | 91% | 1-13 AY094494 |
| AM279212 | <i>Flavobacteriaceae</i> | June | Pacific coast | 90% | SPOTSAPR01_5m145 DQ009104 |
| AM279168 | <i>Flavobacteriaceae</i> | August | Pacific coast | 99% | SPOTSAPR01_5m210 DQ009083 |
| AM279202 | <i>Flavobacteriaceae</i> | June | Atlantic coast | 98% | PB2.10 DQ071089 |

Clones are listed according to enzymatic restriction patterns.

(Table 3). Four hundred and thirty-eight of the 798 planktonic marine *Bacteroidetes* sequences originated from isolates and 360 from uncultured organisms. The two categories formed clearly separated subsets of sequence types, and only 15 of the 321 operationally defined genera harboured sequences from both cultured and uncultured *Bacteroidetes*. Accordingly, the Bray–Curtis, Chao–Jaccard and Chao–Sorensen estimators indicated low similarity (0.06–0.18) of cultured and uncultured *Bacteroidetes* genera.

Rarefaction analyses of currently available sequences from cultured and uncultured planktonic marine *Bacteroidetes* suggested that the known diversity of sequences derived from clones was significantly higher at similarity levels of 95% and 97% (Fig. 3). By contrast, both groups were equally diverse at the level of unique sequence types (99% similarity). In addition, rarefaction analyses also indicated that the isolation of *Bacteroidetes* from new genera was less probable than the retrieval of sequences from new uncultured genera. Both non-

Table 2. Clustering of 16S rRNA sequence types from coastal North Sea waters samples at different levels of similarity and their contribution to all pelagic marine sequences > 1200 base pairs that are available in the RDP database (release March 2006).

| Number of sequences | Number of groups | | |
|---------------------------------|------------------|----------------|----------------|
| | 95% similarity | 97% similarity | 99% similarity |
| Total | 53 | 26 | 35 |
| Novel | 41 | 21 | 30 |
| Contribution to known diversity | 5.4% | 6.9% | 6.0% |

The term 'Novel' refers to sequence types or groups that could not be found or defined from sequence information available in the RDP database.

Table 3. Observed and predicted diversity of all 16S rRNA sequences > 1200 base pairs of pelagic marine *Bacteroidetes* available in the RDP database (release March 2006) and categorized at different levels of similarity.

| Similarity | Category | Observed | | | Predicted | |
|------------|------------|----------|------------|------------|-----------|-------|
| | | Groups | Singletons | Doubletons | ACE | Chao1 |
| 95% | Isolates | 149 | 90 | 17 | 372 | 387 |
| | Uncultured | 187 | 127 | 24 | 507 | 523 |
| 97% | Isolates | 193 | 129 | 20 | 586 | 609 |
| | Uncultured | 211 | 154 | 26 | 647 | 667 |
| 99% | Isolates | 289 | 221 | 30 | 1073 | 1103 |
| | Uncultured | 248 | 199 | 26 | 978 | 1010 |

Sequences produced during this study are included in the data set.

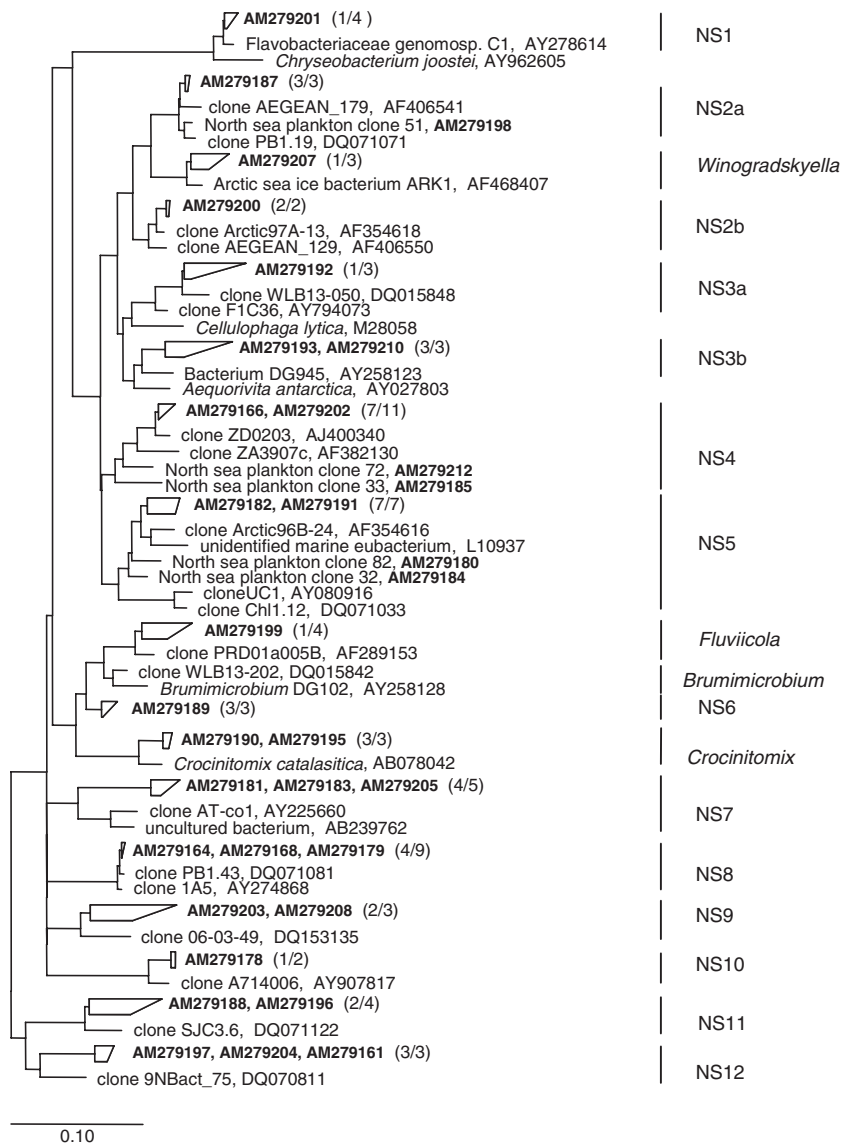


Fig. 1. Phylogenetic relationships of almost complete 16S rRNA gene sequences of *Bacteroidetes* from coastal North Sea bacterioplankton. Accession numbers are given for one representative per unique sequence type. The numbers in parenthesis indicate the contribution of sequences from this study to all sequences in the clades that are shown like wedges. Bar: 10% of estimated sequence divergence.

parametric diversity estimators (ACE and Chao1) predicted similar total numbers of genera, species and unique sequence types of cultured and uncultured *Bacteroidetes* (Table 3).

A closer look at the clustering patterns at 95% similarity revealed that cultured and uncultured *Bacteroidetes* from the marine plankton also differed in their respective distribution of the numbers of sequences per genus. Seventy-five per cent of the cultured genera of *Bacteroidetes* contained up to three sequences, there were substantially more genera containing between seven and 12 sequences, and the maximal number of phylotypes within a genus was 34. By contrast, 75% of uncultured genera harboured two or less sequences and the highest frequency of sequences within a single genus was 13 (Table 3).

Diversity of Bacteroidetes from various aquatic environments

From the > 11 000 *Bacteroidetes* sequences imported from RDP II, 1377 could be assigned to one of the habitat categories specified in the *Experimental procedures* section. Sixty-four per cent of these sequences were affiliated with the class *Flavobacteria*, 23% with the class *Sphingobacteria* and 10% with the class *Bacteroidetes*. The remaining 3% could not be assigned to any of the three classes.

The categorized data set indicated high habitat specificity of *Bacteroidetes* at the level of operationally defined genera (Fig. 4, filled bars). Only 15% of the genera contained sequences from two or more different habitat types. Interestingly, these genera harboured almost half of

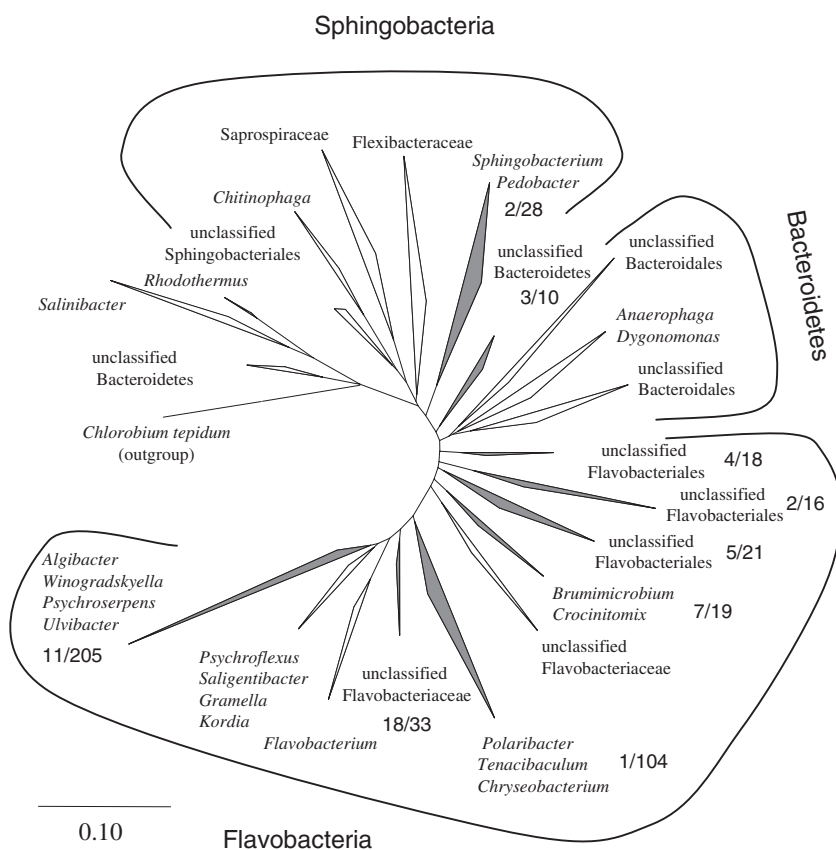


Fig. 2. Distribution of 798 almost complete 16S rRNA gene sequences of *Bacteroidetes* from marine bacterioplankton across different phylogenetic lineages. Clades that contain sequences from this study are shaded in grey. The numbers indicate the contribution of sequences from this study to all sequences in the various clades.

all available sequences of aquatic *Bacteroidetes*, mostly originating from isolates (Fig. 4, open bars). Of the marine planktonic genera that also occurred in other habitats, 23 were also found in freshwaters, 17 in sediments, and 11 and 8 genera were associated with algae and aquatic animals respectively. Only two genera featured sequences from four different origins (freshwater, marine plankton, sediments and associated with aquatic animals). These genera contained 10% of all environmental *Bacteroidetes* sequences. Specifically they harboured numerous isolates identified as *Flavobacterium* and *Chryseobacterium*.

Genera of environmental *Bacteroidetes* containing > 10 sequences were taxonomically categorized by the RDP II classification system and sequences within these genera were split according to their habitat type and origin (cultured versus uncultured) (Table 4). Three additional subcategories were introduced for this comparison: coastal, estuaries and sea ice. Twenty-five per cent of all sequences of environmental *Bacteroidetes* were distributed across 15 genera (Table 4). Typically, the majority of sequences within these genera originated from isolates. Only two genera (harbouring sequences from the clusters NS4 and NS8, respectively; see also Fig. 1) were entirely composed of uncultured *Bacteroidetes*. Forty–60% of the

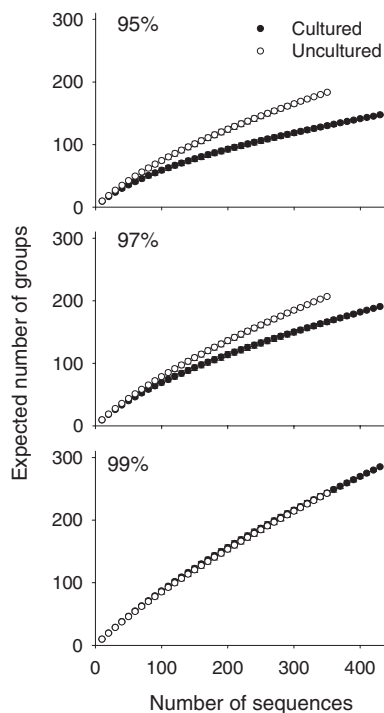


Fig. 3. Rarefaction analysis of cultured and uncultured groups of *Bacteroidetes* at different levels of sequence similarity. Note that the ranges of the 95% confidence intervals are smaller than the size of the symbols.

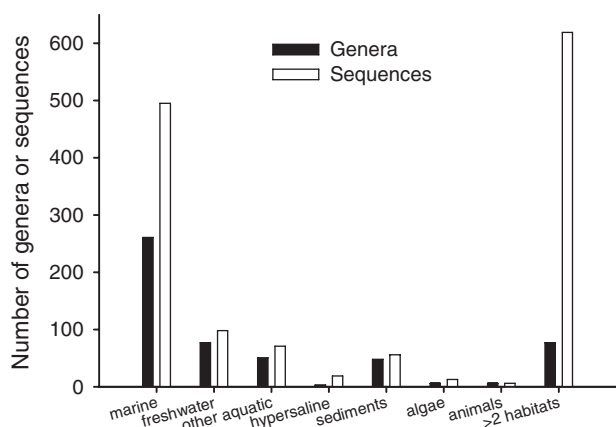


Fig. 4. Abundance of *Bacteroidetes* genera in different aquatic habitats (filled bars) and numbers of individual sequences that are affiliated with these genera (open bars). The following categories are discriminated: marine (marine planktonic), freshwater (freshwater planktonic), other aquatic (not possible to define the nature of the water body), hypersaline environments, sediments (includes both freshwater and marine), algae and animals (sequences retrieved from microbes associated with these aquatic eukaryotes).

sequences in these uncultured genera were produced in our study.

Discussion

Representation of Bacteroidetes in clone libraries

The relative proportion of *Bacteroidetes* sequences in our PCR-derived 16S rRNA gene clone libraries (August 2000: 15.3%; June 2001: 15.5%) lies at the upper end of values typically reported from coastal environments [e.g. Delaware Estuary: 5–13% (Cottrell and Kirchman, 2000b; Kirchman *et al.*, 2003; Cottrell *et al.*, 2005); German Wadden Sea: 8% (Stevens *et al.*, 2005)]. Extraordinary high proportions of *Bacteroidetes* sequence types (66% of all clones) have been reported from a coastal environment over the course of a phytoplankton bloom (O'Sullivan *et al.*, 2004). The fractions of *Bacteroidetes* at station Helgoland Roads were 39% (24 August 2000) and 20% (14 June 2001) of total bacterial cell counts, as determined by FISH with probes CF319a and EUB I-III (all *Bacteria*) (Daims *et al.*, 1999) (J. Pernthaler and C. Beardsley, unpubl. data). Our results thus confirm that *Bacteroidetes* tend to be underrepresented in 16S rRNA gene clone libraries from marine waters, e.g. compared with some lineages of *Proteobacteria* (Cottrell and Kirchman, 2000a). A previous clone library from our sampling site did not contain any *Bacteroidetes*-related sequences, although representatives from this lineage were isolated from the same sample and 20–30% of total bacterioplankton cells were *Bacteroidetes* according to FISH counts (Eilers *et al.*, 2000; 2001). The reverse primer used in that

Table 4. Groups of aquatic *Bacteroidetes* that contain more than 10 sequences > 1200 base pairs, and distribution of sequences within these groups across different aquatic habitats.

| | Flavobacterium | Chryseobacterium | Psychroserpens | Polaribacter | Arenibacter | Ulvibacter | Maribacter | Algoriphagus | NSB | Tenacibaculum | Gelidbacter | NS4 | Cellulophaga | Salinibacter |
|-------------------|----------------|------------------|----------------|--------------|-------------|------------|------------|--------------|--------|---------------|-------------|--------|--------------|--------------|
| Marine | 29 (27) | 1 (1) | 13 (13) | 27 (23) | 7 (7) | 6 (0) | 28 (24) | 13 (13) | 7 (0) | 5 (5) | 5 (5) | 5 (0) | 8 (8) | |
| Coastal | 1 (1) | | | 1 (1) | | | 1 (1) | | 4 (0) | | | 7 (0) | 3 (3) | |
| Estuaries | 4 (4) | | | | 1 (1) | | | | 2 (0) | | | | | |
| Freshwater | 77 (67) | 2 (2) | | | | | | 4 (13) | | | 1 (1) | | | |
| Undefined aquatic | 4 (3) | | | | | 2 (0) | | | | | | | | |
| Algae | | | 1 (1) | 1 (1) | 3 (3) | 1 (1) | 15 (15) | 5 (13) | | 1 (1) | | | | |
| Animals | 6 (5) | 20 (20) | 1 (1) | | | | | | | 6 (6) | | | | |
| Sediments | 6 (6) | 1 (0) | 3 (0) | | 1 (1) | | | | | | 5 (5) | | | |
| Sea ice | | | 1 (0) | 1 (0) | | 2 (0) | | | | | | | | |
| Saline | | | 19 (15) | 30 (25) | 12 (12) | 11 (1) | 44 (40) | 22 (22) | 13 (0) | 12 (12) | 11 (11) | 12 (0) | 11 (11) | 11 (11) |
| Total | 127 (113) | 28 (24) | | | | | | | | | | | | |

The groups were defined by a sequence similarity $\geq 95\%$ (genera). The numbers of isolated representatives are given in parenthesis. The last row shows the total number of sequences within each group.

study (1542R) targets less than a quarter of *Bacteroidetes* sequence types in the current RDP database. By contrast, the GM3–GM4 primer set (this study) theoretically should allow similar recovery of *Bacteroidetes* and *Proteobacteria*: GM3 and GM4 match to 75% and 90% of *Bacteroidetes* sequences in the RDP database that include the primer target region and to 73% and 91% of proteobacterial sequences. Thus, the underrepresentation of *Bacteroidetes* in clone libraries generated with this primer set must be related to other aspects of DNA recovery and/or PCR (von Wintzingerode *et al.*, 1997).

In order to investigate if there could be a systematic bias of diversity obtained by PCR-based and by PCR-independent molecular approaches we attempted to compare our data with *Bacteroidetes* sequences in metagenomic libraries of coastal and open ocean waters (Venter *et al.*, 2004; Cottrell *et al.*, 2005; DeLong *et al.*, 2006). However, no systematic conclusion could be reached because of limitations of the available data sets: only three almost complete (> 1200 base pairs) 16S rRNA gene sequences of *Bacteroidetes* were found in the data set of Venter and colleagues (Venter *et al.*, 2004), one such long sequence was present in the data set of DeLong and colleagues (DeLong *et al.*, 2006), and none in the library of Cottrell and colleagues (Cottrell *et al.*, 2005).

The high coverage indices of both clone libraries (74% and 91%) suggest that our strategy of cloning with general bacterial primers and subsequent PCR screening was appropriate to obtain a comprehensive set of *Bacteroidetes* sequence types present in coastal North Sea surface waters in August 2000 and June 2001. Although our specific primer, the oligonucleotide CF319a, only recovers 36% of all *Bacteroidetes* sequences > 1200 base pairs in the RDP database, it matches to > 90% of all currently available sequence types from marine habitats. However, it should be noted that members of the typically marine genus *Algoriphagus* would be missed by our screening method, because the 16S rRNA gene of these bacteria features three mismatches to the oligonucleotide CF319a. In addition, there are several genera of mainly freshwater *Sphingobacteria* that might also be missed by our screening (e.g. *Flectobacillus*, *Hongiella* and *Flameovirga*). Consequently, other pairs of primers should be considered to survey the diversity of *Bacteroidetes* in non-marine environments.

Local diversity of *Bacteroidetes*

There was almost no overlap between the clone libraries from the two different sampling dates (Table 1). This indicates a pronounced temporal variation of the local diversity of *Flavobacteria* at our sampling site. A comparison with another clone library from the same environment

sampled in early March 2003 (Sekar *et al.*, 2004) supports this conclusion: only four of the 23 sequences from that library were identical to sequences from our libraries. High temporal variability of the diversity of *Bacteroidetes* has been reported in several environments. Schauer and colleagues (Schauer *et al.*, 2003) described seasonal succession of *Bacteroidetes* in the Mediterranean Sea over a 1 year period using denaturing gradient gel electrophoresis (DGGE), and Bano and Hollibaugh (Bano and Hollibaugh, 2002) reported seasonality of *Cytophaga* and *Polaribacter* spp. in Arctic ocean samples analysed by DGGE and clone libraries. Isolation of *Bacteroidetes* from a single environment (eutrophic lake, salt marsh) at different sampling dates has also led to the recovery of predominantly unique strains (Jaspers *et al.*, 2001; Lydell *et al.*, 2004).

Some sequence types from our libraries have been found previously in other marine environments, indicating their cosmopolitan distribution (Table 1): Pacific coast (Brown *et al.*, 2005), Delaware estuary (Kirchman *et al.*, 2003), open ocean (Zubkov *et al.*, 2002), Arctic regions (Bano and Hollibaugh, 2002) and deep sea habitats (Kormas *et al.*, 2006). By contrast, half of our sequence types represent novel genera if grouped with all marine *Bacteroidetes* sequences > 1200 base pairs from the RDP database (Table 2), and the closest neighbours of half of our sequence types in the GenBank database had similarities of < 95% (Table 1). In order to estimate the typical range of newly recovered local diversity we performed a comparison with two other studies that have focused on the *Bacteroidetes*. In a PCR-based clone library from a coastal site (20 long sequences) 45% of genera, 50% of species and 80% of sequence types represented novel groups (Kirchman *et al.*, 2003), thus adding 3% to total diversity of marine *Bacteroidetes*. By contrast, a relatively low proportion of sequence types from an open ocean library (Abell and Bowman, 2005a) represented novel clusters: 12.5% of 22 almost complete sequence types were affiliated with novel genera, 10% with novel species and 35% with novel unique sequence types. Although our comparison is restricted to only three studies, it suggests that coastal environments might feature not only higher abundances but also a higher diversity of *Bacteroidetes* than the open ocean.

Considering the high fraction of newly obtained genera and species in our library, it is theoretically conceivable that the local diversity of *Bacteroidetes* in coastal marine waters might be higher than of other bacterial phyla that occur in similar abundances in this environment, e.g. the *Alphaproteobacteria* (Eilers *et al.*, 2001; Kirchman *et al.*, 2003). We, thus, compared the local diversity of the two above-mentioned groups in a study from a coastal site that provides an appropriate data set (Acinas *et al.*, 2004). In this investigation, similar numbers of 16S rRNA gene

sequences of *Alphaproteobacteria* and *Bacteroidetes* were obtained (131 and 171 respectively). Interestingly, the Chao1 diversity index predicts that the expected number of genera, species and unique sequence types of *Bacteroidetes* at that site would be approximately twice as high (97, 101 and 150 respectively) as the number of genera, species and unique sequence types of *Alphaproteobacteria* (51, 59 and 71 respectively). A more rigorous comparison of the global diversity of different microbial phyla in marine waters might be required to substantiate these findings. However, this goes beyond the scope of our study.

Global diversity of pelagic marine Bacteroidetes

All our analyses were performed on sequences > 1200 base pairs from the latest available release of the RDP data set. The limitation to almost complete sequences is a necessary precondition for the grouping (Seguritan and Rohwer, 2001) and phylogenetic (Hedges, 2002) algorithms required to examine and predict *Bacteroidetes* diversity at different levels of sequence similarity. A different approach for the assessment of diversity of marine bacteria was chosen by Hagström and colleagues (Hagström *et al.*, 2002). These authors assembled partial sequences into consensus contigs at a similarity level of 97%. Despite these different approaches, our estimates of currently known marine *Bacteroidetes* species fit to the outcome of that study: Hagström and colleagues found 97 cultured and 66 uncultured *Bacteroidetes* species in the GenBank database queried in 2001. Taking into account (i) the approximate fourfold increase of available 16S rRNA gene sequences since that time and (ii) the slopes of the rarefaction curves at 97% sequence similarity (Fig. 3), our investigation reaches comparable conclusions (193 and 211 cultured and uncultured *Bacteroidetes* species, respectively, Table 3).

However, while these authors have argued that much (if not all) of the marine microbial species richness has been sampled (Hagström *et al.*, 2002; Pommier *et al.*, 2005) we present different lines of evidence (i.e. high local diversity in our clone libraries, rarefaction analyses, frequencies of singletons and doubletons) suggesting a significant undersampling of the global diversity of pelagic marine *Bacteroidetes* both at the genus and species level (Tables 2 and 3, Fig. 3). Although this undersampling may affect the performance of diversity estimators (Hughes and Hellmann, 2005), the here used non-parametric estimators can nevertheless provide lower boundaries of the expected 'real' diversity (Bohannon and Hughes, 2003). Both the Chao1 and ACE estimators predict approximately 1200 species of marine planktonic *Bacteroidetes*, which is three times the currently known number (Table 3).

Diversity of cultured and uncultured pelagic marine Bacteroidetes

None of the 16S sequence types in our clone libraries was closely related to *Flavobacteria* that have been previously isolated from the same environment (Eilers *et al.*, 2001). Several studies have combined the isolation-based identification of environmental *Bacteroidetes* with cultivation-independent approaches, e.g. from Arctic pack ice (Brinkmeyer *et al.*, 2003), estuarine enrichment cultures (Kisand and Wikner, 2003) or salt marshes (Lydell *et al.*, 2004). These investigations have all reached similar conclusions, i.e. that the cultured and uncultured diversity of environmental *Bacteroidetes* was essentially non-overlapping. In addition, targeted attempts to isolate representatives from those *Bacteroidetes* lineages that are abundant in the environment often fail (e.g. O'Sullivan *et al.*, 2004).

Our analysis of the global diversity of pelagic marine *Bacteroidetes* confirmed that there is a fundamental phylogenetic separation between cultured and uncultured lineages from this phylum. Only 14 of 325 *Bacteroidetes* 'genera' were composed of both isolates and uncultured representatives, and the estimated similarity between the subsets of cultured and uncultured *Bacteroidetes* ranged between 10% and 18% (Chao–Jaccard and Chao–Sorensen indices respectively). Rarefaction analysis moreover indicated that the expected total diversity of uncultured *Bacteroidetes* genera and species was significantly higher than of cultured *Bacteroidetes* at the current sampling state (i.e. database size), whereas there was no difference between unique sequence types (Fig. 3). This suggests that there are a limited number of readily culturable genera and species of pelagic marine *Bacteroidetes* from which numerous strains have been isolated.

It should be noted that the exclusion of partial sequences < 1200 base pairs from our analysis might have artificially decreased the overlap between the diversity of isolates and uncultured *Bacteroidetes*. Nevertheless, the apparent dissimilarity of the two sets of sequence types probably also reflects the respective biases of either approach. On the one hand, PCR-based molecular approaches have various pitfalls that limit their potential to assess microbial diversity in its entirety (Clayton *et al.*, 1995; von Wintzingerode *et al.*, 1997; Ashelford *et al.*, 2005). On the other hand, most *Bacteroidetes* isolates that are represented by almost complete 16S rRNA gene sequences in the RDP database have been recovered by the classical spread plate technique. Novel cultivation strategies have led to the successful isolation of new *Bacteroidetes* species and of strains from lineages previously known from clone libraries only: Kaeberlein and colleagues (Kaeberlein *et al.*, 2002) isolated a new species of *Sphingobacteria* by

placing marine sediment bacteria in diffusion chambers that allowed the exchange of metabolites with the original sediment. Using gel encapsulation of cells and highly parallel cultivation under low nutrient conditions Zengler and colleagues (Zengler *et al.*, 2002) obtained an isolate that is the only cultured representative of a monophyletic lineage of marine *Flavobacteria*. Our results stress the importance of such creative approaches in order to isolate strains from the > 50% of known species of pelagic marine *Bacteroidetes* that are presently not cultured (Table 3).

Distribution of Bacteroidetes across different aquatic environments

Most of the *Bacteroidetes* sequence types retrieved from aquatic habitats other than marine plankton were also affiliated with *Flavobacteria*, although *Sphingobacteria* were more common in freshwater habitats. *Bacteroidetes* from marine planktonic habitats accounted for approximately half of the known diversity at the genus level (Fig. 4). Because our analyses indicate that the diversity of *Bacteroidetes* within this habitat is still undersampled (Table 3), it is likely that such undersampling may be even more pronounced in other habitats. For example, only 10% of available *Bacteroidetes* sequences originated from marine and freshwater sediments, although representatives of this phylum may occur in high numbers both in coastal marine sediments (Llobet-Brossa *et al.*, 1998; Musat *et al.*, 2006) and in riverine benthic organic matter (Fazi *et al.*, 2005).

A relatively small set of *Bacteroidetes* genera appears to have successfully radiated across different habitat types. Only 15% of genera originated from two or more different environments (Fig. 4, filled bars). Interestingly, the relatively small proportion of aquatic *Bacteroidetes* genera that occur in more than one habitat accounted for almost half of all analysed sequences (Fig. 4, open bars). Certainly, this is in part so because several of the habitat-crossing genera harbour readily culturable representatives (Table 4). For example, numerous strains of the genera *Flavobacterium* and *Chryseobacterium* have been retrieved from freshwaters, marine plankton, sediments and diseased aquatic animals (Bowman *et al.*, 1996; Eilers *et al.*, 2000; Kisand *et al.*, 2002; Brinkmeyer *et al.*, 2003). It is conceivable that the conspicuous distribution of these groups across various aquatic habitat types is related to an 'opportunistic' life strategy that also favours cultivation by classical plating techniques (Eilers *et al.*, 2000). However, an equal number of habitat-crossing genera have been retrieved with molecular methods. Recently, Carlos Pedrós-Alió proposed a conceptual framework suggesting that essentially only the abundant microbial taxa would be retrieved with molecular techniques (Pedrós-Alió, 2006). Thus, it would be interesting to investigate whether those genera of wide-

spread distribution are actually members of the core diversity responsible for the carbon and energy fluxes in aquatic environments.

Experimental procedures

Sampling, construction of clone libraries

Surface water samples were collected from the German Bight of the North Sea on 24 August 2000 and 14 June 2001 at station Helgoland Roads (54°11'N, 7°54'E). Portions of 10 ml were fixed with paraformaldehyde (2% final concentration) for 2–24 h and filtered onto membrane filters (type GTTP; pore size 0.2 µm; diameter 47 mm; Millipore). Filters were stored at –20°C until further processing.

Almost complete bacterial 16S rRNA genes were amplified via PCR using the primers GM3F and GM4R (Muyzer *et al.*, 1993). Polymerase chain reaction was performed as previously described, using small pieces of the above-mentioned preparation on membrane filters as templates (Kirchman *et al.*, 2001; Warnecke *et al.*, 2004). Polymerase chain reaction products were purified using the QIAquick PCR purification kit (Qiagen, Hilden, Germany) and ligated into the pCR4-TOPO vector using the TOPO TA cloning kit (Invitrogen, Groningen, the Netherlands). Competent *Escherichia coli* cells JM109 (Promega, Mannheim, Germany) were transformed and plated on Luria–Bertoni (LB) agar plates containing 50 µg ml⁻¹ of ampicillin. Single colonies were transferred into microtiter plates (MTP) containing ampicillin-amended LB medium (50 µg ml⁻¹) and cultivated overnight at 37°C. Glycerol (50% final concentration) was added to each well, and the MTPs were stored at –80°C until further processing.

Screening of clone libraries

Clones in the MTPs were screened for *Bacteroidetes*-related sequence types by PCR with the primers CF319a (5'-TGGTCCGTGTCTCAGTAC-3') (Manz *et al.*, 1996) and CFB563 (5'-ATTGGGTTAAAGGGTCC-3') (Weller *et al.*, 2000). A database search revealed that > 90% of currently known 16S rRNA gene sequences of planktonic marine *Bacteroidetes* are targeted by these oligonucleotides. The accumulation of PCR products was detected as previously described (Warnecke *et al.*, 2004) with an ABI SDS 7700 instrument (Taqman, Applied Biosystems, Foster City, USA), using the double-stranded DNA-binding dye SYBR Green I. Polymerase chain reactions were performed at the following conditions: initial denaturation at 50°C for 2 min and at 95°C for 10 min, followed by 30 cycles consisting of denaturation (15 s at 95°C), annealing (30 s at 58°C) and extension (1 min at 72°C). These conditions were determined using fully sequenced inserts with zero to two mismatches to the primers. Each MTP featured one well with a positive control (i.e. a previously sequenced clone affiliated to *Bacteroidetes*) and one well with a no-template control. Relative fluorescence values of SYBR Green I at the end of amplification were used to identify clones with *Bacteroidetes* 16S rRNA inserts. Prior to applying the screening procedure to all clones, a random subset of 50 clones was selected to confirm that the optimized Taqman protocol yielded identical results

as regular PCR and subsequent gel electrophoresis. Clones harbouring *Bacteroidetes* sequences were analysed by restriction fragment length polymorphism (RFLP) with two enzymes (HaeIII and RsaI) performed overnight at 37°C. At least one representative of each RFLP pattern was selected for sequence analysis.

Sequencing

Plasmids were isolated from clones with the QIAprep Spin Miniprep kit (Qiagen). Sequencing reactions were performed using the ABI BigDye® chemistry and an ABI 3100 genetic analyser (Applied Biosystems, Foster City, USA) according to the manufacturer's instructions. The following primers were used to obtain almost complete 16S rRNA gene sequences: GM1F (Muyzer *et al.*, 1993), M13F (5'-GTAAAACGACGGCCAG-3') and M13R (5'-CAGGAAACAGCTATGAC-3'). Partial sequences were assembled and manually corrected using the software Sequencher (Gene Codes, MI, USA). The assembled sequences were checked for chimera origin with the free software Mallard and Pintail (<http://www.cf.ac.uk/bios/research/biosoft/>) (Ashelford *et al.*, 2005). The coverage index of the clone collection was estimated by the equation $C = (1 - n_i/N) \times 100$, where N represents the number of all clones carrying *Bacteroidetes* inserts and n_i the number of unique RFLP patterns (O'Sullivan *et al.*, 2002). 16S rRNA gene sequences produced during this study were deposited in GenBank under the accession numbers AM279161–AM279213.

Phylogenetic reconstructions

Phylogenetic analyses were performed using the ARB software package (Ludwig *et al.*, 2004). For the reconstruction of phylogenetic trees, only nearly complete 16S rRNA sequences (i.e. longer than 1200 nucleotides) were considered. The ARB database (release January 2005; <http://www.arb-home.de>) was completed with 11 131 *Bacteroidetes* sequences available in the RDP II (release March 2006; <http://rdp.cme.msu.edu/>) (Cole *et al.*, 2005). The 10 closest relatives of the newly produced sequences were searched and automatically pre-aligned using the ARB tool Fast_Aligner. Alignments were subsequently improved manually considering the secondary structure of the rRNA molecule. The respective ARB tools were used to perform maximum parsimony (MP), neighbour-joining (NJ) and maximum likelihood (ML) analyses on various subsets of the data set. The resulting phylogenetic trees were compared manually. The final consensus tree shows bifurcations only if branchings appeared to be stable and well separated from neighbouring branchings in the majority of analyses. Multifurcations were introduced if tree topologies could not be resolved unambiguously. In addition, a phylogenetic tree of all available *Bacteroidetes* sequences from the marine plankton was calculated by MP on a larger data set of 1380 almost complete sequences from this lineage.

Diversity analysis

Bacteroidetes-related phylotypes from various aquatic habitats were extracted from the sequence-associated informa-

tion of the imported RDP data set using the ARB 'Search and Query' function and various keywords. The following categories were discriminated: marine planktonic (coastal, open and deep ocean), freshwater planktonic (lakes, streams, springs and rivers), other aquatic (not possible to define the nature of the water body), hypersaline environments, sediments (includes both freshwater and marine), associated with algae and associated with aquatic animals. In addition, the collection was split into sequences from cultured and uncultured *Bacteroidetes*. Prior to further analysis the data set was checked for chimeric or otherwise anomalous sequences using Mallard and Pintail (Ashelford *et al.*, 2005).

Sequences of *Bacteroidetes* of aquatic origin (including the ones from this study) were subsequently analysed using the software FastGroup II (<http://phage.sdsu.edu/research/tools/fastgroup/>), which allows to group sequences according to a defined level of similarity (Yu *et al.*, 2006). The grouping criteria were set to compare sequence stretches of at least 800 bp including the highly variable regions V3–V7 (Neefs *et al.*, 1990). Grouping was performed at similarity levels of 95%, 97% and 99%. These values were chosen to approximate the taxonomic categories genus (Schloss and Handelsman, 2005) and species (Stackebrandt and Goebel, 1994) and to set a conservative threshold for unique sequence types. Analyses were performed either on the complete data set or on subsets split according to habitat types or culturability. Rarefaction curves of sequence types from the pelagic marine environment were calculated from the results of the FastGroup analysis using the software aRarefactWin (<http://www.uga.edu/~strata/software/Software.html>). The diversity indexes Chao1 and ACE (Hughes *et al.*, 2001) and the similarity indexes Bray–Curtis, Chao–Jaccard and Chao–Sorensen (Chao *et al.*, 2005) were calculated for all planktonic marine *Bacteroidetes* using the software EstimateS (Colwell, 2005) (<http://viceroy.eeb.uconn.edu/EstimateS>).

Acknowledgements

This study was supported by the European Union (EVK3-2001-00194 BASICS), by a grant from the German Academic Exchange Program awarded to C.A., and by the Max-Planck Society. Many thanks to Thierry Lombardot for kindly providing the 16S sequences from the env_nt data set and for fruitful discussions about bioinformatic sequence analysis. Kevin E. Ashelford, Steven M. Holland, R.K. Colwell and Forest Rohwer are, respectively, acknowledged for providing the free softwares Pintail and Mallard, aRarefactWin, EstimateS and FastGroup. The authors would also like to thank two anonymous reviewers whose helpful comments led to a substantial improvement of an earlier version of this manuscript.

References

- Abell, G.C.J., and Bowman, J.P. (2005a) Ecological and biogeographic relationships of class *Flavobacteria* in the Southern Ocean. *FEMS Microbiol Ecol* **51**: 265–277.
- Abell, G.C.J., and Bowman, J.P. (2005b) Colonization and community dynamics of class *Flavobacteria* on diatom

- detritus in experimental mesocosms based on Southern Ocean seawater. *FEMS Microbiol Ecol* **53**: 379–391.
- Acinas, S.G., Klepac-Ceraj, V., Hunt, D.E., Pharino, C., Ceraj, I., Distel, D.L., and Polz, M.F. (2004) Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **430**: 551–554.
- Ashelford, K.E., Chuzhanova, N.A., Fry, J.C., Jones, A.J., and Weightman, A.J. (2005) At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol* **71**: 7724–7736.
- Bano, N., and Hollibaugh, J.T. (2002) Phylogenetic composition of bacterioplankton assemblages from the Arctic Ocean. *Appl Environ Microbiol* **68**: 505–518.
- Bohannon, B.J., and Hughes, J. (2003) New approaches to analyzing microbial biodiversity data. *Curr Opin Microbiol* **6**: 282–287.
- Bowman, J.P., Cavanagh, J., Austin, J.J., and Sanderson, K. (1996) Novel *Psychrobacter* species from Antarctic ornithogenic soils. *Int J Syst Bacteriol* **46**: 841–848.
- Brinkmeyer, R., Knittel, K., Jurgens, J., Weyland, H., Amann, R., and Helmke, E. (2003) Diversity and structure of bacterial communities in arctic versus antarctic pack ice. *Appl Environ Microbiol* **69**: 6610–6619.
- Brown, M.V., Schwabach, M.S., Hewson, I., and Fuhrman, J.A. (2005) Coupling 16S-ITS rDNA clone libraries and automated ribosomal intergenic spacer analysis to show marine microbial diversity: development and application to a time series. *Environ Microbiol* **7**: 1466–1479.
- Chao, A., Chazdon, R.L., Colwell, R.K., and Shen, T.J. (2005) A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecol Lett* **8**: 148–159.
- Clayton, R., Sutton, G., Hinkle, P., Jr, Bult, C., and Fields, C. (1995) Intraspecific variation in small-subunit rRNA sequences in GenBank: why single sequences may not adequately represent prokaryotic taxa. *Int J Syst Bacteriol* **45**: 595–599.
- Cole, J.R., Chai, B., Farris, R.J., Wang, Q., Kulam, S.A., McGarrell, D.M., *et al.* (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res* **33**: D294–D296.
- Colwell, R.K. (2005) *EstimateS: Statistical Estimation of Species Richness and Shared Species from Samples*. Version 7.5. User's Guide and application. URL <http://purl.oclc.org/estimates>
- Cottrell, M.T., and Kirchman, D.L. (2000a) Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence *in situ* hybridization. *Appl Environ Microbiol* **66**: 5116–5122.
- Cottrell, M.T., and Kirchman, D.L. (2000b) Natural assemblages of marine proteobacteria and members of the *Cytophaga-Flavobacter* cluster consuming low- and high-molecular-weight dissolved organic matter. *Appl Environ Microbiol* **66**: 1692–1697.
- Cottrell, M.T., Waidner, L.A., Yu, L., and Kirchman, D.L. (2005) Bacterial diversity of metagenomic and PCR libraries from the Delaware River. *Environ Microbiol* **7**: 1883–1895.
- Daims, H., Bruhl, A., Amann, R., Schleifer, K.H., and Wagner, M. (1999) The domain-specific probe EUB338 is insufficient for the detection of all *Bacteria*: development and evaluation of a more comprehensive probe set. *Syst Appl Microbiol* **22**: 434–444.
- DeLong, E.F., Franks, D.G., and Alldredge, A.L. (1993) Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. *Limnol Oceanogr* **38**: 924–934.
- DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.U., *et al.* (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.
- Eilers, H., Pernthaler, J., Glöckner, F.O., and Amann, R. (2000) Culturability and *in situ* abundance of pelagic bacteria from the North Sea. *Appl Environ Microbiol* **66**: 3044–3051.
- Eilers, H., Pernthaler, J., Peplies, J., Glöckner, F.O., Gerdt, G., and Amann, R. (2001) Isolation of novel pelagic bacteria from the German Bight and their seasonal contribution to surface picoplankton. *Appl Environ Microbiol* **67**: 5134–5142.
- Fazi, S., Amalfitano, S., Pernthaler, J., and Puddu, A. (2005) Bacterial communities associated with benthic organic matter in headwater stream microhabitats. *Environ Microbiol* **7**: 1633–1640.
- Grossart, H.P., Levold, F., Allgaier, M., Simon, M., and Brinkhoff, T. (2005) Marine diatom species harbour distinct bacterial communities. *Environ Microbiol* **7**: 860–873.
- Hagström, A., Pinhassi, J., and Zweifel, U.L. (2000) Biogeographical diversity among marine bacterioplankton. *Aquat Microb Ecol* **21**: 231–244.
- Hagström, A., Pommier, T., Rohwer, F., Simu, K., Stolte, W., Svensson, D., and Zweifel, U.L. (2002) Use of 16S ribosomal DNA for delineation of marine bacterioplankton species. *Appl Environ Microbiol* **68**: 3628–3633.
- Hedges, S.B. (2002) The origin and evolution of model organisms. *Nat Rev Genet* **3**: 838–849.
- Hughes, J.B., and Hellmann, J.J. (2005) The application of rarefaction techniques to molecular inventories of microbial diversity. In *Methods in Enzymology*. Leadbetter, J.R. (ed.). San Diego, CA, USA: Academic Press, pp. 292–308.
- Hughes, J.B., Hellmann, J.J., Ricketts, T.H., and Bohannon, B.J.M. (2001) Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl Environ Microbiol* **67**: 4399–4406.
- Jaspers, E., Nauhaus, K., Cypionka, H., and Overmann, J. (2001) Multitude and temporal variability of ecological niches as indicated by the diversity of cultivated bacterioplankton. *FEMS Microbiol Ecol* **36**: 153–164.
- Kaeberlein, T., Lewis, K., and Epstein, S.S. (2002) Isolating 'uncultivable' microorganisms in pure culture in a simulated natural environment. *Science* **296**: 1127–1129.
- Kirchman, D.L. (2002) The ecology of *Cytophaga-Flavobacteria* in aquatic environments. *FEMS Microbiol Ecol* **39**: 91–100.
- Kirchman, D.L., Yu, L.Y., Fuchs, B.M., and Amann, R. (2001) Structure of bacterial communities in aquatic systems as revealed by filter PCR. *Aquat Microb Ecol* **26**: 13–22.

- Kirchman, D.L., Yu, L.Y., and Cottrell, M.T. (2003) Diversity and abundance of uncultured *Cytophaga*-like bacteria in the Delaware Estuary. *Appl Environ Microbiol* **69**: 6587–6596.
- Kisand, V., and Wikner, J. (2003) Combining culture-dependent and -independent methodologies for estimation of richness of estuarine bacterioplankton consuming riverine dissolved organic matter. *Appl Environ Microbiol* **69**: 3607–3616.
- Kisand, V., Cuadros, R., and Wikner, J. (2002) Phylogeny of culturable estuarine bacteria catabolizing riverine organic matter in the Northern Baltic Sea. *Appl Environ Microbiol* **68**: 379–388.
- Kormas, K.A., Tivey, M.K., Von Damm, K., and Teske, A. (2006) Bacterial and archaeal phylotypes associated with distinct mineralogical layers of a white smoker spire from a deep-sea hydrothermal vent site (9°N, East Pacific Rise). *Environ Microbiol* **8**: 909–920.
- Llobet-Brossa, E., Rossellò-Mora, R., and Amann, R. (1998) Microbial community composition of Wadden Sea sediments as revealed by fluorescence-*in-situ*-hybridization. *Appl Environ Microbiol* **64**: 2691–2696.
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.
- Lydell, C., Dowell, L., Sikaroodi, M., Gillevet, P., and Emerson, D. (2004) A population survey of members of the phylum *Bacteroidetes* isolated from salt marsh sediments along the East Coast of the United States. *Microb Ecol* **48**: 263–273.
- Manz, W., Amann, R., Ludwig, W., Vancanneyt, M., and Schleifer, K.-H. (1996) Application of a suite of 16S rRNA-specific oligonucleotide probes designed to investigate bacteria of the phylum *Cytophaga-Flavobacter-Bacteroides* in the natural environment. *Microbiol* **142**: 1097–1106.
- Musat, N., Werner, U., Knittel, K., Kolb, S., Dodenhof, T., van Beusekom, J.E.E., *et al.* (2006) Microbial community structure of sandy intertidal sediments in the North Sea, Sylt-Romo Basin, Wadden Sea. *Syst Appl Microbiol* **29**: 333–348.
- Muyzer, G., de Waal, E.C., and Uitterlinden, A.G. (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59**: 695–700.
- Neeffs, J.M., Vandeppeer, Y., Hendriks, L., and Dewachter, R. (1990) Compilation of small ribosomal-subunit RNA sequences. *Nucleic Acids Res* **18**: 2237–2317.
- O'Sullivan, L.A., Weightman, A.J., and Fry, J.C. (2002) New degenerate *Cytophaga-Flexibacter-Bacteroides*-specific 16S ribosomal DNA-targeted oligonucleotide probes reveal high bacterial diversity in River Taff epilithon. *Appl Environ Microbiol* **68**: 201–210.
- O'Sullivan, L.A., Fuller, K.E., Thomas, E.M., Turley, C.M., Fry, J.C., and Weightman, A.J. (2004) Distribution and culturability of the uncultivated 'AGG58 cluster' of *Bacteroidetes* phylum in aquatic environments. *FEMS Microbiol Ecol* **47**: 359–370.
- Pedrós-Alió, C. (2006) Marine microbial diversity: can it be determined? *Trends Microbiol* **14**: 257–263.
- Pinhassi, J., Sala, M.M., Havskum, H., Peters, F., Guadayol, O., Malits, A., and Marrase, C.L. (2004) Changes in bacterioplankton composition under different phytoplankton regimens. *Appl Environ Microbiol* **70**: 6753–6766.
- Pommier, T., Pinhassi, J., and Hagstrom, A. (2005) Biogeographic analysis of ribosomal RNA clusters from marine bacterioplankton. *Aquat Microb Ecol* **41**: 79–89.
- Riemann, L., Steward, G.F., and Azam, F. (2000) Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. *Appl Environ Microbiol* **66**: 578–587.
- Schauer, M., Balague, V., Pedros-Alio, C., and Massana, R. (2003) Seasonal changes in the taxonomic composition of bacterioplankton in a coastal oligotrophic system. *Aquat Microb Ecol* **31**: 163–174.
- Schloss, P.D., and Handelsman, J. (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**: 1501–1506.
- Seguritan, V., and Rohwer, F. (2001) FastGroup: a program to dereplicate libraries of 16S rDNA sequences. *BMC Bioinformatics* **2**: 9.
- Sekar, R., Fuchs, B., Amann, R., and Pernthaler, J. (2004) Flow sorting of marine bacterioplankton after fluorescence *in situ* hybridization. *Appl Environ Microbiol* **70**: 6210–6219.
- Simon, M., Glöckner, F.O., and Amann, R. (1999) Different community structure and temperature optima of heterotrophic picoplankton in various regions of the Southern Ocean. *Aquat Microb Ecol* **18**: 275–284.
- Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., *et al.* (2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci USA* **103**: 12115–12120.
- Stackebrandt, E., and Goebel, B. (1994) Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* **44**: 846–849.
- Stevens, H., Stubner, M., Simon, M., and Brinkhoff, T. (2005) Phylogeny of *Proteobacteria* and *Bacteroidetes* from oxic habitats of a tidal flat ecosystem. *FEMS Microbiol Ecol* **54**: 351–365.
- Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.
- Warnecke, F., Pernthaler, J., and Amann, R. (2004) Actinobacterial 16S rRNA genes from freshwater habitats cluster in four distinct lineages. *Environ Microbiol* **6**: 242–253.
- Webster, N.S., Wilson, K.J., Blackall, L.L., and Hill, R.T. (2001) Phylogenetic diversity of bacteria associated with the marine sponge *Rhopaloeides odorabile*. *Appl Environ Microbiol* **67**: 434–444.
- Weller, R., Glöckner, F.O., and Amann, R. (2000) 16S rRNA-targeted oligonucleotide probes for the *in situ* detection of members of the phylum *Cytophaga-Flavobacterium-Bacteroides*. *Syst Appl Microbiol* **23**: 107–114.
- von Wintzingerode, F., Göbel, U.B., and Stackebrandt, E. (1997) Determination of microbial diversity in environmen-

- tal samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21**: 213–229.
- Yu, Y.N., Breitbart, M., McNairnie, P., and Rohwer, F. (2006) FastGroupII: a web-based bioinformatics platform for analyses of large 16S rDNA libraries. *BMC Bioinformatics* **7**: 57.
- Zengler, K., Toledo, G., Rappe, M., Elkins, J., Mathur, E.J., Short, J.M., and Keller, M. (2002) Cultivating the uncultured. *PNAS* **99**: 15681–15686.
- Zubkov, M.V., Fuchs, B.M., Tarran, G.A., Burkill, P.H., and Amann, R. (2002) Mesoscale distribution of dominant bacterioplankton groups in the northern North Sea in early summer. *Aquat Microb Ecol* **29**: 135–144.