

Northumbria Research Link

Citation: Edwards, Jennifer, Smith, Darren, Connolly, John, McDonald, James, Cox, Michael, Joint, Ian, Edwards, Clive and McCarthy, Alan (2010) Identification of Carbohydrate Metabolism Genes in the Metagenome of a Marine Biofilm Community Shown to Be Dominated by Gammaproteobacteria and Bacteroidetes. *Genes*, 1 (3). pp. 371-384. ISSN 2073-4425

Published by: MDPI

URL: <http://dx.doi.org/10.3390/genes1030371>
<<http://dx.doi.org/10.3390/genes1030371>>

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/id/eprint/4265/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

Article

Identification of Carbohydrate Metabolism Genes in the Metagenome of a Marine Biofilm Community Shown to Be Dominated by Gammaproteobacteria and Bacteroidetes

Jennifer L. Edwards¹, Darren L. Smith¹, John Connolly¹, James E. McDonald¹, Michael J. Cox¹, Ian Joint², Clive Edwards¹ and Alan J. McCarthy^{1,*}

¹ Microbiology Research Group, School of Biological Sciences, BioSciences Building, University of Liverpool, Crown Street, Liverpool, L69 7ZB, UK; E-Mails: jenedw@live.co.uk (J.L.E.), smithdx@liverpool.ac.uk (D.L.S.), john76@liverpool.ac.uk (J.C.), j.mcdonald@bangor.ac.uk (J.E.M.), mikeyj.cox@gmail.com (M.J.C.), kj26@liverpool.ac.uk (C.E.)

² Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, PL1 3DH, UK; E-Mail: irj@pml.ac.uk

* Author to whom correspondence should be addressed; E-Mail: aj55m@liverpool.ac.uk; Tel.: +44 151 795 4574; Fax: +44 151 795 4410.

Received: 21 September 2010; in revised form: 5 October 2010 / Accepted: 11 October 2010 / Published: 26 October 2010

Abstract: Polysaccharides are an important source of organic carbon in the marine environment and degradation of the insoluble and globally abundant cellulose is a major component of the marine carbon cycle. Although a number of species of cultured bacteria are known to degrade crystalline cellulose, little is known of the polysaccharide hydrolases expressed by cellulose-degrading microbial communities, particularly in the marine environment. Next generation 454 Pyrosequencing was applied to analyze the microbial community that colonizes and degrades insoluble polysaccharides *in situ* in the Irish Sea. The bioinformatics tool MG-RAST was used to examine the randomly sampled data for taxonomic markers and functional genes, and showed that the community was dominated by members of the *Gammaproteobacteria* and *Bacteroidetes*. Furthermore, the identification of 211 gene sequences matched to a custom-made database comprising the members of nine glycoside hydrolase families revealed an extensive repertoire of functional genes predicted to be involved in cellulose utilization. This demonstrates that

the use of an *in situ* cellulose baiting method yielded a marine microbial metagenome considerably enriched in functional genes involved in polysaccharide degradation. The research reported here is the first designed to specifically address the bacterial communities that colonize and degrade cellulose in the marine environment and to evaluate the glycoside hydrolase (cellulase and chitinase) gene repertoire of that community, in the absence of the biases associated with PCR-based molecular techniques.

Keywords: 454 pyrosequencing; next generation sequencing; marine polysaccharide degradation; glycoside hydrolases; metagenomics; marine bacteria; cellulose biofilm

1. Introduction

It is established that bacteria are mainly responsible for the consumption of organic matter in the marine environment, and the carbon cycle is largely dependent upon this remineralization of biomass. The polysaccharides, proteins and lipids produced as a result of carbon fixation together form the core of the particulate organic matter (POM) and dissolved organic matter (DOM) pools in the marine environment. POM and DOM are mineralized and oxidized by heterotrophic marine bacteria to generate microbial biomass and inorganic products [1]. Biogeochemical processes within the marine environment as a whole are dynamic and complex, and are poorly understood [2]. Furthermore, microbial community structure varies significantly between coastal and open waters, at different water depths, and also seasonally. Through 16S rRNA gene sequence analysis, marine microbial populations have been cataloged to show that particle-attached and free-living communities differ, with the former colonized by bacteria (related to *Cytophaga spp*, *Planctomyces spp* and the γ Proteobacteria), while the planktonic fraction is rich in members of the α Proteobacteria [3]. Whole genome shotgun sequencing (WGS) of filtered sea water from the Sargasso Sea also confirmed the predominance of Proteobacteria (primarily α , β and γ subgroups) followed by the Firmicutes, Cyanobacteria and Bacteroidetes when a range of phylogenetic markers were used for comparison [4]. More recently, metagenomic and metatranscriptomic analysis of ocean surface water has shown the Cyanobacteria (genus *Prochlorococcus*) and *Alphaproteobacteria* (genus *Roseobacter*) to be the best represented taxonomic groups [5]. However, few studies have directly addressed structural polysaccharide degradation in marine ecosystems, and have focused entirely on the isolation and identification of bacterial species, including members of the genus *Pseudoalteromonas*. [6,7]. More recently, *Saccharophagus degradans* and *Teredinibacter turnerae* have emerged as strong examples of two well characterized marine bacterial species involved in polysaccharide degradation, and whole genome annotations have revealed an extensive repertoire of relevant functional genes [8,9]. There is a diversity of polysaccharide structures and sources in the marine environment, but cellulose is well represented and likely to persist in POM due to the recalcitrance imparted by its crystalline structure.

The Glycoside Hydrolases (GHs) are modular enzymes that hydrolyse glycosidic bonds of carbohydrates, with classification based on amino acid sequence and predicted three-dimensional structure. Such enzymes may contain single or multiple catalytic modules (GH) together with single or multiple non-catalytic carbohydrate-binding modules (CBMs) [10-12]. While the screening of large

insert, vector based metagenomic libraries from other environments has resulted in the occasional identification of cellulase genes, these have typically been present at very low frequencies. Only four cellulase positive clones were identified amongst *ca.* 100,000 cosmids prepared from a sample of compost DNA [13] and 70,000 clones of ~40 kbp constructed from soil DNA yielded only one cellulase positive clone [14]. These data suggest that fosmid and cosmid-derived metagenomic libraries of environmental samples do not provide an adequate assessment of the functional diversity of GH genes, even for microbial communities that would be expected to retain a significant capability for polysaccharide biodegradation. Extraction of DNA fragments of adequate size from environmental samples is a limiting step in large insert library based metagenomics, whereas 454 pyrosequencing is a high throughput alternative to identifying functional genes present in microbial communities from relatively short fragments of DNA. A number of studies have successfully targeted environments known to be rich in polysaccharide-degrading microorganisms, including the hindgut microbiota of a wood feeding termite, bovine rumen contents and an enriched switchgrass composting system [15-17].

The majority of marine bacteria have yet to be isolated in culture, resulting in a paucity of information on community structure and function, particularly for colonized organic matter representing the primary stage of carbon mineralization in the sea. Metagenomic studies offer the potential to gain an unprecedented insight into such communities but, as has been demonstrated in previous DNA sequencing projects [4], marine microbial communities are much too diverse to generate sequence information that provides adequate coverage of the total community. In order to circumvent the problem of under-representation of genes for environmentally important primary functions, we have employed an *in situ* enrichment technique previously used to both identify and isolate cellulolytic organisms in freshwater lakes [18,19]. Here, *in situ* colonized cellulose ‘bait’ is used as the source of biological material for metagenomic studies directed at organisms and genes involved in this primary step in carbon recycling in the Irish Sea.

2. Results and Discussion

2.1. Metagenome Analysis

Cellulose baits were tethered to a Cefas SMART buoy in the Irish Sea, and recovered after one month *in situ* incubation. Total DNA was extracted from the microbial biofilm that colonized the cotton string and 454 pyrosequencing was performed on the community DNA. This metagenomic sequencing generated 223,263 reads of DNA sequence, containing 48,338,140 bp of DNA with a size range of 8-375 bp. The raw read data were assembled into 26,860 contiguous sequences (contigs) with a size range of 93-26,859 bp, comprising 6,841,343 bp in total. Of the assembled contigs in the dataset, the majority were less than 1 kb in length, suggesting significant heterogeneity within the sample. One contig of ~26 Kb was removed from the dataset and analysed separately as it was identified to have sequence similarity with a bacteriophage (data not shown). The phylogenetic diversity within the dataset was assessed using the SEED MG-RAST (Metagenome Rapid Annotation Using Subsystem Technology) [20]. Taxonomic information on the sequences within the metagenomic data was obtained by comparing the contiguous sequences against the Greengenes 16S rRNA gene database [21]. Using an expected cut off (e-value) of 1×10^{-5} and a minimum alignment of 50 bp, a total of

18 partial 16S rRNA gene fragments were identified. Fourteen of these were assigned to the Bacteria, of which four are affiliated with the phylum *Bacteroidetes* (*Flavobacterium*; *Ulvibacter*; *Roseivirga* and unclassified *Cytophagaceae*) and ten with the phylum Proteobacteria (*Sulfitobacter*; unclassified *Rhodobacteraceae* (3); *Glaciecola* (3); *Teredinibacter*; *Cellvibrio* and unclassified *Gammaproteobacteria*) (Table 1). The remaining four contigs identified as fragments of 16S rRNA genes could not be assigned. The occurrence of 16S rRNA genes in the pyrosequenced dataset reported was therefore one in every 1,413 contigs. 16S rRNA gene-based phylogenetic studies have become routine in microbial ecology, and through such studies it is recognized that *Bacteroidetes* (particularly members of the genus *Cytophaga*) and members of the alpha and gamma lineages of the Proteobacteria colonize and contribute to the mineralization of organic aggregates, while noting that the community structure of particle-attached and free-living bacteria differ [2,22-25]. Due to the inherent complexity of marine biofilms [26], it has become clear that recent metagenomic studies have barely scratched the surface of the total phylogenetic and functional diversity of the marine microbial community. There is a clear advantage in including an *in situ* enrichment step for capturing functional genes that are involved in influential processes in marine carbon cycling as well as the dominant bacterial groups residing within the polysaccharide colonizing community.

Table 1. Identification of 16S rRNA genes in the Irish Sea cellulose biofilm DNA 454 pyrosequenced dataset by the Greengenes 16S rRNA gene database.

Sequence ID	Alignment Length	E-value	% Identity	Bit Score	Fragment (Start - End)	Taxonomy Assignment	Best Hit ID
<u>Proteobacteria</u>							
contig16635	85	1.00E-40	100	168	1 - 85	<i>Sulfitobacter</i>	165917
contig26360	150	2.51E-51	93	202	1 - 149	Unclassified <i>Rhodobacteraceae</i>	142124
contig26572	162	7.94E-70	96	264	4 - 163	Unclassified <i>Rhodobacteraceae</i>	70710
contig26707	130	2.51E-63	99	242	1 - 130	Unclassified <i>Rhodobacteraceae</i>	113926
contig25574	235	2.51E-130	100	466	50 - 284	<i>Glaciecola</i>	108683
contig26573	526	0.00E+00	99	1003	1 - 526	<i>Glaciecola</i>	80428
contig26860	151	1.26E-80	100	299	1 - 151	<i>Glaciecola</i>	170839
contig26820	696	0.00E+00	93	1015	6 - 701	<i>Teredinibacter</i>	144812
contig00070	247	1.26E-81	93	305	738 - 983	<i>Cellvibrio</i>	98921
contig12909	577	0.00E-00	93	858	209 - 784	Unclassified <i>Gammaproteobacteria</i>	151615
<u>Bacteroidetes</u>							
contig00061	892	0.00E+00	90	1094	1988 - 2879	<i>Roseivirga</i>	102384
contig26228	139	2.51E-66	98	252	1 - 138	<i>Flavobacterium</i>	154970
contig26430	241	5.01E-130	99	464	1 - 241	<i>Ulvibacter</i>	80102
contig26765	436	5.01E-140	89	500	60 - 495	Unclassified <i>Cytophagaceae</i>	2577

Best Hit is the ProkMSA Greengenes reference ID.

Further information on the taxonomic diversity residing within this Irish Sea metagenome dataset was obtained by investigating protein-encoding genes as taxonomic markers. This was achieved by comparing the assembled contigs against the SEED-nr database [27]. A total of 14,179 contigs could be assigned to known functional genes, of which 14,020 were to the Bacteria (98.9% of positive

matches), with Eukaryota, Viruses and Archaea accounting for 0.65%, 0.18% and 0.29% of the positive hits, respectively. Taxonomic assignment of the biofilm community was dominated by members of the Proteobacteria (8847; 61% of the assignments to Bacteria) and Bacteroidetes (4465; 32% of assignments to Bacteria) (Figure 1). Although only 14 contigs were identified as fragments of bacterial 16S rRNA genes, the markers of choice for taxonomic assignments, it is encouraging that the predominance of Proteobacteria and Bacteroidetes discovered therein (Table 1) is mirrored by the SEED MG-RAST analysis of the distribution of taxonomic assignments in over 14,000 contigs from protein coding genes (Figure 1) in which Gammaproteobacteria and Flavobacteria dominate the assignments at the class level (Figure 2). Notably the emergence of the carbohydrate degrading *Microbulbifer/Teredinibacter/Saccharophagus* group of the Gammaproteobacteria [28] and their role in marine polysaccharide degradation is in keeping with the results found here. This is also true of the well established role of Bacteroidetes, formerly known as the *Cytophaga/Flavobacteria/Bacteroides* group, and their dominance in the marine environment particularly in coastal regions [29,30] and their role in phytoplankton colonization [31] and polysaccharide degradation [30]. The conclusion that this biofilm colonizing cellulose suspended in the Irish Sea is dominated by these two major bacterial groups can therefore be made with some confidence.

Figure 1. Summary of Phylum level taxonomic assignment of 14,020 assembled contigs matching protein encoding genes (PEGs).

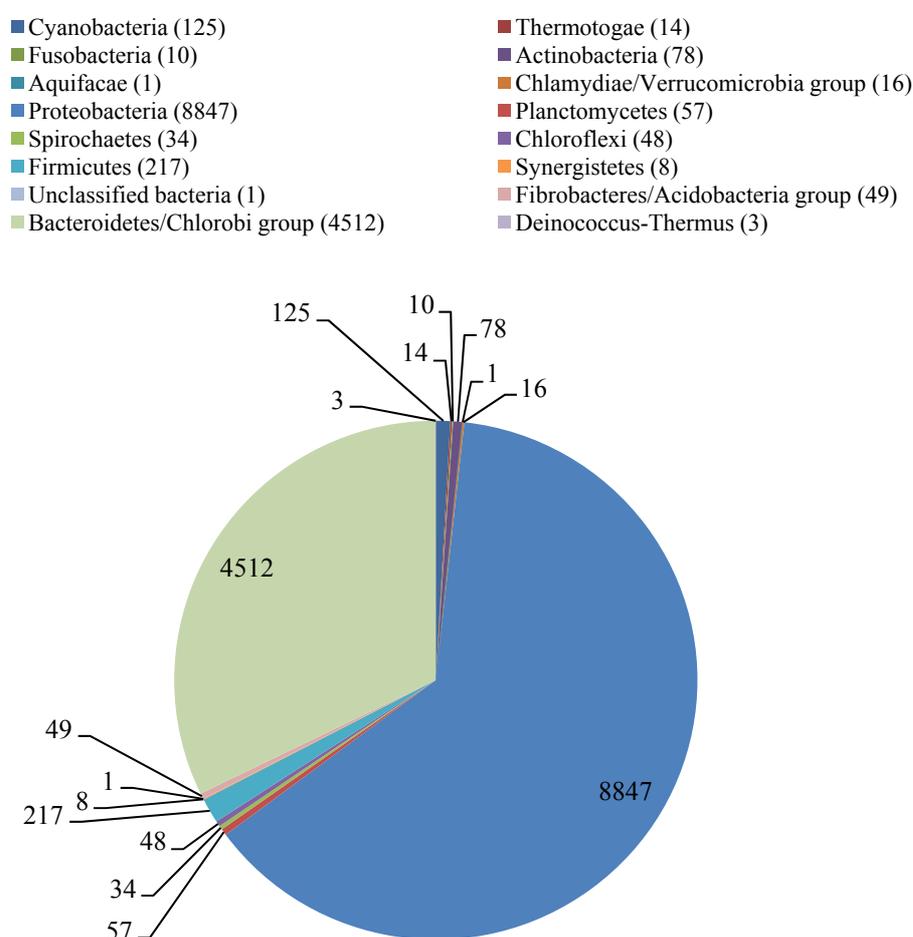
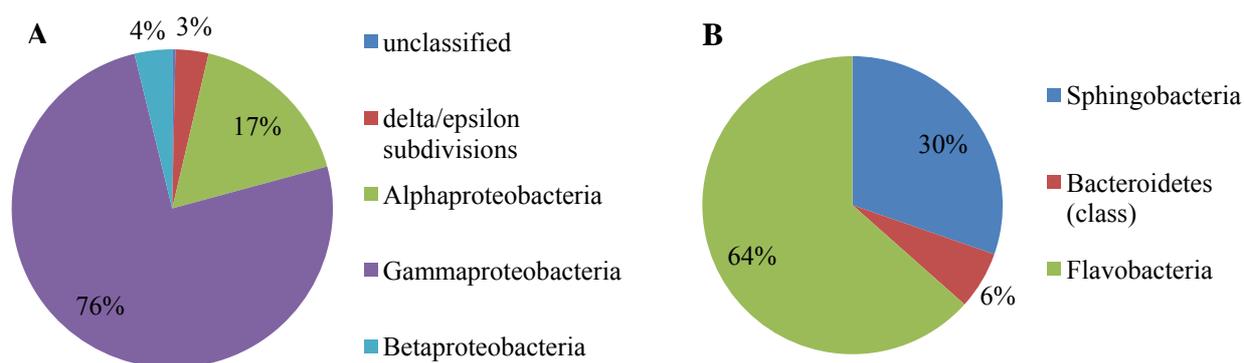


Figure 2. Taxonomic assignment of 13,312 sequence contigs matching Protein encoding genes (PEGs) of the Proteobacteria and the Bacteroidetes. The percentage of contigs matching PEGs assigned at the Class level using the SEED MG-RAST (Metagenome Rapid Annotation Using Subsystem Technology) are shown for (A) Proteobacteria (8847; 61% of assignments to the Bacteria) and (B) Bacteroidetes (4465; 32% of assignments to Bacteria). Values are shown as percentages of the total contigs assigned to each class within the two Phyla.



2.2. Polysaccharide hydrolases

Glycoside Hydrolase families containing representatives of selected functions such as endoglucanase and chitinase were identified and catalytic domains downloaded from the Pfam database. This customized database comprised a total of 14,332 protein sequences distributed across twelve of the GH families (5, 6, 7, 8, 9, 12, 16, 18, 19, 45, 48 and 61). Comparison of the 26,859 metagenome pyrosequencing-derived contigs to the constructed GH database provided 211 hits at an E value restriction of 1×10^{-5} (Table 2; a complete list of all those contigs with matches to sequences in the constructed GH database is provided in Supplementary Table 1). The most frequently occurring GH families were GH5 (56; 27% of total matches), 8 (40; 19%), 9 (30; 14%) and 16 (64; 30%). GH5 is one of the largest and most diverse of the GH families, with several known enzymatic functions. GH families 8 and 9 primarily contain endoglucanases and cellobiohydrolases, whilst GH16 proteins show Endo-1,3(4)- β glucanase and xyloglucanase activities against carbohydrates that are usually intertwined with cellulose in plant cell wall material [10,32,33]. The contigs that were identified to have sequence similarity were then compared against the NCBI nr database using BLASTX [34]. Sequence similarity hits that were most frequent were proteins from the Gram negative, rod shaped aerobic Gammaproteobacterium *S. degradans* (27 matches), a marine cellulolytic bacterium, *Teredinibacter turnerae* [8,28] (17 matches), a Gammaproteobacterium closely related to *S. degradans* that occurs as an intracellular endosymbiont in the gills of wood boring bivalves [9] and the Gram-negative, aerobic, rod-shaped gliding Bacteroidete *Cytophaga hutchinsonii* (48 matches), a cellulolytic bacterium originally isolated from soil [35]. Taken together, these account for 44% of all the hits to the

GH database summarized in Table 2; a complete list of all hits against the NCBI nr database is provided in Supplementary Table 2). All three of these named species are known to be actively cellulolytic, containing a plethora of enzymes involved in the hydrolysis of polysaccharides [8,9,35].

Table 2. Summary of the number of 454 sequencing contigs matching each GH family.

CAZy family	Pfam ID	Number of hits
GH5	PF00150	56
GH6	PF01341	5
GH7	PF00840	0
GH8	PF01270	40
GH9	PF00759	30
GH12	PF01670	1
GH16	PF00722	64
GH18	PF00704	10
GH19	PF00182	3
GH45	PF02015	0
GH48	PF02011	2
GH61	PF03443	0

Searches were performed using Blastx for all 26,859 assembled contigs against the customized library of downloaded GH families using an E value cut-off of 1×10^{-5} . Pfam ID is the protein family [36] and CAZy (carbohydrate active enzymes database family [10]). A comprehensive list of all hits and values can be found in Supplementary Table 1.

The quantity of contigs providing sequence similarity to GH supports the use of *in situ* cellulose baits as a means of enriching the metagenome in genes that encode polysaccharide-degrading functions. Ninety-eight of the sequence matches to the constructed GH database are unique, whereby one contig matches one sequence in the database, whereas the remainder had sequence similarity with at least two other contiguous sequences (Supplementary Table 3). For example, 13 contigs matched a GH8 family protein from the Bacteroidete *C. hutchinsonii* (Q11PI8_CYTH3), nine further contigs also hit another GH8 family protein from *C. hutchinsonii* (Q11VQ4_CYTH3) and nine contigs hit a GH16 family protein from the Gammaproteobacterium *S. degradans* (Q21KX4_SACD2). Only 10 (<10%) contigs matched proteins of GH family 18, and three contigs matched to GH19 proteins (Table 2), the two families that contain all known chitinases (<http://www.cazy.org>). Two contigs matched the same protein, C1S930_9SPHI, from the chitinolytic Bacteroidete, *Chitinophaga pinensis*. A few culture-independent molecular biological studies have addressed GHs in the marine environment, but these have focused mainly on chitinases, due to the fact that marine vibrios are often chitinolytic and traditionally amongst the most readily isolated and cultivated marine bacteria [37]. Cottrell *et al.* [38] screened metagenomic libraries of coastal and estuarine water DNA using fluorogenic analogues of chitin and cellulose to identify a number of genes involved in chitin hydrolysis, but none that expressed proteins with activity against the cellulose analogue. The diversity and abundance of β -1,4 endoglucanases within the GH5 family in the North Atlantic Ocean has been investigated by designing primers for a 350 bp fragment constituting ~one-third of the gene. Relative abundance was determined for three locations using qPCR, and found to positively correlate with chlorophyll concentrations. Analysis of clone libraries showed that the GH5 family genes were more diverse in

samples from coastal water than those from the open ocean [39]. Cottrell *et al.* [40] screened a fosmid library constructed from prokaryotic DNA from the Western Arctic Ocean. PCR primer sets were designed for the most abundant type of endoglucanase identified in the Sargasso Sea WGS dataset [4], but subsequent functional analysis revealed that the gene encoded a peptidase. Functional molecular ecology studies targeting polysaccharide degradation in the marine environment are clearly fragmentary in nature.

2.3. Scanning Electron Microscopy analysis of colonized cellulose bait

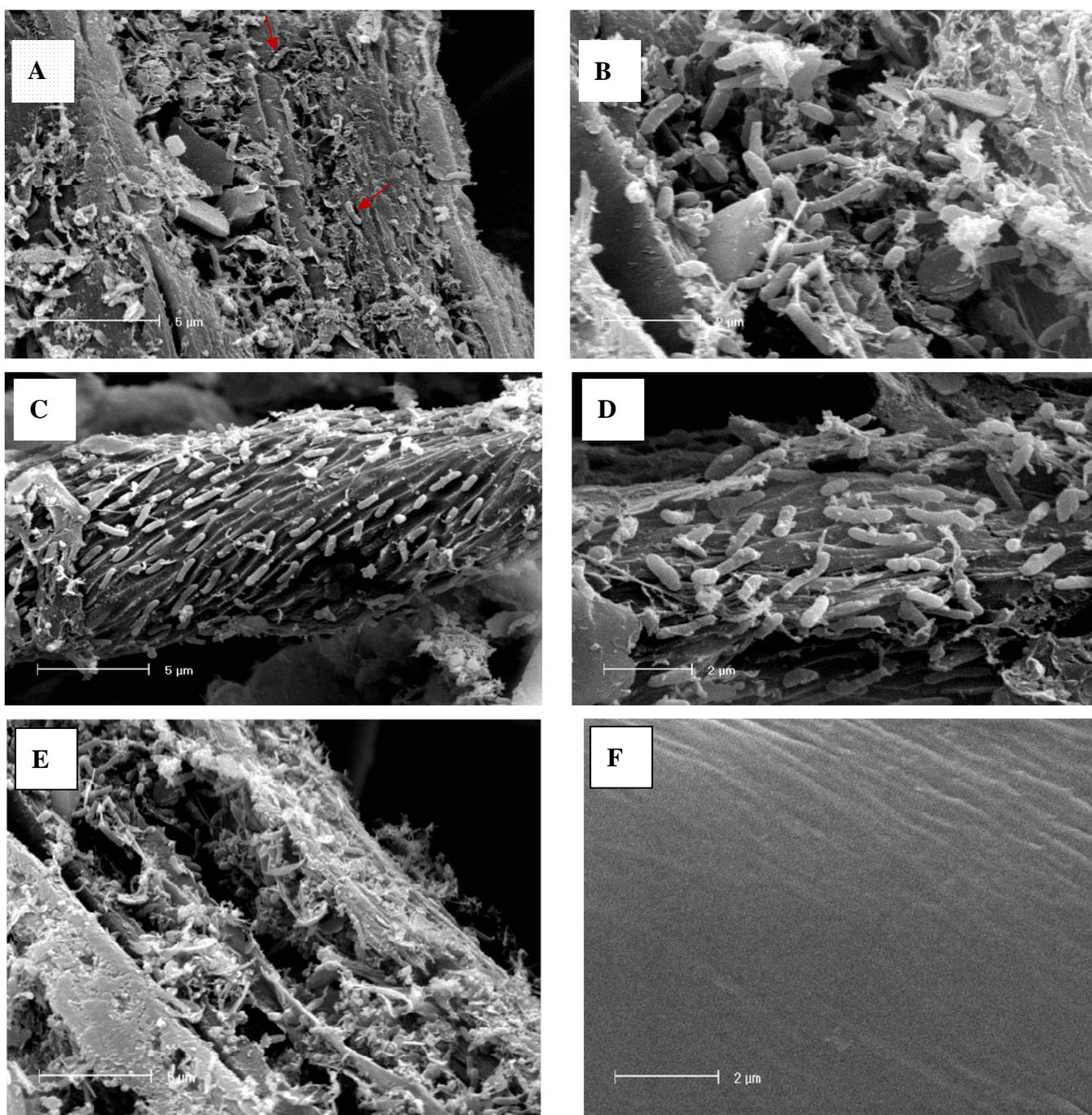
The surface of colonized cellulose recovered from the same sampling site in the Irish Sea was visualized using SEM. The cellulose bait was heavily colonized (Figure 3 a-c and e), with predominantly rod shaped bacteria (Figure 3 c, d). Microorganisms were often arranged in rows on the cellulose surface (Figure 3c) with pockets visible where degradation had occurred (Figure 3). In areas of heavy colonization by the dense biofilm matrix, visible signs of degradation were observed where the cotton surface had been eroded (Figure 3 a, b). In comparison the surface of cotton string not colonized by microorganisms is smooth in appearance (Figure 3f). Protuberances are apparent on the surface of many rod-shaped cells (Figure 3d) one explanation for which could be the presence of polycellulosomes, the macromolecular structures responsible for cellulase activity in a number of bacterial species [41]. *S. degradans*, the marine aerobic polysaccharide-degrading species whose proteins gave matches to over 10% of the GH hits above, produces surface structures in the presence of cellulose [28], and a number of hydrolytic enzymes are thought to be consorted in such *S. degradans* complexes [8]. It has been suggested through genome sequence analysis of *S. degradans* and *T. turnerae* that Gram-negative bacteria use lipoproteins to anchor carbohydrate active enzymes to the outer membrane, playing a similar role to that of cellulosomes that are usually associated with Gram-positive bacteria [8,9]. *S. degradans* and *C. hutchinsonii* which both make up a large proportion of hits to GHs are known to be Gram-negative aerobic rod-shaped bacteria [28,34]. The morphology of which is in keeping with that seen in the SEM pictures (Figure 3 a-e).

3. Experimental Section

3.1. Sampling

Cotton yarn (0.7 cm diameter) (Lancashire cotton best twist, from Texere Yarns, Bradford, UK) was used as the cellulose bait. Approximately 1 m of yarn was placed in customized nylon mesh bags (10 cm × 10 cm). ‘Baits’ were tethered to a Cefas SMART (<http://www.cefas.co.uk>) buoy in the Eastern Irish Sea in surface water (53° 27′ N 3° 38.6′ W) for a period of 1 month.

Figure 3. Scanning electron micrographs of the microbial community colonizing cellulose ‘bait’ *in situ* in the Irish Sea. (A) A cellulose fiber heavily colonized by a biofilm matrix. Rod-shaped bacteria can clearly be seen occupying grooves or pits (arrow) providing visible evidence of degradation by these cellulolytic bacteria; (B) a close image (bar = 2 μ m) of a hollowed out region of a cellulose fiber with abundant rod-shaped bacteria; (C) rod-shaped bacteria arranged in rows around the cellulose surface with grooves visible where bacterial cells have formed pits on the surface; (D) a closer view of a cellulose fiber where rod shaped bacteria are located in grooves on the surface; (E) surface of cotton string heavily colonized by biofilm matrix showing extensive degradation of the cellulose surface, in comparison to (F) uncolonized cotton string.



3.2. Metagenome Analysis

3.2.1. DNA extraction [42]

Nucleic acids were extracted by placing 0.5 g (wet weight) string (retrieved from the Irish Sea following one month *in situ*, April, 2008) in a Q-biogene purple top multimix tube (lysing matrix E). 0.5 ml hexadecyltrimethylammonium bromide (CTAB) buffer [prepared by mixing equal volumes of 10 % (w/v) CTAB in 0.7 M NaCl with 240 mM potassium phosphate buffer, pH 8.0] was added along with phenol-chloroform-isoamyl alcohol (25:24:1; pH 8.0). Cells were lysed by bead beating in a Ribolyser for 30 s at a speed of 5.5 m/s, and the aqueous phase containing nucleic acids separated by centrifugation ($16,000 \times g$) for 5 min at 4 °C. The aqueous phase was transferred to a fresh microfuge tube and phenol removed by mixing an equal volume of chloroform-isoamyl alcohol (24:1), followed by centrifugation at ($16,000 \times g$) for 5 min. Nucleic acids were obtained by precipitation of the top layer by the addition of 2 volumes of 30% polyethylene glycol (PEG) solution (30% polyethylene glycol and 1.6 M NaCl), incubated overnight at 4 °C. The precipitated nucleic acids were collected by centrifugation ($16,000 \times g$) for 15 min. The supernatant was removed and the pellet washed with 200 μ l 70% ice cold ethanol and air dried prior to resuspension in 50 μ l sterile ddH₂O.

3.2.2. Pyrosequencing and sequence analysis

DNA was sequenced using the 454 Corporation's GS-FLX instrument at the NERC-funded Advanced Genomics Facility at the University of Liverpool (<http://www.liv.ac.uk/agf/>). 26,859 assembled contigs were uploaded in a FastA format to the MG-RAST server at the SEED [20,43] on 15 February 2010 under the name Irish_Sea_Metagenome and was assigned the Metagenome ID: 4446437.3.

3.2.3. Glycoside Hydrolase database construction

The GHs representing classes of GHs of functional interest (e.g. endoglucanase and chitinase) were identified by analyzing the GH families in the Carbohydrate Active Enzyme (CAZy) web resource. All of the protein sequences of catalytic domains of GH families 5, 6, 7, 8, 9, 12, 16, 18, 19, 45, 48 and 61 were downloaded (February 2010) from the Pfam database [36]. These families were chosen based on known functions, including mainly cellulase and chitinase ability (families 18 and 19). All metagenome-derived contigs were used as a query and compared against the GH database using Blastx. Only those hits with an E value cut-off of 1×10^{-5} were recorded.

3.3. Scanning Electron Microscopy (SEM) of colonized cellulose samples

Samples of cellulose bait were collected from the mooring site in the Irish Sea (January, 2009) and refrigerated during transport to the laboratory. The samples were gently rinsed with ddH₂O, immersed in excess absolute Ethanol (Sigma) which had been pre cooled to -80 °C, and the samples returned to -80 °C overnight. The samples were then removed and placed into a universal bottle containing pre-cooled absolute ethanol, and again returned to -80 °C until required. Specimens were dried from absolute ethanol in carbon dioxide using a Polaron E3000 critical point dryer, glued to stubs,

sputter-coated with 60 % gold-palladium in a Polaron E 5100 coater and viewed in a Philips 501B scanning electron microscope at accelerating voltages of 7.2 and 15 kV [44]. Final sample preparation and primary microscopic examination of the samples was carried out by Cornelis Veltkamp & Carmel Pinnington at the Department of Earth and Ocean Sciences, University of Liverpool.

4. Conclusions

Cellulose is the most abundant polysaccharide on Earth and the degradation of this recalcitrant substrate is therefore an important driver of the carbon cycle. Although a number of cultured bacterial species are known to degrade cellulose, there is a paucity of information on true community function and structure, particularly in the marine environment [45]. Metagenomic analysis of the biofilm that developed on cellulose immersed in the Irish Sea for one month has revealed a community dominated by members of the Gammaproteobacteria and Bacteroidetes, supported by both protein encoding and 16S rRNA gene distribution in the 454 pyrosequence dataset. Significantly, a total of 211 genes were identified as potentially involved in the polysaccharide degrading process, reflecting the metagenome of a microbial community enriched in cellulolytic microorganisms. SEM supported the metagenomic data in showing a biofilm dominated by small ($\leq 1 \mu\text{m}$) rod-shaped bacteria that form erosion pits in the cellulose surface. There was a relatively high frequency of sequence similarity matches to genes of the marine cellulose-degrading bacterial species *S. degradans* and *T. turnerae* and the cellulolytic soil bacterium *C. hutchinsonii*, both of which have sequenced genomes and cellulase systems that have been well characterized recently. However it is important to note that the available sequence information in databases greatly under represents the true diversity present in environmental microbial communities and matches are to the closest relatives in databases. The metagenomic data here provide evidence for the existence of a pool of potentially exploitable polysaccharide hydrolases in the marine microbial community, revealed by an *in situ* approach to generating metagenomes that are enriched in genes that encode these enzymes.

Acknowledgements

This study was funded by the Natural Environment Research Council (NERC). The 454 pyrosequencing and assembly was performed by the NERC-funded Advanced Genomics Facility at the University of Liverpool (<http://www.liv.ac.uk/agf/>). We wish to thank John Howarth & Phil Knight at the Proudman Oceanography Laboratory (POL) for providing access to the Cefas SMART buoys and to the crew and scientists of the RV Prince Madog, and the Liverpool Bay Coastal Observatory. For the SEM work we wish to thank Cornelis Veltkamp & Carmel Pinnington at the Department of Earth & Ocean Sciences, University of Liverpool.

References and Notes

1. Nagata, T. Organic matter-bacteria interactions in seawater. In *Microbial Ecology of the Oceans*, 2nd ed.; Kirchman, D.L., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2008, p. 207-242.
2. DeLong, E.F.; Karl, D.M. Genomic perspectives in microbial oceanography. *Nature* **2005**, *437*, 336-342.

3. Delong, E.F.; Franks, D.G.; Alldredge, A.L. Phylogenetic diversity of aggregate-attached vs free-living marine bacterial assemblages. *Limnol. Ocean.* **1993**, *38*, 924-934.
4. Venter, J.C.; Remington, K.; Heidelberg, J.F.; Halpern, A.L.; Rusch, D.; Eisen, J.A.; Wu, D.; Paulsen, I.; Nelson, K.E.; Nelson, W. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **2004**, *304*, 66-74.
5. Frias-Lopez, J.; Shi, Y.; Tyson, G.W.; Coleman, M.L.; Schuster, S.C.; Chisholm, S.W.; Delong, E.F. Microbial community gene expression in ocean surface waters. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 3805-3810.
6. Garsoux, G.; Lamotte, J.; Gerday, C.; Feller, G. Kinetic and structural optimization to catalysis at low temperatures in a psychrophilic cellulase from the Antarctic bacterium *Pseudoalteromonas haloplanktis*. *Biochem. J.* **2004**, *384*, 247-253.
7. Zeng, R.; Xiong, P.; Wen, J. Characterization and gene cloning of a cold-active cellulase from a deep-sea psychrotrophic bacterium *Pseudoalteromonas* sp. DY3. *Extremophiles* **2006**, *10*, 79-82.
8. Weiner, R.M.; Taylor, L.E., 2nd; Henrissat, B.; Hauser, L.; Land, M.; Coutinho, P.M.; Rancurel, C.; Saunders, E.H.; Longmire, A.G.; Zhang, H.; *et al.* Complete genome sequence of the complex carbohydrate-degrading marine bacterium, *Saccharophagus degradans* strain 2-40 T. *PLoS Genet.* **2008**, *4*, e1000087.
9. Yang, J.C.; Madupu, R.; Durkin, A.S.; Ekborg, N.A.; Peadarallu, C.S.; Hostetler, J.B.; Radune, D.; Toms, B.S.; Henrissat, B.; Coutinho, P.M. *et al.* The complete genome of *Teredinibacter turnerae* T7901: an intracellular endosymbiont of marine wood-boring bivalves (shipworms). *PLoS One* **2009**, *4*, e6085.
10. CAZy Home Page. <http://www.cazy.org> (accessed on 11 July 2010).
11. Cantarel, B.L.; Coutinho, P.M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **2009**, *37*, D233-238.
12. Henrissat, B.; Davies, G. Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* **1997**, *7*, 637-644.
13. Pang, H.; Zhang, P.; Duan, C.J.; Mo, X.C.; Tang, J.L.; Feng, J.X. Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase. *Curr. Microbiol.* **2009**, *58*, 404-408.
14. Kim, S.J.; Lee, C.M.; Han, B.R.; Kim, M.Y.; Yeo, Y.S.; Yoon, S.H.; Koo, B.S.; Jun, H.K. Characterization of a gene encoding cellulase from uncultured soil bacteria. *FEMS Microbiol. Lett.* **2008**, *282*, 44-51.
15. Allgaier, M.; Reddy, A.; Park, J.I.; Ivanova, N.; D'Haeseleer, P.; Lowry, S.; Sapro, R.; Hazen, T.C.; Simmons, B.A.; VanderGheynst, J.S. *et al.* Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. *PLoS One* **2010**, *5*, e8812.
16. Brulc, J.M.; Antonopoulos, D.A.; Miller, M.E.; Wilson, M.K.; Yannarell, A.C.; Dinsdale, E.A.; Edwards, R.E.; Frank, E.D.; Emerson, J.B.; Wacklin, P. *et al.* Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 1948-1953.

17. Warnecke, F.; Luginbuhl, P.; Ivanova, N.; Ghassemian, M.; Richardson, T.H.; Stege, J.T.; Cayouette, M.; McHardy, A.C.; Djordjevic, G.; Aboushadi, N. *et al.* Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **2007**, *450*, 560-565.
18. de Menezes, A.B.; Lockhart, R.J.; Cox, M.J.; Allison, H.E.; McCarthy, A.J. Cellulose degradation by micromonosporas recovered from freshwater lakes and classification of these actinomycetes by DNA gyrase B gene sequencing. *Appl. Environ. Microbiol.* **2008**, *74*, 7080-7084.
19. McDonald, J.E.; de Menezes, A.B.; Allison, H.E.; McCarthy, A.J. Molecular biological detection and quantification of novel *Fibrobacter* populations in freshwater lakes. *Appl. Environ. Microbiol.* **2009**, *75*, 5148-5152.
20. Meyer, F.; Paarmann, D.; D'Souza, M.; Olson, R.; Glass, E.M.; Kubal, M.; Paczian, T.; Rodriguez, A.; Stevens, R.; Wilke, A. *et al.* The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **2008**, *9*, 386-399.
21. DeSantis, T.Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E.L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G.L. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **2006**, *72*, 5069-5072.
22. DeLong, E.F. Modern microbial seascapes. *Nat. Rev. Microbiol.* **2007**, *5*, 755-757.
23. Rath, J.; Wu, K.Y.; Herndl, G.J.; DeLong, E.F. High phylogenetic diversity in a marine-snow-associated bacterial assemblage. *Aquat. Microb. Ecol.* **1998**, *14*, 261-269.
24. Simon, M.; Grossart, H.P.; Schweitzer, B.; Ploug, H. Microbial ecology of organic aggregates in aquatic ecosystems. *Aquat. Microb. Ecol.* **2002**, *28*, 175-211.
25. Smith, D.C.; Simon, M.; Alldredge, A.L.; Azam, F. Intense hydrolytic enzyme-activity on marine aggregates and implications for rapid particle dissolution. *Nature* **1992**, *359*, 139-142.
26. Giovannoni, S.J.; Stingl, U. Molecular diversity and ecology of microbial plankton. *Nature* **2005**, *437*, 343-348.
27. Overbeek, R.; Begley, T.; Butler, R.M.; Choudhuri, J.V.; Chuang, H.Y.; Cohoon, M.; de Crecy-Lagard, V.; Diaz, N.; Disz, T.; Edwards, R. *et al.* The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* **2005**, *33*, 5691-5702.
28. Ekborg, N.A.; Gonzalez, J.M.; Howard, M.B.; Taylor, L.E.; Hutcheson, S.W.; Weiner, R.M. *Saccharophagus degradans* gen. nov., sp. nov., a versatile marine degrader of complex polysaccharides. *Int. J. System. Evol. Microbiol.* **2005**, *55*, 1545-1549.
29. Johansen, J.E.; Nielsen, P.; Sjöholm, C. Description of *Cellulophaga baltica* gen. nov., sp. nov. and *Cellulophaga fucicola* gen. nov., sp. nov. and reclassification of [*Cytophaga*] *lytica* to *Cellulophaga lytica* gen. nov., comb. nov. *Int. J. Syst. Bacteriol.* **1999**, *49* (Pt 3), 1231-1240.
30. Kirchman, D.L. The ecology of *Cytophaga-Flavobacteria* in aquatic environments. *FEMS Microbiol. Ecol.* **2002**, *39*, 91-100.
31. Gomez-Pereira, P.R.; Fuchs, B.M.; Alonso, C.; Oliver, M.J.; van Beusekom, J.E.; Amann, R. Distinct flavobacterial communities in contrasting water masses of the north Atlantic Ocean. *ISME J.* **2010**, *4*, 472-487.
32. Jovanovic, I.; Magnuson, J.K.; Collart, F.; Robbertse, B.; Adney, W.S.; Himmel, M.E.; Baker, S.E. Fungal glycoside hydrolases for saccharification of lignocellulose: outlook for new discoveries fueled by genomics and functional studies. *Cellulose* **2009**, *16*, 687-697.

33. Strohmeier, M.; Hrmova, M.; Fischer, M.; Harvey, A.J.; Fincher, G.B.; Pleiss, J. Molecular modeling of family GH16 glycoside hydrolases: potential roles for xyloglucan transglucosylases/hydrolases in cell wall modification in the poaceae. *Protein Sci.* **2004**, *13*, 3200-3213.
34. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403-410.
35. Xie, G.; Bruce, D.C.; Challacombe, J.F.; Chertkov, O.; Detter, J.C.; Gilna, P.; Han, C.S.; Lucas, S.; Misra, M.; Myers, G.L. *et al.* Genome sequence of the cellulolytic gliding bacterium *Cytophaga hutchinsonii*. *Appl. Environ. Microbiol.* **2007**, *73*, 3536-3546.
36. Pfam Home Page. <http://pfam.sanger.ac.uk> (accessed on 11 July 2010).
37. Grimes, D.J.; Johnson, C.N.; Dillon, K.S.; Flowers, A.R.; Noriega, N.F.; Berutti, T. What genomic sequence information has revealed about *Vibrio* ecology in the ocean-v review. *Microb. Ecol.* **2009**, *58*, 447-460.
38. Cottrell, M.T.; Moore, J.A.; Kirchman, D.L. Chitinases from uncultured marine microorganisms. *Appl. Environ. Microbiol.* **1999**, *65*, 2553-2557.
39. Elifantz, H.; Malmstrom, R.R.; Cottrell, M.T.; Kirchman, D.L. Assimilation of polysaccharides and glucose by major bacterial groups in the Delaware Estuary. *Appl. Environ. Microbiol.* **2005**, *71*, 7799-7805.
40. Cottrell, M.T.; Yu, L.; Kirchman, D.L. Sequence and expression analyses of *Cytophaga*-like hydrolases in a western arctic metagenomic library and the Sargasso Sea. *Appl. Environ. Microbiol.* **2005**, *71*, 8506-8513.
41. Bayer, E.A.; Shimon, L.J.; Shoham, Y.; Lamed, R. Cellulosomes-structure and ultrastructure. *J. Struct. Biol.* **1998**, *124*, 221-234.
42. Griffiths, R.I.; Whiteley, A.S.; O'Donnell, A.G.; Bailey, M.J. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl. Environ. Microbiol.* **2000**, *66*, 5488-5491.
43. The SEED Home Page. <http://metagenomics.theseed.org> (accessed on 11 July 2010).
44. Veltkamp, C.J.; Chubb, J.C.; Birch, S.P.; Eaton, J.W. A simple freeze dehydration method for studying epiphytic and epizoic communities using the scanning electron-microscope. *Hydrobiologia* **1994**, *288*, 33-38.
45. Fuhrman, J.A.; Hagstrom, A. Bacterial and archaeal community structure and its patterns. In *Microbial Ecology of the Oceans*, 2nd ed.; Kirchman, D.L., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2008; p. 45-90.