Diversity and abundance of glycosyl hydrolase family 5 in the North Atlantic Ocean

Hila Elifantz, Lisa A. Waidner, Vanessa K. Michelou, Matthew T. Cottrell & David L. Kirchman

College of Marine and Earth Studies, University of Delaware, Lewes, DE, USA

Correspondence: David L. Kirchman, College of Marine and Earth Studies, University of Delaware, Lewes, DE 19958, USA. Tel.: +1 302 645 4375; fax: +1 302 645 4007; e-mail: kirchman@udel.edu

Present address: Hila Elifantz, Department of Microbiology, University of Massachusetts, Amherst, MA 01003-9298, USA. Lisa A. Waidner, Delaware Biotechnological Institute, University of Delaware, Newark, DE 19711, USA.

Received 16 June 2007; revised 17 September 2007; accepted 22 September 2007.
First published online 7 January 2008.
DOI:10.1111/j.1574-6941.2007.00429.x

Editor: Gary King

Keywords: glycosyl hydrolase; polysaccharides; functional genes; carbon cycling; qPCR.

Abstract

The diversity and abundance of glycosyl hydrolase family 5 (GH5) were studied in the North Atlantic Ocean. This family was chosen because of the large number of available sequences from cultured bacteria, the variety of substrates it targets, and the high number of similar sequences in the Sargasso Sea environmental genome database. Three clone libraries of a GH5 subcluster were constructed from the Mid-Atlantic Bight and the eastern and western North Atlantic Ocean. The two North Atlantic Ocean libraries did not differ from each other but both were significantly less diverse than the Mid-Atlantic Bight library. The abundance of GH5 genes estimated by quantitative PCR was positively correlated with chlorophyll concentrations in the eastern part of a transect from Fort Pierce, Florida, to the Azores and in a depth profile, suggesting that the supply of labile organic material selects for GH5-bearing bacteria in these waters. However, the data suggest that only < 1% of all bacteria harbor the GH5 subcluster. These and other data suggest that the hydrolysis of polysaccharides requires complicated multi-enzyme systems.

Introduction

Polysaccharides make up one of the largest identified components of organic carbon in marine environments (Benner et al., 1992; Skoog & Benner, 1997). Molecular analysis of these polysaccharides suggests that they are heteropolymers (Amon & Benner, 2003) originating from phytoplankton (Biersmith & Benner, 1998; Aluwihare & Repeta, 1999) with glucose as the most common monomer (Borch & Kirchmann, 1997). Other polysaccharides that are possibly abundant in marine environments include chitin (Biersmith & Benner, 1998; Aluwihare & Repeta, 1999) with glucose as the most common monomer (Borch & Kirchmann, 1997). Other polysaccharides that are possibly abundant in marine environments include chitin (McCarthy et al., 1997; Aluwihare et al., 2005) and various structural polymers from algae (Giordano et al., 2006). While the monosaccharide composition of polysaccharides in the oceans is known, the linkages and overall structure are not well understood.

Before assimilation of low-molecular-weight byproducts, polysaccharides must be hydrolyzed by glycosyl hydrolases. Glycosyl hydrolase family 5 (GH5) is one of the largest and most diverse families of glycosyl hydrolases (Collins et al., 2005; Cottrell et al., 2005). GH5 enzymes target various substrates, including chitin, mannan, cellulose, xylan, glucan, and lichenin (Coutinho & Henrissat, 1999). Most of the enzymes in this family are from bacteria (60%), with only a few sequences from eukaryotes. The remaining 38% of the genes are from archaea. The remaining 38% of the genes are from eukaryotes. The biochemically characterized GH5 enzymes are mostly from cultured soil bacteria, phytopathogens, and anaerobic bacteria from extreme environments (Lynd et al., 2002; Collins et al., 2005). A few GH5 genes from cultured marine bacteria, however, have been characterized.

Of the cultured marine bacteria known to contain GH5 endoglucanases, most are Gammaproteobacteria (Violot et al., 2003; Garsoux et al., 2004; Taylor et al., 2006). The celG gene from Pseudoalteromonas haloplanktis, which was
isolated from Antarctic seawater, encodes an endoglucanase that is active at cold temperatures (Violot et al., 2003). An initial sequence analysis placed this gene in the GH5 family, and further structural and biochemical analyses confirmed this assignment (Violot et al., 2003; Garsoux et al., 2004). A similar cold-active endoglucanase occurs in a deep-sea psychrophilic bacterium, Pseudoalteromonas sp. DY3, which was isolated from the east Pacific (Zeng et al., 2006). A recent study found that the marine gammaproteobacterium, Saccharophagus degradans, has 12 predicted endoglucanases, 10 of which have GH5 modules (Taylor et al., 2006). Overall, this bacterium has 10 distinct carbohydrate-degrading systems, potentially enabling polysaccharide degradation by a single bacterium rather than by a consortium (Taylor et al., 2006). Although informative, studies of cultured bacteria probably do not cover the diversity of glycosyl hydrolases in the oceans.

The abundance and diversity of functional genes provides insights into the potential of microbial communities to participate in biogeochemical cycles. In contrast to functional genes in the nitrogen cycle (Zehr et al., 2003), diversity studies of genes involved in organic carbon degradation are limited to a few hydrolases, including fungal glycosyl hydrolases (Jacobsen et al., 2005), prokaryotic chitinases (Cottrell et al., 2000; LeCleir et al., 2004; Hobel et al., 2005), and a cellulase (Grant et al., 2004). Recent metagenomic studies have provided sequence data about hydrolases (Cottrell et al., 2005), but these data have not been used extensively so far. Furthermore, to date, only one study has explored the abundance of a glycosyl hydrolase (a chitinase) in an aquatic environment (Xiao et al., 2005). Previous studies have used quantitative PCR (qPCR) to estimate the abundance of a few other functional genes, including nitorgenases (nifH) (Church et al., 2005), nitrate reductases (nasa) (Allen et al., 2005), and anoxygenic photosynthesis genes (pfum) (Schwalbach & Fuhrman, 2005; Waidner & Kirchman, 2007). In general, the diversity and abundance of functional genes in uncultured microorganisms is not well known, especially compared with the 16S rRNA gene.

The aim of this study was to assess the diversity and abundance of β-1,4-endoglucanase genes from the GH5 family in coastal water of the Mid-Atlantic Bight and in the North Atlantic Ocean. The diversity of GH5 genes was examined in three locations using clone libraries of amplicons generated with new degenerate primers for these genes. The abundance of GH5 genes was assayed using qPCR along a transect between Fort Pierce, Florida, and the Azores. It was found that the abundance of GH5 correlated positively with chlorophyll concentration in the eastern part of the North Atlantic Ocean and in a depth profile. It was also found that the diversity of GH5 was greater in coastal water than in the open ocean.

Materials and methods

Sampling and environmental parameters

Samples for bacterial community DNA were collected from the surface of the Mid-Atlantic Bight shelf-break in November 2002 (38° 19.61′N, 73° 28.36′W) and from the surface of a transect in the North Atlantic Ocean in June 2005. Two samples from the transect (32° 59.98′N, 65° 59.85′W and 37° 42.70′N, 38° 54.20′W) and one from the Mid-Atlantic Bight shelf break (38° 19.61′N, 73° 28.36′W) were used for the construction of GH5 gene libraries, and the rest of the samples were used in qPCR analyses. In addition, bacterial community DNA was collected from a depth profile in the western North Atlantic Ocean (29° 46.8′N, 74° 52.2′W).

In order to isolate DNA from the bacterial size fraction, water samples were prefiltered through 3 μm and 0.8 μm pore-size Isopore ATTP filters (Millipore). After prefiltration, bacterial biomass was collected onto a 0.22 μm pore-size Durapore GV filter (Millipore), preserved in sucrose lysis buffer (20 mM EDTA, 400 mM NaCl, 0.75 M sucrose, 50 mM Tris-HCl, pH 9), and stored at −80 °C until further processing. DNA was extracted using the phenol–chloroform method as described previously (Fuhrman et al., 1988). DNA concentrations were estimated with PicoGreen (Invitrogen) using a POLARstar fluorometer (BMG LAB-TECH) and λ DNA (Invitrogen) as a standard. Bacterial production was estimated from 3H-leucine incorporation using the microcentrifuge method (Kirchman, 2001). A conversion factor of 1.5 kgC mole−1 of incorporated leucine was assumed (Kirchman, 2001). Chlorophyll was measured by fluorometry as described elsewhere (Welschmeyer, 1994).

PCR primer design

Eighty-three sequences of bacterial endo-1,4-β-glucanases from family 5 (GH5) were retrieved from the Carbohydrate-Active enZymes (CAZy) database (Coutinho & Henrissat, 1999). The predicted amino acid sequences of these genes were aligned using CLUSTAL W in MEGALIGN (DNASTAR). Sixty-nine sequences could be clustered into eight groups, which were determined by a bootstrap value of over 50 in a neighbor-joining tree analysis. The percent identity of the amino acid sequences in each of these clusters was over 40%. Similar sequences in five of the eight clusters were found in the Sargasso Sea database (Venter et al., 2004) using BLASTX (Table 1). Cluster 1 was chosen for primer design because it had the largest number of sequences from the Sargasso Sea and from cultured marine bacteria. However, due to the high variability in the sequences, only the eight sequences from the Sargasso Sea were used for the primer design. The new primers, GH5-558F1 (5′-GGAWCMCKGARTGGT-CACA-3′) and GH5-959R1 (5′-ATGATATTTTTTGCCAG-3′), bracket a 350-bp fragment of the GH5 gene. This
Table 1. Clusters of GH5 from cultured bacteria according to amino acid similarity

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Phylogenetic group</th>
<th>Number of sequences</th>
<th>Number of similar sequences in the Sargasso Sea database*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Actinobacteria</td>
<td>2</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Gammaproteobacteria</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Firmicutes</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Gammaproteobacteria</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Actinobacteria</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Gammaproteobacteria</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Alphaproteobacteria</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Thermotoga</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fibrobacter</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Firmicutes</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Gammaproteobacteria</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Alphaproteobacteria</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Actinobacteria</td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Gammaproteobacteria</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Actinobacteria</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>Actinobacteria</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Firmicutes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Deinococci</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gammaproteobacteria</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

GHS gene sequences from the CA2y database were compared with sequences in the Sargasso Sea database (Venter et al., 2004) using BLASTX.

*The number of sequences is the total for the cluster, not per phylogenetic group which could not be determined for the Sargasso Sea sequences. The number includes only homologs with a BLAST score above 100.

Three of these 12 Gammaproteobacteria were isolated from the oceans.

fragment is located at the C terminal of the predicted protein and makes up about one third of the gene. It includes one of the two glutamates in the active site (Wang et al., 1993). Numbers in the primer names correspond to their locations on the Sargasso Sea clone AACY01003003.

Cloning and sequencing of Atlantic GH5 genes

Three libraries of GH5 Cluster 1 genes were constructed using primers GH5-588F1 and GH5-959R1. For each library, five PCR reactions (25 μL each) were performed. Each reaction contained 1X PCR buffer, 2.5 mM MgCl₂, 0.2 mM of each dNTP, 0.5 U Taq (Promega), 0.2 pmol of each primer (MWG), and 100 ng DNA. PCR conditions were as follows: 95 °C for 3 min, 30 cycles of (95 °C for 15 s, 50 °C for 15 s, 72 °C for 30 s), 72 °C for 2 min. Five microclones from each reaction were analyzed on a 3% agarose gel to confirm amplification. The remaining 20 μL of five reactions from each location were pooled and concentrated using Microcon YM-100 (Millipore). The concentrated PCR products were subjected to electrophoresis on a 3% low melt agarose gel (SeaKem, BMA). The band at 350 bp was excised and purified using GeneClean (Q-BIOgene) according to the manufacturer’s instructions. PCR products were cloned using the TOPO TA Cloning Kit for sequencing (Invitrogen) into One Shot Mach-1-T1 chemically competent Escherichia coli (Invitrogen) according to the manufacturer’s instructions.

Ninety-six clones were picked from each library, and insert sizes were confirmed by amplification with M13 primers. The PCR products were purified with QiaQuick (Qiagen) (Mid-Atlantic Bight) or Agencourt AMPure magnetic beads (Beckman Coulter) (Western North Atlantic Ocean). Cleaned products were used in sequencing reactions with the BigDye v. 3.1 kit (ABI) according to the manufacturer’s recommendations. Sequencing of the Mid-Atlantic Bight and western North Atlantic libraries was performed on an SCE2410 capillary DNA sequencer (SpectruMedix). Purification and sequencing of the eastern North Atlantic library was performed by the High-Throughput Genomic Unit, Department of Genome Sciences, at the University of Washington.

Diversity analysis

Sequences of 85, 82, and 79 clones from the western North Atlantic, eastern North Atlantic and the Mid-Atlantic Bight, respectively, were suitable for further analysis. The seqman program (DNASTAR) was used to check for base calling mistakes by inspection of the electropherograms. The priming areas from both ends of the sequences were eliminated before analysis. A distance matrix for each library was created using the dnadist program of the phylip package (Felsenstein, 1989). The distance matrix was used in the dotur software (Schloss & Handelsman, 2005) to calculate gene types [analogous to operational taxonomic units (OTU)], diversity indexes, and rarefaction curves. The libraries were compared with each other using UniFrac (Lozupone & Knight, 2005). Neighbor-joining phylogenetic trees of nucleotide acid sequences with bootstrap values based on 100 replicates were constructed with MEGA 3 (Kumar et al., 2004).

The sequence data from these clone libraries have been deposited in Genbank under accession numbers EF679809-EF680053.

qPCR

A new set of nondegenerate primers was designed for GH5 abundance analysis based on the sequences retrieved from the western North Atlantic Ocean library. Two regions with high conservation were used as the priming sites. The GH5-46F (5’-ATTGCTTATACCCTTCACTTTA-3’) and GH5-290R (5’-GTACTTGCGCCAGGAATGACAAT-3’) primers generated a qPCR amplicon of 244 bp. The BACT1 primer set was used to quantify 16S rRNA genes from bacteria (Suzuki et al., 2000).
GH5 and 16S rRNA gene amplifications were performed in 25 µL reactions containing 1X SYBR Green master mix (Stratagene), 40 µM of each primer, and 0.3 µM ROX reference dye. The MgCl₂ concentrations for 16S rRNA gene and GH5 qPCR were 2.5 mM and 4.5 mM, respectively. The DNA concentrations were in the range of 20–200 pg reaction⁻¹. The quadruplicate reactions per sample were analyzed with an ABI 7500 Real Time PCR System (Applied Biosystems). The cycling parameters for GH5 quantification were as follows: 95 °C for 10 min, 50 cycles of (95 °C for 15 min, 50 °C for 0.30 min, 72 °C for 0.45 min), 72 °C for 2 min, followed by a dissociation step to check for nonspecific amplification. The amplification of 16S rRNA genes was the same as GH5, except that only 30 cycles were used, and the annealing temperature was 60 °C. Amplification products for GH5 were cloned and 24 clones were sequenced as described above to confirm that the new primers amplified only GH5 genes.

Genomic DNA of *E. coli* and a GH5 clone from the Mid-Atlantic Bight library were used as standards for the 16S rRNA gene and GH5 in the qPCR, respectively. Several clones from both the Mid-Atlantic Bight and the North Atlantic libraries were tested. The amplification efficiency estimated from the slope (~3.8) of the Cₑ curve was about 83%. The amplification efficiency of 16S rRNA gene was lower, about 70%. The qPCR assays could detect as little as a single copy of both the GH5 and 16S rRNA genes.

Results

Design of degenerate GH5 primers

Gene sequences for β-1,4-endoglucanases of GH5 from cultured bacteria were pooled from the CAZy database (Coutinho & Henrissat, 1999). Multiple sequence alignments (CLUSTAL W) revealed high variability between sequences. The GH5 genes were grouped into eight clusters according to amino acid sequence identity (above 40% identity and bootstrap values over 50). Six clusters contained genes from a diverse array of phylotypic groups (Table 1). However, only one cluster (Cluster 1) had genes from marine bacteria (affiliated with *Gammaproteobacteria*), while the rest of the sequences were found in bacteria from other environments. Sequences from each cluster were used to query the Sargasso Sea database (Venter et al., 2004) with BLASTX. Only three clusters had greater than five significant homologs in the Sargasso Sea database (score above 100) among the Sargasso Sea sequences with Cluster 1 having the largest number (34) (Table 1). Therefore, this cluster was chosen for further investigation.

A neighbor-joining tree was constructed using the predicted amino acid sequences of Cluster 1 and the eight most similar environmental sequences from the Sargasso Sea database (Fig. 1). The predicted amino acid sequences from the Sargasso Sea formed a distinct subcluster. This subcluster was least similar to amino acid sequences from *Firmicutes* and most closely related to enzymes from *Gammaproteobacteria*. These included genes isolated from marine bacteria, such as celG from *Pseudoalteromonas haloplanktis* and celX from *Pseudoalteromonas* sp. DY3. The primer set used for analyses of North Atlantic GH5 diversity was designed based on the Sargasso Sea subcluster.

Diversity analysis of GH5 clone libraries

Three GH5 clone libraries from surface water of the Mid-Atlantic Bight and the North Atlantic Ocean were constructed using the primers GH5-558F1 and GH5-5959R1. Rarefaction analysis at the 97% similarity level suggested that the diversity of GH5 genes in the Mid-Atlantic Bight library was higher than that of the eastern and western North Atlantic libraries (Fig. 2). The Shannon diversity index, which was calculated using the DOTUR software package (Schloss & Handelsman, 2005), indicated a higher diversity of gene sequences in the Mid-Atlantic Bight library than in the North Atlantic libraries as well (Table 2). Additionally, the number of gene types in the Mid-Atlantic Bight library was higher than in the North Atlantic Ocean libraries when percent similarity was 95% or higher (values for 97% similarity are indicated in Table 2). However, the evenness index indicated that gene types are represented more evenly in the Mid-Atlantic Bight library than in either of the North Atlantic libraries. A UniFrac analysis (Lozupone & Knight, 2005) indicated that the GH5-like genes in both North Atlantic Ocean libraries were significantly different from the genes in the Mid-Atlantic Bight library (*P* < 0.05). However, the genes in the two libraries from the North Atlantic Ocean were not different (*P* > 0.05).

A neighbor-joining tree was constructed to further examine the genes amplified in this study and to compare them with sequences from the Sargasso Sea and cultured bacteria (Fig. 3). Sequences that were more than 93% identical at the nucleic acid sequence were grouped together (Fig. 3 and Table 3). Most of the GH5 genes collected during this study could be grouped into two major subclusters (Atlantic GH5 Subclusters 1 and 2) (Fig. 3 and Table 3). The number of sequences in both subclusters was similar in all three libraries. Five additional subclusters (Atlantic GH5 Subclusters 3–7) contained sequences that were also closely related to those from the Sargasso Sea (Fig. 3). Four of these subclusters were composed mostly of sequences from the Mid-Atlantic Bight (Table 3). Four sequences from the Sargasso Sea, which were used for designing the new primers, fell into Atlantic GH5 Subclusters 1 and 6 (Fig. 3). Surprisingly, only one sequence from the three Atlantic GH5
libraries was 100% identical to a sequence from the Sargasso Sea database. Overall, however, BLASTX analysis of all Atlantic GH5 clones against these Sargasso Sea sequences revealed that the percent identity was 93% or higher (average E score of $3.9 \times 10^{-61}$). In contrast, the percent identity between GH5 in this study and the predicted amino acid sequences from cultured bacteria was much lower (46–52% with an average E score of $1.10 \times 10^{-26}$).
Diversity and abundance of glycosyl hydrolase family 5

The abundance of GH5 genes in the North Atlantic Ocean was determined using qPCR with specific primers based on sequences from the western North Atlantic library. The percent of sequences in both eastern and western North Atlantic libraries that were 100% similar to the forward and reverse primer sequences was 80–90%. These primers were used in qPCR assays on samples from a transect in the North Atlantic Ocean between Fort Pierce, Florida, and the Azores. In most locations, the abundance of GH5 genes was 0–100 copies ng\(^{-1}\) DNA, occasionally exceeding 200 copies ng\(^{-1}\) DNA (Fig. 4a and b). Overall, the abundance of GH5 did not correlate with chlorophyll (\(r = 0.08; P > 0.05; n = 17\)). However, when the transect was divided into two parts, GH5 gene abundance was correlated with chlorophyll in the eastern half of the transect (\(r = 0.68; P < 0.05; n = 9\)) and maxima in GH5 abundance coincided with maxima in chlorophyll (Fig. 4b). In contrast, the relationship between GH5 gene abundance and chlorophyll was not significant in the western half of the transect (\(r = 0.37; P > 0.05; n = 8\)) (Fig. 4a).

The percent of bacteria containing GH5 genes was calculated from the ratio of GH5 copies to 16S rRNA gene copies, assuming one copy of GH5 per genome and 1.9 copies of 16S rRNA gene per genome. The first assumption is the average copies per genome of cultured bacteria used in the sequence alignment taken from the CAZy database. The latter assumption was based on the ratio of the 16S rRNA gene to six single copy genes in the Sargasso Sea (Venter et al., 2004) as discussed previously (Campbell et al., 2008). In most locations, < 0.2% of all bacteria harbored this GH5 subgroup in the North Atlantic surface water (Fig. 4c and d).

The correlation coefficients between the percent of bacteria with GH5 and chlorophyll were similar to those for the relationship between GH5 copies ng\(^{-1}\) DNA and chlorophyll (data not shown). About 0.5% of all bacteria had GH5 where chlorophyll was maximal (Fig. 4d).

GH5 gene abundance, bacterial production, and chlorophyll concentrations were also examined in a depth profile in the western North Atlantic (29° 46.8’N, 74° 52.2’W). GH5 copies varied from about 10 to 260 copies ng\(^{-1}\) DNA (Fig. 5a), with maxima at 72 and 130 m. At 72 m, this GH5 maximum coincided with high chlorophyll concentrations (Fig. 5a) and bacterial production (Fig. 5b). The correlation between GH5 copies ng\(^{-1}\) DNA and chlorophyll concentration was significant (\(r = 0.76, P < 0.05; n = 7\)). When the GH5 copies were normalized to 16S rRNA gene copies (Fig. 5b) or to volume filtered (data are not shown), this correlation was even higher (\(r = 0.96, P < 0.05; n = 7\)). The correlation between percent of bacteria with GH5 and bacterial production was also significant (\(r = 0.89; P < 0.05; n = 7\)).

**Discussion**

The specific pathways used by bacteria to process DOM components in the oceans are not fully resolved. Various genes involved in the nitrogen (Braker et al., 2000; Zehr et al., 2003; Jenkins et al., 2006) and sulfur (Cottrell & Cary, 1999; Minz et al., 1999) cycles as well as genes involved in phototrophy (Kolber et al., 2000; Beja et al., 2002; Waidner & Kirchman, 2005) have been studied. However, only a few gene families that participate in the breakdown of polysaccharides have been examined in a few aquatic systems

---

**Table 2. Diversity indices for three GH5-like gene libraries from the Mid-Atlantic Bight, and the western and the eastern North Atlantic**

<table>
<thead>
<tr>
<th>Index</th>
<th>Mid-Atlantic Bight</th>
<th>Western North Atlantic</th>
<th>Eastern North Atlantic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shannon</td>
<td>2.48 (± 0.19)</td>
<td>1.22 (± 0.23)</td>
<td>1.38 (± 0.23)</td>
</tr>
<tr>
<td>Gene types(^1)</td>
<td>17</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Evenness(^1)</td>
<td>0.87</td>
<td>0.58</td>
<td>0.66</td>
</tr>
<tr>
<td>Coverage(^1)</td>
<td>0.96</td>
<td>0.95</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Indices were calculated with sequences grouped at the 97% similarity level. The 95% confidence intervals are in parentheses for the Shannon index.

\(^1\) Western and eastern North Atlantic correspond to libraries that were created with samples from 32° 59.98’N, 65° 59.85’W and 37° 42.70’N, 38° 54.20’W, respectively.

\(^1\) Gene types are similar to operational taxonomic units for phylogenetic markers.

\(^1\) Evenness was calculated as H/nS where H is the Shannon diversity index and S is the number of gene types. Coverage was calculated as 1 – (n/N) and where n is the number of singletons and N is total number of clones.

---

**Abundance of GH5 genes in the North Atlantic Ocean**

The abundance of GH5 genes in the North Atlantic Ocean was determined using qPCR with specific primers based on sequences from the western North Atlantic library. The percent of sequences in both eastern and western North Atlantic libraries that were 100% similar to the forward and reverse primer sequences was 80–90%. These primers were used in qPCR assays on samples from a transect in the North Atlantic Ocean between Fort Pierce, Florida, and the Azores. In most locations, the abundance of GH5 genes was 0–100 copies ng\(^{-1}\) DNA, occasionally exceeding 200 copies ng\(^{-1}\) DNA (Fig. 4a and b). Overall, the abundance of GH5 did not correlate with chlorophyll (\(r = 0.08; P > 0.05; n = 17\)). However, when the transect was divided into two parts, GH5 gene abundance was correlated with chlorophyll in the eastern half of the transect (\(r = 0.68; P < 0.05; n = 9\)) and maxima in GH5 abundance coincided with maxima in chlorophyll (Fig. 4b). In contrast, the relationship between GH5 gene abundance and chlorophyll was not significant in the western half of the transect (\(r = 0.37; P > 0.05; n = 8\)) (Fig. 4a).

The percent of bacteria containing GH5 genes was calculated from the ratio of GH5 copies to 16S rRNA gene copies, assuming one copy of GH5 per genome and 1.9 copies of 16S rRNA gene per genome. The first assumption is the average copies per genome of cultured bacteria used in the sequence alignment taken from the CAZy database. The latter assumption was based on the ratio of the 16S rRNA gene to six single copy genes in the Sargasso Sea (Venter et al., 2004) as discussed previously (Campbell et al., 2008). In most locations, < 0.2% of all bacteria harbored this GH5 subgroup in the North Atlantic surface water (Fig. 4c and d).

The correlation coefficients between the percent of bacteria with GH5 and chlorophyll were similar to those for the relationship between GH5 copies ng\(^{-1}\) DNA and chlorophyll (data not shown). About 0.5% of all bacteria had GH5 where chlorophyll was maximal (Fig. 4d).

GH5 gene abundance, bacterial production, and chlorophyll concentrations were also examined in a depth profile in the western North Atlantic (29° 46.8’N, 74° 52.2’W). GH5 copies varied from about 10 to 260 copies ng\(^{-1}\) DNA (Fig. 5a), with maxima at 72 and 130 m. At 72 m, this GH5 maximum coincided with high chlorophyll concentrations (Fig. 5a) and bacterial production (Fig. 5b). The correlation between GH5 copies ng\(^{-1}\) DNA and chlorophyll concentration was significant (\(r = 0.76, P < 0.05; n = 7\)). When the GH5 copies were normalized to 16S rRNA gene copies (Fig. 5b) or to volume filtered (data are not shown), this correlation was even higher (\(r = 0.96, P < 0.05; n = 7\)). The correlation between percent of bacteria with GH5 and bacterial production was also significant (\(r = 0.89; P < 0.05; n = 7\)).

**Discussion**

The specific pathways used by bacteria to process DOM components in the oceans are not fully resolved. Various genes involved in the nitrogen (Braker et al., 2000; Zehr et al., 2003; Jenkins et al., 2006) and sulfur (Cottrell & Cary, 1999; Minz et al., 1999) cycles as well as genes involved in phototrophy (Kolber et al., 2000; Beja et al., 2002; Waidner & Kirchman, 2005) have been studied. However, only a few gene families that participate in the breakdown of polysaccharides have been examined in a few aquatic systems.
As polysaccharides comprise a major part of the labile DOM in the oceans (Benner et al., 1992), a better understanding of the diversity and abundance of the hydrolytic enzymes involved in processing polysaccharides is necessary.

Sequences of functional genes isolated from environmental DNA are usually different from those of similar genes in cultured bacteria, and GH5 genes are no exception. The GH5 sequences from the North Atlantic Ocean clustered separately from GH5 of cultured bacteria. Other functional genes in the environment that are different from their cultured counterparts include nitrate reductases in sediments (Braker et al., 2000), bisulfate reductases of the microbial community associated with the annelid Alvinella pompejana (Cottrell & Cary, 1999), and various glycosyl hydrolases (Cottrell et al., 2000; LeCleir et al., 2004; Jacobsen et al., 2005). The differences between environmental and cultured bacteria may reflect neutral changes (no effect on enzyme activity) or subtle changes in enzymatic mechanisms and controls. However, a large variation in sequences...
Table 3. Atlantic GH5 groups in the GH5 gene tree (see Fig. 3). The percentage of sequences affiliated with a group from each library is in parentheses.

<table>
<thead>
<tr>
<th>Atlantic GH5 Subcluster</th>
<th>Mid-Atlantic Bight</th>
<th>Western North Atlantic*</th>
<th>Eastern North Atlantic*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>36 (46%)</td>
<td>52 (62%)</td>
<td>47 (57%)</td>
</tr>
<tr>
<td>2</td>
<td>30 (38%)</td>
<td>29 (35%)</td>
<td>23 (28%)</td>
</tr>
<tr>
<td>3</td>
<td>5 (6%)</td>
<td>1 (1%)</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>3 (4%)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>2 (2%)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>2 (2%)</td>
<td>12 (15%)</td>
</tr>
<tr>
<td>7</td>
<td>3 (4%)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>79</td>
<td>84</td>
<td>82</td>
</tr>
</tbody>
</table>

*Western and eastern North Atlantic libraries were created from samples taken at 32° 59.98'N, 65° 59.85'W and 37° 42.70'N, 38° 54.20'W, respectively.

may reflect large differences in function, the extreme being enzymes from cultured and uncultured bacteria that hydrolyze different substrates. Cottrell et al. (2005) found that the sequence of a cellulase from a cultured bacterium was similar to an enzyme from an uncultured bacterium in the Arctic Ocean, but activity assays later revealed that the Arctic enzyme was a peptidase.

The diversity of GH5 genes was higher in coastal waters of the Mid-Atlantic Bight than in the open North Atlantic Ocean, probably due to fundamental differences between the two environments. The North Atlantic Ocean selects for a primary production community dominated by cyanobacteria such as Prochlorococcus and Synechococcus (Lomas & Bates, 2004). In contrast, the Mid-Atlantic Bight is influenced by upwelling and nutrient inputs from terrestrial environments, leading to a dominance of diatoms and dinoflagellates and overall higher productivity (Falkowski

<table>
<thead>
<tr>
<th>Longitude (° W)</th>
<th>GHS5 copies –1 ng DNA</th>
<th>Chlorophyll (µg L⁻¹)</th>
<th>% of bacteria with GH5</th>
<th>Bacterial production (pmol C L⁻¹ h⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>−75</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>−70</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>−65</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>−60</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig. 4. Abundance of GH5 subgroup genes expressed as copies/ng DNA (a, b) or as % of bacteria with the GH5 subgroup, based on 16S rRNA gene (c, d) and chlorophyll concentration along a transect between Florida and the Azores. (a and c) Western part of the transect. (b and d) Eastern part of the transect. Errors were calculated from quadruplicate qPCR assays, but the error bars are smaller than the symbols.
et al., 1994). These differences in the phytoplankton community probably result in variation in polysaccharides in the DOC pool (Myklestad, 1995; Biersmith & Benner, 1998). The composition of aldoses in the DOM produced by diatoms (Skeletonema costatum) and cyanobacteria (Synechococcus bacillaris) is different (Biersmith & Benner, 1998). This variation in the amount and composition of polysaccharides likely explains the diversity of GH5 enzymes that was observed. This diversity may be even higher because particle and phytoplankton-associated bacteria were excluded in order to minimize 16S rRNA genes from plastids.

This difference in gene diversity between aquatic regions is not unique to the current study. Group I chitinases differ from each other in various environments, such as water column and sediments (LeCleir et al., 2004). The variation in nitrite reductase genes isolated from sediments from Puget Sound and from the Washington continental margin was explained by a combination of a higher input of labile organic matter in Puget Sound and large time and distance scales of the continental margin (Braker et al., 2000). Diversity analysis revealed that N$_2$O reductase genes from soils in a forest and neighboring meadow vary, probably due to the vegetation (Rich et al., 2003).

Quantitative estimates of gene abundance may provide some insights into the potential function of bacterial communities and the controls of that function. One hypothesis is that the abundance of functional genes such as GH5 varies independently of environmental parameters and that environmental factors regulate expression of these genes rather than their abundance within the community. For example, even though nitrogenase genes (nifH) are highly diverse in the Chesapeake Bay, only two phylotypes are expressed in that environment (Short & Zehr, 2007), implying selection at the level of transcription rather than the gene level. In contrast, it was observed that the abundance of GH5 genes was partially correlated with chlorophyll in surface water and a depth profile, implying some selection for GH5-bearing bacteria. However, the low correlation between GH5 abundance and chlorophyll in other samples is consistent with the hypothesis that selection acts on gene expression rather than abundance, assuming chlorophyll is an adequate tracer for polysaccharide supply (which may not be the case). The abundance of other functional genes also varies with environmental parameters (Allen et al., 2005; Church et al., 2005; Waidner & Kirchman, 2007). When these genes do correlate with relevant environmental parameters, it seems more likely that gene abundance is an accurate indication of the actual function of a microbial community.

Regardless of location, overall, the estimated abundance of GH5 genes was low. The exclusion of particle- and phytoplankton-associated bacteria may contribute to the low abundance estimates. Another reason for the low estimate is the specificity of the primers used in this study. Only one subcluster of the GH5 family was targeted, although it contained the highest number of genes from cultured bacteria and of similar sequences in the Sargasso Sea database. The relative number of bacteria with genes from the other GH5 subclusters can be estimated by comparing the qPCR results with the distribution of genes in the Sargasso Sea (Table 1). Assuming that the relationship between the qPCR and metagenomic estimates for Cluster 1 applies to the other subclusters, the relative abundance of the other genes would be $c.0.4\%$. Altogether, these calculations suggest that about 1% of all bacteria in the North Atlantic have GH5 genes. It is not possible to calculate the

![Fig. 5. Abundance of GH5 subgroup genes expressed as copies ng$^{-1}$ DNA (a) and % of bacteria with the GH5 subgroup, based on 16S rRNA gene (b) and chlorophyll concentration in a depth profile from the western North Atlantic Ocean (29°46.8’N 74°52.2’W). Errors were calculated from quadruplicate qPCR assays, but the error bars are smaller than the symbols.](image-url)
relative abundance for another glycosyl hydrolase family (chitinases) from the only other study similar to the present one (Xiao et al., 2005), although a metagenomic approach suggested that chitinases are in 0.1–5% of all bacteria in the Delaware Estuary (Cottrell et al., 1999).

In spite of potential limitations due to qPCR primer specificity, it was suspected that the low copy number estimates for GH5 are correct. Even though it is one of the largest glycosyl hydrolase families, the GH5 family is only one of many such families (110 in the CAZY database as of August 2007) (http://www.cazy.org/). Diversity in polysaccharide composition and structure probably leads to diverse glycosyl linkages and thus glycosyl hydrolases. The abundance of other glycosyl hydrolase families has not been studied so far and thus the number of enzymes needed for polysaccharide degradation in the marine environment remains unknown. However, studies of cultured bacteria may shed some light. A whole-genome study suggests that the marine bacterium Saccharophagus degradans has multiple enzymatic systems for polysaccharide degradation and is capable of using various polysaccharides in monocuplute (Taylor et al., 2006). In contrast, various bacteria isolated from seawater could not degrade more than one or two types of polymers, suggesting that several types of bacteria are needed to degrade polymers in the marine environment (Martínez et al., 1996).

This study provides insights into the potential capacity of the microbial community to hydrolyze polysaccharides in marine environments. The diversity analysis suggests that the sequences of these functional genes differ with oceanic regime, which may indicate variability in function. The low abundance of the GH5 genes is probably one indication of the diversity of glycosyl hydrolases that are necessary for hydrolysis of polysaccharides in the oceans. To answer this and other questions, more studies are needed on the abundance and expression of glycosyl hydrolases.

Acknowledgements

The authors would like to thank David Hutchins for the invitation to participate in the North Atlantic Spring Bloom cruise in 2005 and Steve Wilhelm for the chlorophyll data. The authors also thank Thomas Hanson and two anonymous reviewers for comments on early drafts of this manuscript. This work was supported by DOE grant DE-FG02-97ER62479 and NSF grant OCE-0452377. H.E was supported by a Fulbright doctoral fellowship.

References


Coutinho PM & Henrissat B (1999) Carbohydrate-active enzymes: an integrated database approach. Recent Advances...


