

Capturing Single Cell Genomes of Active Polysaccharide Degraders: An Unexpected Contribution of *Verrucomicrobia*

Manuel Martinez-Garcia^{1‡}, David M. Brazel^{1,2}, Brandon K. Swan¹, Carol Arnosti³, Patrick S. G. Chain^{4,5}, Krista G. Reitenga^{4,5}, Gary Xie^{4,5}, Nicole J. Poulton¹, Monica Lluesma Gomez¹, Dashiell E. D. Masland¹, Brian Thompson¹, Wendy K. Bellows¹, Kai Ziervogel³, Chien-Chi Lo^{4,5}, Sanaa Ahmed^{4,5}, Cheryl D. Gleasner^{4,5}, Chris J. Detter^{4,5}, Ramunas Stepanauskas^{1*}

1 Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, Main, United States of America, **2** Colby College, Waterville, Main, United States of America, **3** Department of Marine Sciences, University of North Carolina, Chapel Hill, North Carolina, United States of America, **4** Genome Science Group, Los Alamos National Laboratory, Los Alamos, New Mexico, United States of America, **5** Microbial and Metagenome Program, Joint Genome Institute, Walnut Creek, California, United States of America

Abstract

Microbial hydrolysis of polysaccharides is critical to ecosystem functioning and is of great interest in diverse biotechnological applications, such as biofuel production and bioremediation. Here we demonstrate the use of a new, efficient approach to recover genomes of active polysaccharide degraders from natural, complex microbial assemblages, using a combination of fluorescently labeled substrates, fluorescence-activated cell sorting, and single cell genomics. We employed this approach to analyze freshwater and coastal bacterioplankton for degraders of laminarin and xylan, two of the most abundant storage and structural polysaccharides in nature. Our results suggest that a few phylotypes of *Verrucomicrobia* make a considerable contribution to polysaccharide degradation, although they constituted only a minor fraction of the total microbial community. Genomic sequencing of five cells, representing the most predominant, polysaccharide-active *Verrucomicrobia* phylotype, revealed significant enrichment in genes encoding a wide spectrum of glycoside hydrolases, sulfatases, peptidases, carbohydrate lyases and esterases, confirming that these organisms were well equipped for the hydrolysis of diverse polysaccharides. Remarkably, this enrichment was on average higher than in the sequenced representatives of *Bacteroidetes*, which are frequently regarded as highly efficient biopolymer degraders. These findings shed light on the ecological roles of uncultured *Verrucomicrobia* and suggest specific taxa as promising bioprospecting targets. The employed method offers a powerful tool to rapidly identify and recover discrete genomes of active players in polysaccharide degradation, without the need for cultivation.

Citation: Martinez-Garcia M, Brazel DM, Swan BK, Arnosti C, Chain PSG, et al. (2012) Capturing Single Cell Genomes of Active Polysaccharide Degraders: An Unexpected Contribution of *Verrucomicrobia*. PLoS ONE 7(4): e35314. doi:10.1371/journal.pone.0035314

Editor: Jacques Ravel, Institute for Genome Sciences, University of Maryland School of Medicine, United States of America

Received: November 23, 2011; **Accepted:** March 13, 2012; **Published:** April 20, 2012

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: This research was supported by the NSF grants DEB-841933 and OCE-821374 to RS, OCE-0848703 to CA and by a Maine Technology Institute research infrastructure grant to the Bigelow Laboratory. The Los Alamos National Laboratory researchers were supported in part by the U.S. Department of Energy Joint Genome Institute through the Office of Science of the U.S. Department of Energy under Contract Number DE-AC02-05CH11231 and grants from the U.S. Defense Threat Reduction Agency under contract numbers B1041531 and B0845311. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: rstepanuskas@bigelow.org

‡ Current address: Department of Physiology, Genetics and Microbiology, University of Alicante, Alicante (Spain)

Introduction

Polysaccharides are major components of biomass and detritus in aquatic ecosystems and their microbial degradation constitutes one of the key bottlenecks in the carbon cycle [1,2]. Better understanding of the microbial types and their biochemical machinery involved in the degradation of polysaccharides is also of special interest for cost-effective biofuel production from terrestrial plants and algae [3–5]. Laboratory-based experiments on cultured isolates have been traditional sources of information on polysaccharide-degrading microbial taxa and enzymes, but they represent only a minor fraction of the active players in the carbon cycling in nature [6]. Culture-independent methods, such

as microautoradiography coupled to fluorescent *in situ* hybridization, have provided valuable insights into the uptake rates of some organic compounds by broad microbial phylogenetic groups [7]. More recently, deep metagenomic sequencing has been proven effective in high-throughput discovery of individual polysaccharide hydrolysis genes [8–10]. However, methodological limitations have so far hindered unambiguous identification of microbial taxa responsible for specific hydrolytic processes in the environment and the recovery of entire carbohydrate degradation pathways from members of the microbial “uncultured majority”.

To address this challenge, we developed a novel research approach, which relies on fluorescent labeling of polysaccharides of interest [11] and the use of these polysaccharides in samples

taken directly from the environment to label uncultured microbial cells involved in polysaccharide hydrolysis. Subsequent single-cell genomic DNA amplification and sequencing then yields detailed insight into the metabolic potential of the labeled microorganisms. We employed this approach to analyze freshwater and coastal bacterioplankton for degraders of laminarin and xylan, two of the most abundant storage and structural polysaccharides in nature [12]. Bacterial breakdown of these polysaccharides has been widely demonstrated in aquatic environments [2], but the identity of specific microbes performing this process *in situ* has remained largely unknown [13], due to the challenges outlined above.

Results and Discussion

Overall strategy

Cells that probe positive for a specific polysaccharide are detected and separated from the rest of the natural microbial assemblage by fluorescence-activated cell sorting. Individual, polysaccharide-positive cells are deposited into microplates and subjected to high-throughput single-cell genomic DNA amplification and sequencing [14–18]. We refer to this technique as Fluorescent Substrate Single Amplified Genome Analysis (FS-SAGA).

Optimization of conditions for cell probing with fluorescent polysaccharides

Bacteria-size particles with green fluorescence were detected in aquatic samples that were amended with either 4 or 40 μM fluoresceinamine-labeled laminarin (Figure 1). The number of putative fluorescent cells in 4 μM laminarin treatments increased between 5 and 12 minutes and was stable for the remaining two hours of incubation. The heat-killed control had over 60-fold lower abundance of particles with elevated fluorescence in the gated area compared to the live treatments. No fluorescent particles were detected in the gated area in the live control treatment without the addition of the fluoresceinamine-labeled laminarin. Compared to the 4 μM laminarin treatments, 40 μM treatments had significantly higher background fluorescence, obscuring the demarcation of labeled microbial cells. Thus, a 12–120 minute incubation with 4 μM fluoresceinamine-labeled laminarin was optimal for bacterioplankton probing.

Phylogenetic composition of single amplified genomes

Using a combination of single cell fluorescence-activated cell sorting, whole genome multiple displacement amplification and subsequent PCR and sequencing of the 16S rRNA genes, we generated and identified 414 coastal and 68 freshwater single amplified genomes (SAGs; Figures 2 and S1, Table S1). In both environments, the composition of SAGs generated from cells labeled with the generic DNA stain SYTO-9 was consistent with prior findings of total bacterioplankton composition using other culture-independent techniques. The SAR11 cluster, *Bacteroidetes*, and *Gammaproteobacteria* dominated coastal SAGs (Figure 2B) [19], while *Betaproteobacteria Polynucleobacter* spp., *Actinobacteria* acI, *Alphaproteobacteria* LD12 cluster and *Bacteroidetes* dominated freshwater SAGs (Figure S1) [20]. In contrast, SAGs generated from laminarin-labeled cells were dominated by *Verrucomicrobia* in both coastal and freshwater samples (Figure 2B and Figure S1). Other laminarin-positive cells belonged mostly to the *Bacteroidetes*, *Planctomycetes*, and *Gammaproteobacteria* (Figure 2 and S1, S2). Only 11 coastal and 5 freshwater, xylan-positive SAGs produced SSU rRNA gene sequences. The xylan-positive SAGs were dominated by *Verrucomicrobia* and *Gammaproteobacteria* (including the SAR86

cluster) in the coastal sample and by *Verrucomicrobia* in the freshwater sample.

There was a clear phylogenetic separation between coastal and freshwater *Verrucomicrobia* SAGs, with *Verrucomicrobiaceae* dominating the coastal sample and Subdivision 3 dominating the freshwater sample (Figure 3). *Verrucomicrobia* SAGs grouped into ten marine and five freshwater phylotypes sharing $\geq 99\%$ SSU rRNA gene identity within each phylotype. Of them, one marine phylotype (AAA168-F10) and two freshwater phylotypes (AAA202-P16 and AAA204-K13) comprised over 2/3 of polysaccharide-positive SAGs in their respective environments. Sequences that were identical or closely related to the most abundant marine and freshwater phylotypes (AAA168-F10 and AAA202-P16) have been reported from other environments, indicating that they are broadly distributed and are not limited to the samples analyzed in this study (Figure 3A). Remarkably, none of these phylotypes comprised more than 1% of the total bacterioplankton (HNA and LNA fractions). This corroborates our finding that only $\sim 0.1\%$ of bacterioplankton cells retained laminarin and xylan fluorescence in both aquatic environments. Our results suggest unexpected roles of uncultured *Verrucomicrobia* phylotypes as active laminarin and xylan degraders in the coastal and freshwater environments examined in this study.

To determine the composition of metabolically active members of the studied coastal microbial assemblage, we labeled them with the electron transport system (ETS) activity probe 5-cyano-2,3-ditolyltetrazolium chloride (CTC) and the esterase activity probe carboxyfluorescein diacetate (CFDA), both of which are often used in microbial ecology studies [21,22]. SAGs were generated from the labeled cells and identified by their SSU rRNA gene sequencing. The composition of SAGs generated from esterase- and ETS-positive coastal bacterioplankton was similar to each other and was enriched in *Gamma-* and *Alphaproteobacteria* (*Rhodospirillaceae* and *Rhodobacteraceae*) relative to the total bacterioplankton (Figure 2B, S3). In the coastal sample, the most abundant polysaccharide-positive *Verrucomicrobia* phylotype AAA168-F10 constituted 6% of ETS-positive SAGs (Table S2), providing further evidence that this phylotype was a metabolically active member of the microbial assemblage. Due to the potential toxicity of CTC to some microbial cells, its relevance in microbial ecology studies has been actively debated [21–24]. The compositional similarity between ETS-positive and esterase-positive SAGs observed in this study suggests that both probes detect the same taxonomic groups, likely representing the most metabolically active members of the microbial community.

Whole genome analysis

To verify the potential role of the most abundant polysaccharide-positive *Verrucomicrobia* phylotype AAA168-F10, we performed genomic sequencing of five SAGs representing this phylotype, employing a combination of GAIIx (Illumina) and PacBioTM RS (Pacific Biosciences) sequencing technologies. The obtained assemblies ranged 1.0–4.9 Mbp, with estimated 32%–88% genome recovery (Table S3). The fraction of genome encoding various carbohydrate-active enzymes was almost identical in all five SAGs, and the number of glycoside hydrolase genes correlated with the genome size ($R^2 = 0.93$; Figure S4), indicating that the number of glycoside hydrolases was a function of genome coverage in the five sequenced SAGs. The five SAGs shared high degree of average nucleotide identity (ANI; $>97.8\%$) and similar tetranucleotide signature frequencies (>0.96 ; Figure S5), further confirming that the SAGs were closely related. Therefore, we focused our further annotation efforts on SAG AAA168-F10, which had the largest fraction of the genome recovered. First, we searched for

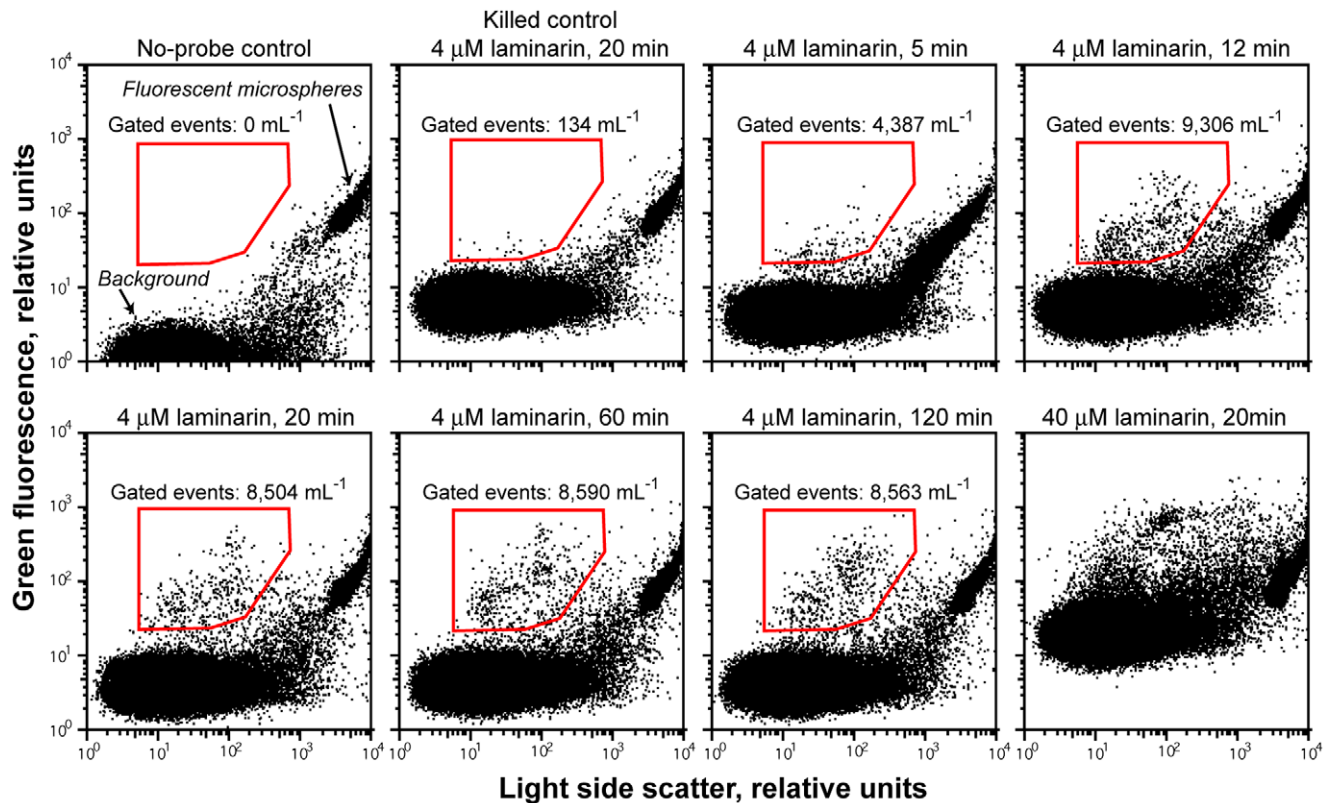


Figure 1. Optimization of cell probing conditions with fluorescently labeled laminarin. Flow cytometric dot plots of heat-killed and live freshwater samples incubated for various lengths of time with 4 or 40 μM fluorescently labeled laminarin. Red polygons indicate gates used to count putative laminarin-positive cells. doi:10.1371/journal.pone.0035314.g001

genes encoding glycoside hydrolases, which catalyze the initial step of converting high molecular weight polysaccharides into oligo- or monosaccharides that are sufficiently small (<600 Da) to be transported into the cell for further processing [25]. We found that *Verrucomicrobia* in general and AAA168-F10 in particular were enriched in glycoside hydrolases (0.91% and 1.2% of total genes, respectively) when compared to the 3,062 publicly available bacterial genomes (Figure 4). On average, about 0.2% of bacterial genes encode glycoside hydrolases. Interestingly, the fraction of these genes in AAA168-F10 and other publicly available *Verrucomicrobia* genomes was on average higher than in *Bacteroidetes*, which are frequently regarded as the most efficient biopolymer degraders [7].

Genome sequence analysis confirmed that AAA168-F10 possesses the genes encoding both laminarinase and xylanase, including their active sites and catalytic residues (Figures 5 and S6). We also detected signal peptide cleavage sites at the N-termini of these proteins, which direct the protein's outward transport across the cellular membrane (Figure S7). The SAG AAA168-F10 genome contained 58 putative glycoside hydrolases representing 15 carbohydrate-active enzyme (CAZy) families [26] (Figure 4). These enzymes are potentially involved in the degradation of complex and diverse biopolymers, including mucopolysaccharides, glycoproteins, peptidoglycan, celluloses, hemicelluloses, and glycogen (Table S4). Furthermore, the AAA168-F10 genome encoded an exceptional number of sulfatases (75 genes; Figure 4). These enzymes have been proposed to be involved in the hydrolysis of sulfate groups to access the carbon skeleton of sulfated polysaccharides, which are major constituents of algal cell

walls [27]. In addition, AAA168-F10 contained a significant number of carbohydrate lyases and esterases (Figure 4), which complement the enzymatic activity of glycoside hydrolases to degrade polysaccharides [26]. We also detected 199 peptidase genes, representing 67 protein families, indicating a vast proteolytic potential (Figure S8). Among detected peptidases are members of the M23B family, which is likely involved in the lysis of bacterial cell wall peptidoglycans [27]. The detected M23B peptidases contained signal peptide cleavage sites indicative of periplasmic or extracellular secretion [28]. Thus, genome sequence analysis provides strong support for the hypothesis that *Verrucomicrobia* phylotypes captured using the FS-SAGA are well equipped for the hydrolysis of diverse polysaccharides and other complex biopolymers.

Prior studies employing cultivation, genomics, metagenomics and radiolabeled DOM uptake assays have suggested the importance of *Bacteroidetes*, *Clostridia*, *Planctomycetes*, *Spirochaetes*, and *Gamma*proteobacteria in polysaccharide degradation in aquatic, soil and cow-rumen environments [7,27,29]. In contrast, very little is known about the metabolism and ecological roles of *Verrucomicrobia*, primarily due to the difficulty in isolation and subsequent paucity of experimental and genomic data. Members of this phylum are widespread in aquatic, terrestrial and intestinal tract environments [30–33] and have been found in association with algae, protozoa, and invertebrate animals [34–36]. To the best of our knowledge, only one prior report shows the ability of *Verrucomicrobia* to degrade polysaccharides, employing cultures isolated from soils [37]. Our study suggests that previously unrecognized, uncultured and relatively rare taxa of *Verrucomicrobia*

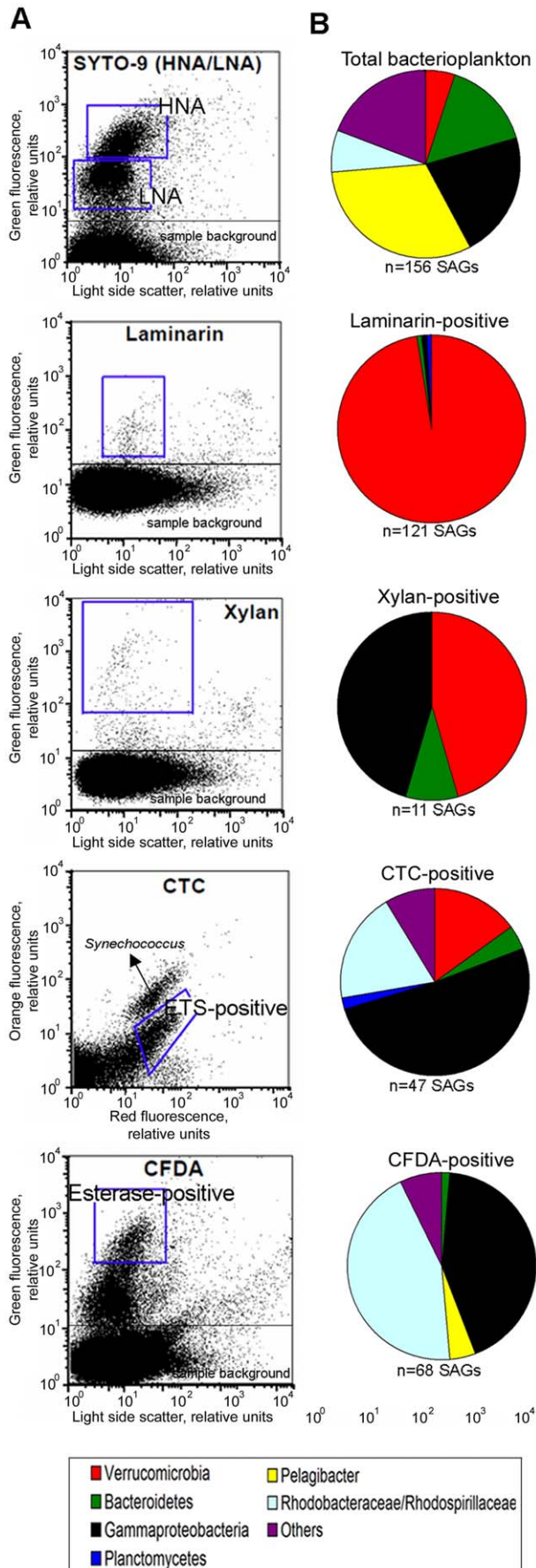


Figure 2. Flow-cytometric sort gates (A) and taxonomic composition (B) of single amplified genomes (SAGs) generated from coastal bacterioplankton using various fluorescent probes. Bacterioplankton were probed with (from top to bottom): 1) nucleic acid stain SYTO-9, targeting high- and low-nucleic acid content cells (HNA and LNA cells) representing a random subset of the entire microbial assemblage; 2) fluorescently-labeled laminarin; 3) fluorescently-labeled xylan; 4) 5-cyano-2,3-ditolyltetrazolium chloride (ETS-active cells) and 5) carboxyfluoresceindiacetate (esterase-active cells). Gates used for cell sorting are indicated in blue. doi:10.1371/journal.pone.0035314.g002

are likely highly active polysaccharide degraders in the studied marine and freshwater environments.

Methodological considerations

The significant enrichment of the sequenced SAGs in genes involved in polysaccharide degradation provides support for FS-SAGA as a useful tool to recover genomes of active polysaccharide degraders, without the need for cultivation. The striking difference in the taxonomic composition of polysaccharide-positive SAGs and esterase- and ETS-positive SAGs provides further evidence that cell labeling by fluorescent polysaccharides targeted microbial groups that express very specific physiological traits rather than general viability. We can rule out the possibility of applied probes labeling *Verrucomicrobia* due to their cell wall peculiarities rather than enzymatic activity, because: a) closely related *Verrucomicrobia* phylotypes, expected to have similar cell wall structure, exhibited highly divergent responses to the same substrate (Figure 3) and b) no cells exhibited fluorescence in killed control treatments (Figure 1).

The FS-SAGA approach offers significant advantages compared to other culture-independent techniques for the discovery and genomic analysis of biopolymer degraders. Compared to metagenomic sequencing, advantages include a) targeting members of natural microbial assemblages that are highly active in the degradation of specific polymers under the studied conditions (*in situ* or manipulated), b) recovery of near-complete genomes, independent of the complexity of the microbial community and the relative abundance of the target taxa, c) a physiology- rather than genetics-based cell targeting, making it independent of existing, limited genetic databases, and d) fast cell probing, removing the risk of biasing microbial composition and confusing primary and secondary responses to the substrate amendment.

The FS-SAGA is not exempt from limitations. We assume that laminarin- and xylan-positive cells retain the fluorescently-labeled polysaccharides on their cell surface, presumably through enzyme-associated carbohydrate binding domains [38]. Some active biopolymer degraders may thus fail to retain the fluorescently-labeled polysaccharide if they lack distinct carbohydrate binding domains, or their enzymes are released into the surrounding medium rather than attached to the cell surface or contained in the periplasm. Moreover, cells that bind a biopolymer but cleave away the fluorescently tagged portion of the polysaccharide may not be labeled despite a high level of activity. Second, taxonomic biases may also be introduced by taxon-specific differences in cell lysis efficiency, SSU rRNA gene primer mismatches during SAG PCR, or interference of fluorescent substrates with downstream molecular analyses. For example, we had low success rate recovering DNA from cells labeled with fluorescent xylan (Figure 2). This may be caused by multiple factors, such as xylan inhibition of cell lysis and DNA amplification, or by some of the sorted fluorescent particles being cellulosome-like enzyme complexes [39] or other non-living particles with associated enzymes. Some support for the latter possibility is provided by the notably

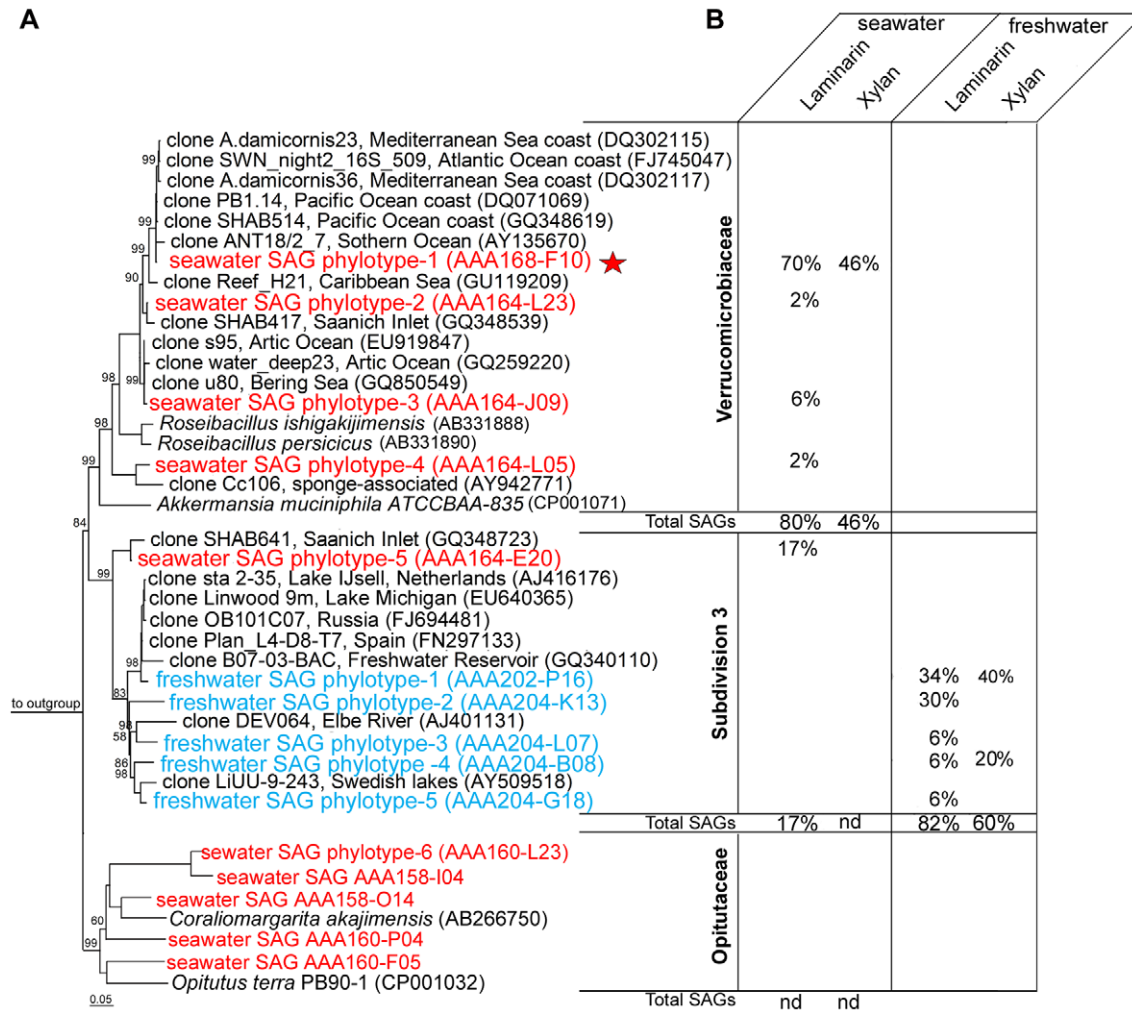


Figure 3. Phylogenetic composition of *Verrucomicrobia* SAGs. (A) Maximum likelihood phylogenetic analysis of the SSU rRNA gene sequences. Bootstrap (1000 replicates) values ≥ 50 are displayed. Each phylotype, indicated in red (coastal) or blue (freshwater) is formed by SAGs with $\geq 99\%$ SSU rRNA gene sequence similarity. Five SAGs from the most abundant putative polysaccharide degrader phylotype in the coastal sample were selected for whole genome sequencing (red star). (B) Phylotype relative abundances in SAG libraries generated using various fluorescent probes. (nd) = not detected in a SAG library. doi:10.1371/journal.pone.0035314.g003

lower light side scatter (a proxy for particle size) among xylan-labeled particles, as compared to laminarin-labeled particles (Figure 2A). Despite these limitations, which may be addressed by future method improvements, FS-SAGA offers a powerful and cost-effective tool to rapidly identify and recover discrete genomes of active players in biopolymer degradation, without the need for cultivation.

Conclusions

We demonstrate the use of FS-SAGA to recover genomes of active laminarin and xylan degraders in coastal and freshwater bacterioplankton, opening new opportunities for basic microbial ecology research and for bioprospecting. Our results indicate unexpected significance in polysaccharide hydrolysis of a few relatively rare, yet widely distributed, planktonic *Verrucomicrobia* phylotypes. The employed method could be readily applied to recover genomes of microorganisms involved in the degradation of diverse polysaccharides in a wide range of environments, utilizing well-established protocols for polysaccharide fluorescent labeling [2] and high-throughput single cell genomics [16–18]. The

spectrum of target substrates may be expanded to other chemical classes, after the development of suitable fluorescent labeling techniques.

Materials and Methods

Optimization of cell probing conditions with fluorescently labeled polysaccharides

A surface water sample was collected from Damariscotta Lake ($44^{\circ}10'38''N$ $69^{\circ}29'12''W$) in Maine, USA, on July 23, 2008 and analyzed within two hours of storage at in situ temperature in the dark. The sample was pre-screened through a $70\ \mu\text{m}$ mesh-size cell strainer (BD), divided into 2 mL aliquots, amended with either $4\ \mu\text{M}$ or $40\ \mu\text{M}$ fluorescein-labeled laminarin (final concentration) and incubated at in situ temperature in the dark. The laminarin was synthesized and labeled with fluoresceinamine as described in detail elsewhere, with about 1 in 148 monomers receiving fluorescent tags [11,40,41]. A subsample of the field sample was brought to boil in a microwave oven, cooled down to room temperature, and then aliquoted and amended with fluorescein-

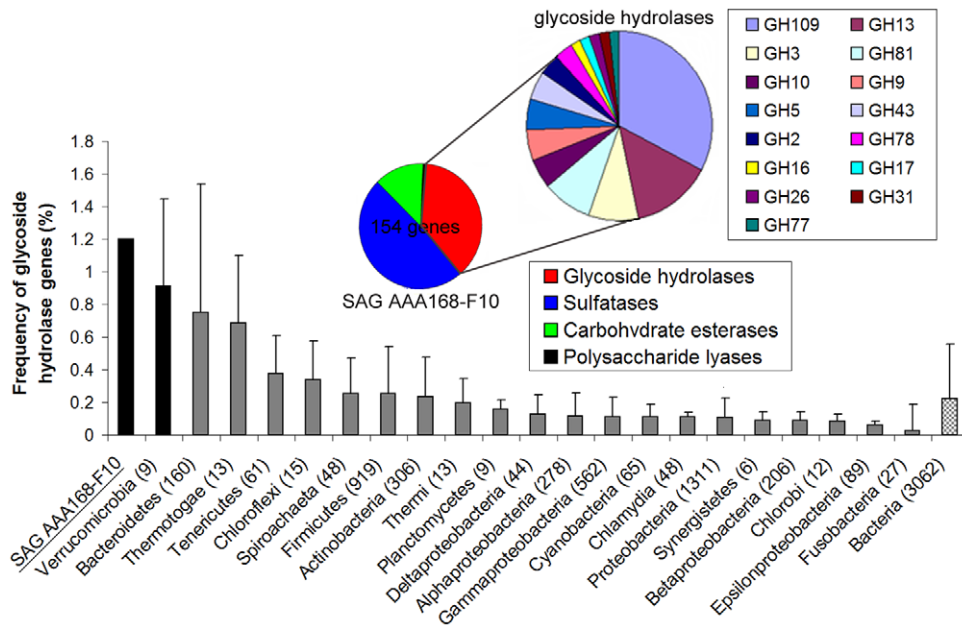


Figure 4. Comparative analysis of genes encoding hydrolytic enzymes in prokaryote genomes. The bar chart indicates the genome-wide frequency of glycoside hydrolase genes in various microbial groups, average \pm standard deviation. The number of publicly available genomes found in the IMG database (as of February 2012) for each taxonomic group is provided in parentheses. The average enrichment of glycoside hydrolases was also estimated for the *Bacteria* domain. The small pie chart shows the number and composition of genes involved in polysaccharide hydrolysis in the *Verrucomicrobia* SAG AAA168-F10. The large pie chart shows CAZY families of glycoside hydrolase genes detected in SAG AAA168-F10. Each glycoside hydrolase family is indicated as GH-xxx, according to CAZY database nomenclature [25]. doi:10.1371/journal.pone.0035314.g004

labeled laminarin as above, to serve as a killed, negative control. After 5, 12, 20, 60 and 120 minutes of incubation, each treatment was analyzed for the abundance of green-fluorescent particles in the prokaryote size range, using light side scatter as a proxy for particle size. Approximately 10^3 mL⁻¹ of 2.15 μ m fluorescent SkyBlue microspheres (Spherotech, Inc., Libertyville, IL) were added to each treatment to serve as internal standards, and their abundance was determined by epifluorescence microscopy. Putative fluorescent microbial cells and fluorescent microspheres were counted in each treatment using a MoFloTM (Beckman Coulter) flow cytometer. The gate for putative fluorescent cells was delineated in the light side scatter interval that is typical for prokaryotes and in the green fluorescence interval above the background fluorescence. The abundance of putative fluorescently labeled cells per mL sample was estimated as the ratio of gated cell-like particles versus the microspheres, multiplied by the abundance of microspheres and corrected for dilution.

Sample collection and cell labeling for the main experiment

Surface water samples were collected from the Gulf of Maine (43°50'40"N 69°38'27"W) and the freshwater, mesotrophic Damariscotta Lake (44°10'38"N 69°29'12"W) in Maine, USA, on July 19, 2009 and analyzed within two hours of storage at *in situ* temperature in the dark. Water samples were pre-screened through a 70 μ m mesh-size cell strainer (BD) and the bacterioplankton cells were labeled, in parallel, using the following fluorescent probes, for subsequent single cell sorting:

1. Fluoresceinamine-labeled polysaccharides laminarin and xylan (4 μ M final concentration, 20–60 min incubation), which were obtained as described above. Only particles with low light side

scatter, likely corresponding to individual prokaryote cells, were sorted.

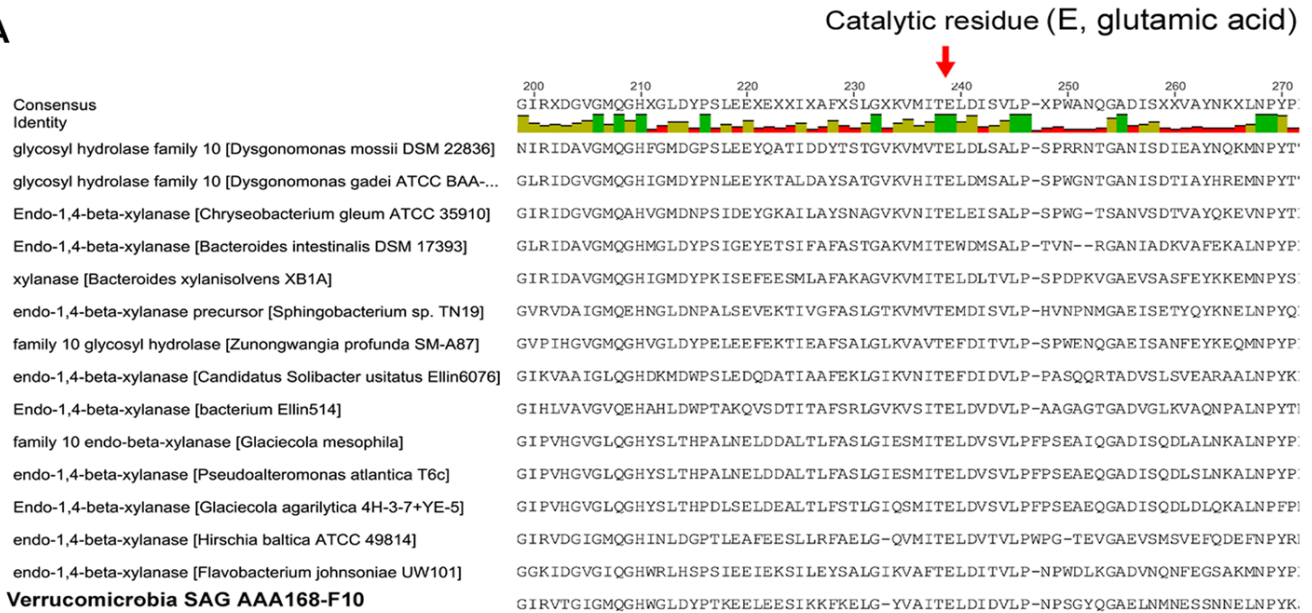
2. SYTO-9 DNA stain (Invitrogen; 5 μ M final concentration; 10–120 min incubation) to label all bacterioplankton cells [15]. The high and low nucleic acid content cells of prokaryotes (HNA and LNA) were sorted and processed separately.
3. The 5-cyano-2,3-ditolyltetrazolium chloride (CTC; Sigma; 5 mM final concentration; 60 min incubation) for detection of prokaryotes with active electron transport system (ETS), indicative of cell's viability [21].
4. The carboxyfluorescein diacetate (CFDA; Invitrogen; 10 μ M final concentration; 20–60 min incubation) for detection of prokaryotes with intracellular esterase activity, as another proxy of cell's viability [23].

No specific permits were required for the described field studies.

Single cell sorting, whole genome amplification and PCR

Microbial cells were sorted with a MoFloTM (Beckman Coulter) flow cytometer equipped with a CyCloneTM robotic arm for droplet deposition into 384-well plates. The cytometer was triggered on side scatter. The “single 1 drop” mode was used for maximal sort purity, which ensures the absence of non-target particles within the target cell drop and the adjacent drops. Under these sorting conditions, sorted drops contain a few 10³ of pL of sample surrounding the target cell [42], so non-target DNA is very low or absent. The accuracy of 10 μ m fluorescent bead deposition into the 384-well plates was verified by microscopically examining the presence of beads in the plate wells. Of the 2–3 plates examined each sort day, <2% wells were found to not contain a bead and only <0.5% wells were found to contain more than one bead, indicating very high purity of single cells. In addition, we verified the lack of DNA contamination in the sheath fluid and in

A



B

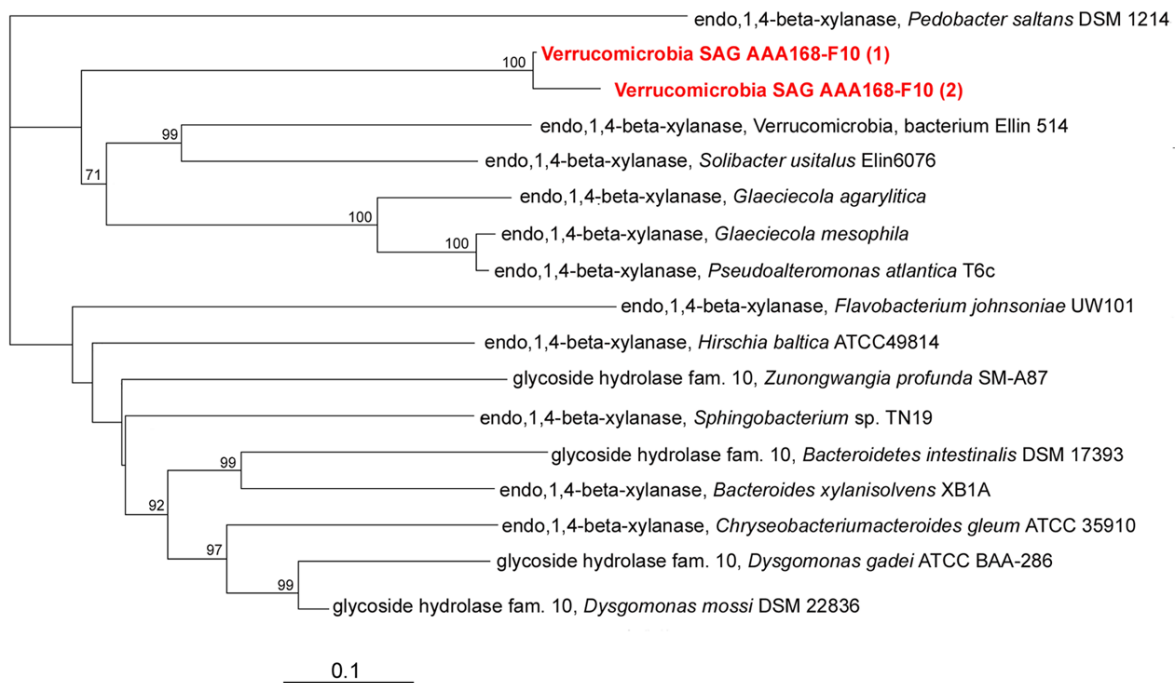


Figure 5. Evidence for the laminarinase gene in the single amplified genome AAA168-F10. (A) Active site, including the catalytic residues responsible for laminarin hydrolysis, derived from Conserved Domain Protein, SWISS-MODEL, and PROSITE databases. (B) Neighbor-joining phylogenetic tree of amino acid sequences, applying the Kimura evolutionary model and indicating bootstrap values above 50. doi:10.1371/journal.pone.0035314.g005

sheath fluid lines by performing real-time multiple displacement amplification with the processed sheath fluid as the template.

Bacterial cells were deposited into 384-well plates containing 0.6 μ L per well of TE buffer. Plates were stored at -80°C until further processing. Of the 384 wells, 315 were dedicated for single cells, 66 were used as negative controls (no droplet deposition) and 3 received 10 cells each (positive controls). The cells were lysed and their DNA was denatured using cold KOH [14]. Genomic DNA

from the lysed cells was amplified using multiple displacement amplification (MDA) [14,43] in 10 μ L final volume. The MDA reactions contained 2 U/ μ L Replphi polymerase (Epicentre), 1 \times reaction buffer (Epicentre), 0.4 mM each dNTP (Epicentre), 2 mM DTT (Epicentre), 50 mM phosphorylated random hexamers (IDT) and 1 μ M SYTO-9 (Invitrogen) (all final concentration). The MDA reactions were run at 30°C for 12–16 h, and then inactivated by 15 min incubation at 65°C . The amplified genomic

DNA was stored at -80°C until further processing. We refer to the MDA products originating from individual cells as single amplified genomes (SAGs). To obtain sufficient quantity of genomic DNA for shotgun sequencing of selected SAGs, the original MDA products were re-amplified using similar MDA conditions as above: eight replicate 125 μL reactions were performed and then pooled together, resulting in ~ 100 μg of genomic dsDNA for each SAG.

The instruments and the reagents were decontaminated for DNA prior to sorting and MDA setup, as previously described [15,44]. Cell sorting and MDA setup were performed in a HEPA-filtered environment. As a quality control, the kinetics of each MDA reaction was monitored by measuring the SYTO-9 fluorescence using FLUOstar Omega (BMG). The critical point (Cp) was determined for each MDA reaction as the time required to produce half of the maximal fluorescence. The Cp is inversely correlated to the amount of DNA template [45]. The Cp values were significantly lower in 1-cell wells compared to 0-cell wells ($p < 0.05$; Wilcoxon Two Sample Test) in each microplate.

The MDA products were diluted 50-fold in sterile TE buffer. Then 0.5 μL aliquots of the dilute MDA products served as templates in 5 μL real-time PCR screens targeting bacterial SSU rRNA genes using primers 27F' and 907R [46,47]. Forward (5'-GTAAAACGACGGCCAGT-3') or reverse (5'-CAGGAAA-CAGCTATGACC-3') M13 sequencing primer was appended to the 5' end of each PCR primer to aid direct sequencing of the PCR products. All PCRs were performed using LightCycler 480 SYBR Green I Master mix (Roche) in a LightCycler[®] 480 II real time thermal cycler (Roche). The real-time PCR kinetics and the amplicon melting curves served as proxies detecting successful target gene amplification. New, 20 μL PCR reactions were set up for the PCR-positive SAGs and the amplicons were sequenced from both ends using M13 targets and Sanger technology by Beckman Coulter Genomics. Single cell sorting, whole genome amplification and real-time PCR screens were performed at the Bigelow Laboratory Single Cell Genomics Center (www.bigelow.org/scgc). Our previous studies and other recent publications using our single cell sequencing technique demonstrate the reliability of our methodology with insignificant levels of DNA contamination in individual cell MDA products [10,15–18,44,48–51].

16S rRNA phylogenetic analysis

The SAG 16S rRNA gene sequences were aligned using the SILVA aligner [52]. Phylogenetic analysis based on maximum likelihood (1000 bootstrap replications) was performed with RAxML version 7.0.3 [53] implemented in ARB package [54], using the reference ARB database 102 containing 460,783 high quality 16S rRNA sequences. The core tree was calculated with the closest reference sequences and then partial sequences from SAGs (742–833 nucleotide positions) were added using the ARB parsimony tool. Those 16S rRNA gene sequences from SAGs that displayed $\geq 99\%$ similarity were grouped into the same phylotype. Quantitative β -diversity analysis was performed to compare the diversity found in the SAG libraries by using the weighted UniFrac model [55]. For that purpose, a neighbor-joining tree (Jukes-Cantor substitution model), including the 16S rRNA gene sequences from SAGs served as the input data for Fast UniFrac analysis. The archaeon *Halobacterium salinarum* (AB074299) served as an outgroup. Genbank accession numbers of the 16S rRNA gene sequences from SAGs are JF488098–JF488633.

Whole genome sequencing

Whole genome sequencing was accomplished using a hybrid approach, combining Illumina short read data with PacBio long read data. One microgram aliquots of amplified single cell genomic DNA were prepared following the Illumina TruSeq DNA Sample Preparation Guide for the GAIIx system (Illumina, Revision A, Nov 2010). The completed libraries were validated using the Qubit (Invitrogen Corporation, Carlsbad, CA) for quantitation. Samples ranged from 37 ng/ μL to 57 ng/ μL . The Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA) was used to determine the size of the PCR enriched fragments for all samples. The size range for the samples was from 320 to 540 base pairs. The libraries were normalized to 10 nM, denatured and diluted to 8 pM in preparation for cluster generation on the Illumina Cluster Station using the Paired End Cluster Generation Kit Version 4. During cluster generation, the SAG libraries were multiplexed onto five lanes of the flowcell, three libraries per lane. The flowcell was run on the Illumina GAIIx using the TruSeq Paired End Sequencing By Synthesis Kit Version 5–GA with a multiplexed recipe for a 110+7+110 cycle run.

For the PacBio RS data, three microgram aliquots of amplified single cell genomic DNA were acoustically sheared in a Covaris E210 (Covaris[®]) to a target fragment size of 2 kb using the shearing conditions provided in the Pacific Biosciences Sample Preparation and Sequencing Guide (Pacific Biosciences, 2010–2011). The protocol for preparing a 2 kb library was subsequently followed, using 1 μg of purified, sheared DNA as starting material. Template concentration was calculated using the Qubit fluorometer and the average size was determined by BioAnalyzer trace analysis and served as input to the Annealing & Binding Calculator v.1.2.1 (Pacific Biosciences, March 2011) to prepare SMRTbell-template annealing and polymerase-template binding reactions, as well as the final dilution of the polymerase-bound template complex for sample plate loading and spike-in of control DNA. Due to the variability of sequence data per SMRT cell, we sequenced 6–20 SMRT cells per sample to achieve estimated genome coverage of at least $10\times$. All cells were sequenced with sequencing movie lengths of 40 minutes. The PacBio reads were filtered to a minimum read length of 100 bp and a minimum read quality score of 0.85.

Assemblies were conducted using the Los Alamos National Laboratory assembly pipeline. Briefly, the Velvet assembler [56] is used for Illumina data using a range of Kmers and coverage cutoffs and the resulting contigs are merged together into a final assembly using in house Perl scripts. This assembly was combined with PacBio data using the PacBio AHA (A Hybrid Assembler) software to incorporate long reads and join contigs. The obtained contigs were subject to another round of assembly, using Sequencher software version 4.10.1 (Gene Codes). Ambiguities were trimmed off the ends and contigs overlapping by at least 100 bp and 98% sequence identity were merged into larger contigs. The resulting draft assemblies were used for subsequent analysis.

To verify the absence of contaminating sequences in the assemblies, tetramer frequencies were extracted from all scaffolds and the Principal Component Analysis (PCA) was then used to extract the most important components of this high dimensional feature matrix [16,48]. Scaffolds representing extremes on the first eight PCs were manually examined for their closest tblastx hits against NCBI nt database, which did not yield any close hits to non-*Verrucomicrobia* genomes, thus providing no evidence of contamination in the assemblies.

Partial genome assemblies of the five sequenced *Verrucomicrobia* SAGs were submitted to Genbank under accession numbers

CAGK00000000, CAGL00000000, CAGM00000000, CAGN00000000, GACO00000000. The raw shotgun sequences of the five *Verrucomicrobia* SAGs were deposited in the NCBI short read archive under accession numbers ERP001168 for Illumina reads and ERP001168 for PacBio reads.

Genome annotation and comparative genomics

Prediction of open reading frames was performed with GenMark [57]. Glycoside hydrolase genes were automatically annotated using the CAZymes Analysis Toolkit applying the association rule learning algorithm [58]. The resulting annotation was carefully revised by using conserved domain BLAST [59], BLASTp against non redundant proteins and the resources of SWISS-MODEL [60], PROSITE [61], and CAZy databases [26]. Bioinformatic resources of the Integrated Microbial Genomes (IMG) system were used to estimate the frequency of glycoside hydrolase genes (E.C. 3.2.1.x; see CAZy database) in the publicly available prokaryote genomes in the IMG database (<http://img.jgi.doe.gov/cgi-bin/m/main.cgi>) as of February 2012. Frequency was calculated for each bacterial genome by dividing the total number of genes annotated as glycoside hydrolases by the total number of annotated genes for that particular genome. Then, the average enrichment of glycoside hydrolases for each bacterial phylum was estimated. Peptidase genes were annotated using MEROPS peptidases database [28].

Estimates of complete genome sizes were obtained using conserved single copy gene (CSCG) analysis [48]. To identify relevant CSCGs, 6 genomes from the *Verrucomicrobia* phylum, currently available at the Joint Genome Institute Integrated Microbial Genomes site [62], were included in the analysis: *Akkermansia muciniphila* ATCC BAA-835, *Coralimargarita akajimensis* DSM 45221, *Methylophilum inferorum* V4, *Opitutus terrae* PB90-1, *Verrucomicrobiales* sp. DG1235, and *Verrucomicrobium spinosum* DSM 4136. Of the COG function distributions listed in these genomes, 273 CSCGs were found to be shared by all 6 finished or draft sequences. Of the 273 identified CSCGs, 87 (31.9%), 151 (55.3%), 168 (61.5%), 199 (72.9%), and 239 (87.6%) were present in the SAGs using rps-blast against the COG database, which correlated with assembly size (1.0 Mb, 2.1 Mb, 2.6 Mb, 3.3 Mb, and 4.9 Mb respectively). The expected genome sizes for each SAG was estimated using the function $G_s = A_s / R_{CSCG}$, where G_s is the expected complete genome size; A_s is the size of the SAG assemblies; R_{CSCG} is the recovery of CSCGs based on COG analysis. Thus, the expected genome sizes of these five SAGs are estimated to be approximately 3.2 Mb, 3.8 Mb, 4.4 Mb, 4.7 Mb, and 5.7 Mb.

Pairwise genome comparisons of average nucleotide identity (ANI) [63] and tetranucleotide signature [64] were performed with Jspecies [65], after genome alignment with MUMmer [66] and BLAST [67].

Supporting Information

Figure S1 Taxonomic composition of freshwater single amplified genomes (SAGs). Bacterioplankton were probed with the nucleic acid stain SYTO-9, representing a random subset of the total microbial assemblage, and with fluoresceinamine-labeled polysaccharides laminarin and xylan. (TIF)

Figure S2 Phylogenetic composition of polysaccharide-positive *Gammaproteobacteria*, *Acidobacteria*, *Bacteroidetes*, *Planctomycetes*, *OP10*, and *OP11*. (A) Maximum likelihood phylogenetic analysis of the SSU rRNA gene sequences. Bootstrap (1000 replicates) values ≥ 50 are displayed. Each

phylogroup, indicated in red (coastal) or blue (freshwater) is formed by SAGs with $\geq 99\%$ SSU rRNA gene sequence similarity. (B) Phylogroup relative abundances in SAG libraries generated using various fluorescent probes. (nd) = not detected in a SAG library. (TIF)

Figure S3 Principal coordinate analysis (PCoA) of weighted pairwise UniFrac distances between SSU rRNA gene sequences from the various coastal bacterial fractions. Included are high nucleic acid content (HNA), low nucleic acid content (LNA), ETS-active and esterase-active cells. A Neighbor-Joining tree employing Jukes-Cantor substitution model served as the input data. (TIF)

Figure S4 Frequency of glycoside hydrolase genes in *Verrucomicrobia* genomes. (A) Relationship between the abundance of glycoside hydrolase genes and genome size in the publicly available *Verrucomicrobia* genomes. (B) Relationship between the abundance of glycoside hydrolase genes and the genome size in the five sequenced *Verrucomicrobia* SAGs. (C) The number of genes encoding carbohydrate-active enzymes, glycoside transferases and sulfatases in the five sequenced SAGs of the phylotype AAA168-F10. (TIF)

Figure S5 Comparative genome analysis of the five sequenced single amplified genomes (SAGs) of the *Verrucomicrobia* phylotype AAA168-F10. Plotted are values of the average nucleotide identity (ANI) and the tetranucleotide frequency signature for each of the pairwise genome comparisons. (TIF)

Figure S6 Evidence for the xylanase gene in the single amplified genome AAA168-F10. (A) Active site, including the catalytic residues responsible for xylan hydrolysis, derived from Conserved Domain Protein, SWISS-MODEL, and PROSITE databases. (B) Neighbor-joining phylogenetic tree of amino acid sequences, applying Kimura evolutionary model and indicating bootstrap values above 50. (TIF)

Figure S7 Signal peptide prediction for laminarinase protein sequence. Prediction of signal peptide was performed with SignalP 3.0 Server. Cleavage site is indicated at the N-terminus of the protein sequence, which is used to direct the protein through the cellular membrane. (TIF)

Figure S8 Peptidase genes encoded by the single amplified genome AAA168-F10. A total of 67 peptidase families were found. Peptidases acting on polypeptides (e.g., family M1) and oligopeptides (e.g., S9), carboxy/aminopeptidases (e.g., M14/M42), dipeptidyl-peptidases (e.g., S15) and endopeptidases (e.g., S01B) are encoded on the AAA168-F10 genome. Annotation was performed using the MEROPS peptidase database. (TIF)

Table S1 Summary of single amplified genomes (SAGs) from which the 16S rRNA gene was recovered. (DOC)

Table S2 Abundance of polysaccharide-positive *Verrucomicrobia* phylotypes among ETS- and esterase-positive, coastal SAGs. (DOC)

Table S3 Assembly statistics. (DOC)

Table S4 Glycoside hydrolase enzymes encoded by SAG AAA168-F10.
(DOC)

Acknowledgments

We thank Michael E. Sieracki and David Emerson for productive discussions and valuable comments.

References

- Biddanda B, Benner R (1997) Carbon, nitrogen, and carbohydrate fluxes during the production of particulate and dissolved organic matter by phytoplankton. *Limnol Oceanogr* 42: 506–518.
- Arnosti C (2011) Microbial extracellular enzymes and the marine carbon cycle. *Ann Rev Mar Sci* 3: 401–425.
- Rubin EM (2008) Genomics of cellulosic biofuels. *Nature* 454: 841–845.
- Zeng X, Danquah MK, Chen XD, Lu Y (2011) Microalgae bioengineering: From CO₂ fixation to biofuel production. *Renew Sust Energ. Rev* 15: 3252–3260.
- Blanch HW, Adams PD, Andrews-Cramer KM, Frommer WB, Simmons BA, et al. (2008) Addressing the need for alternative transportation fuels: The Joint BioEnergy Institute. *ACS Chemical Biology* 3: 17–20.
- Rappe MS, Giovannoni SJ (2003) The uncultured microbial majority. *Ann Rev Microbiol* 57: 369–394.
- Cottrell MT, Kirchman DL (2000) 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.
- Pope PB, Denman SE, Jones M, Tringe SG, Barry K, et al. (2010) Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. *Proc Natl Acad Sci USA* 107: 14793–14798.
- Brulc JM, Antonopoulos DA, Berg Miller ME, Wilson MK, Yannarell AC, et al. (2009) Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc Natl Acad Sci USA* 106: 1948–1953.
- Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, et al. (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* 331: 463–467.
- Arnosti C (2003) Fluorescent derivatization of polysaccharides and carbohydrate-containing biopolymers for measurement of enzyme activities in complex media. *J Chromatogr B* 793: 181–191.
- Kennedy JF, White CA (1988) The plant, algal and microbial polysaccharides. In: Kennedy JF, ed. *Carbohydrate chemistry*. Oxford: Clarendon Press. pp 220–262.
- Alderkamp AC, Van Rijssel M, Bolhuis H (2007) Characterization of marine bacteria and the activity of their enzyme systems involved in degradation of the algal storage glucan laminarin. *FEMS Microbiol Ecol* 59: 108–117.
- Raghunathan A, Ferguson HR, Bornarth CJ, Song WM, Driscoll M, et al. (2005) Genomic DNA amplification from a single bacterium. *Appl Environ Microbiol* 71: 3342–3347.
- Stepanauskas R, Sieracki ME (2007) Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc Natl Acad Sci USA* 104: 9052–9057.
- Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, et al. (2011) Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* 333: 1296–1300.
- Martinez-Garcia M, Swan BK, Poulton NJ, Lluetsma Gomez M, Masland D, et al. (2011) High throughput single cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. *ISME J* doi:10.1038/ismej.2011.84.
- Martinez-Garcia M, Brazel D, Poulton NJ, Swan BK, Lluetsma Gomez M, et al. (2011) Unveiling *in situ* interactions between marine protists and bacteria through single cell sequencing. *ISME J* doi:10.1038/ismej.2011.126.
- Fuhrman JA, Hagstrom A (2008) Bacterial and archaeal community structure and its patterns. In: Kirchman DL, ed. *Microbial ecology of the oceans*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Zwart G, Crump BC, Ageterveld MPK-v, Hagen F, Han S-K (2002) Typical freshwater bacteria: An analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquat Microb Ecol* 28: 141–155.
- Sieracki ME, Cucchi TL, Nicinski J (1999) Flow cytometric analysis of 5-cyano-2,3-ditolyl tetrazolium chloride activity of marine bacterioplankton in dilution cultures. *Appl Environ Microbiol* 65: 2409–2417.
- Gasol JM, Aristegui J (2007) Cytometric evidence reconciling the toxicity and usefulness of CTC as a marker of bacterial activity. *Aquat Microb Ecol* 46: 71–83.
- Hoefel D, Grooby WL, Monis PT, Andrews S, Saint CPv (2003) A comparative study of carboxyfluorescein diacetate and carboxyfluorescein diacetate succinimidyl ester as indicators of bacterial activity. *J Microbiol Meth* 52: 379–388.

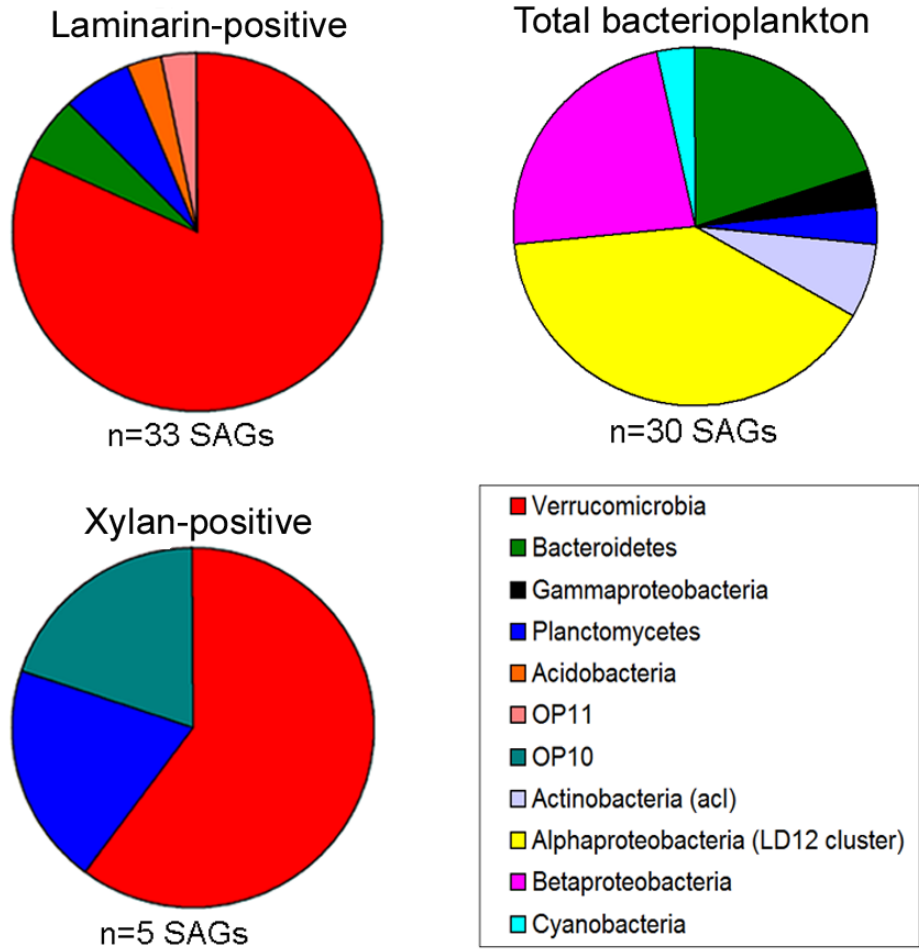
Author Contributions

Conceived and designed the experiments: RS MM. Performed the experiments: RS MM NP PC. Analyzed the data: MM RS PC CA DB. Contributed reagents/materials/analysis tools: RS CA PC KR GX MLG NP DM BT WB KZ CL SA CG CD. Wrote the paper: MM RS DB PC CA BS NP.

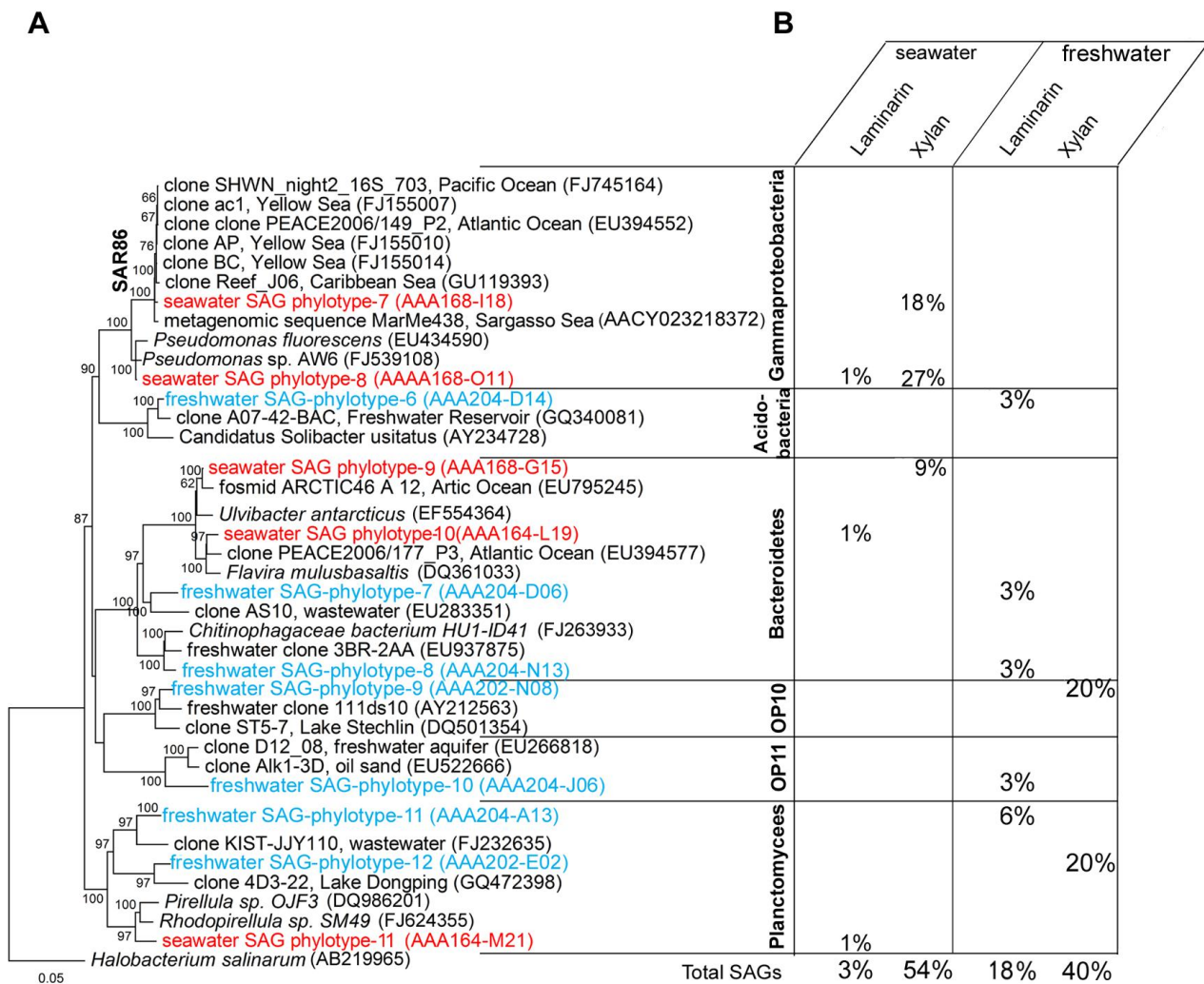
- Ullrich S, Karrasch B, Hoppe HG, Jeskulke K, Mehrens M (1996) Toxic effects on bacterial metabolism of the redox dye 5-cyano-2,3-ditolyl tetrazolium chloride. *Appl Environ Microbiol* 62: 4587–4593.
- Weiss M, Abele U, Weckesser J, Welte W, Schiltz E, et al. (1991) Molecular architecture and electrostatic properties of a bacterial porin. *Science* 254: 1627–1630.
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, et al. (2009) The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acid Res* 37: D233–D238.
- Glöckner FO, Kube M, Bauer M, Teeling H, Lombardot T, et al. (2003) Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. *Proc Natl Acad Sci USA* 100: 8298–8303.
- Rawlings ND, Morton FR, Kok CY, Kong J, Barrett AJ (2008) MEROPS: The peptidase database. *Nucleic Acid Res* 36: D320–D325.
- Weiner RM, Taylor Li LE, Henrissat B, Hauser L, Land M, et al. (2008) Complete genome sequence of the complex carbohydrate-degrading marine bacterium, *Saccharophagus degradans* strain 2-40T. *PLoS Genet* 4.
- Arnds J, Knittel K, Buck U, Winkel M, Amann R (2010) Development of a 16S rRNA-targeted probe set for Verrucomicrobia and its application for fluorescence in situ hybridization in a humic lake. *Syst Appl Microbiol* 33: 139–148.
- Buckley DH, Schmidt TM (2001) Environmental factors influencing the distribution of rRNA from Verrucomicrobia in soil. *FEMS Microbiol Ecol* 35: 105–112.
- Derrien M, Vaughan EE, Plugge CM, de Vos WM (2004) *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *IJSEM* 54: 1469–1476.
- Freitas S, Hatosy S, Fuhrman JA, Huse SM, Mark Welch DB, et al. (2012) Global distribution and diversity of marine Verrucomicrobia. *ISME Journal*, advance online publication.
- Scheuermayer M, Gulder TAM, Bringmann G, Hentschel U (2006) *Rubritalea marina* gen. nov., sp. nov., a marine representative of the phylum ‘Verrucomicrobia’, isolated from a sponge (Porifera). *IJSEM* 56: 2119–2124.
- Bruckner CG, Bahuliker R, Rahalkar M, Schink B, Kroth PG (2008) Bacteria associated with benthic diatoms from Lake Constance: phylogeny and influences on diatom growth and secretion of extracellular polymeric substances. *Appl Environ Microbiol* 74: 7740–7749.
- Petroni G, Spring S, Schleifer K-H, Verni F, Rosati G (2000) Defensive extrusive ectosymbionts of *Euplotidium* (Ciliophora) that contain microtubule-like structures are bacteria related to Verrucomicrobia. *Proc Natl Acad Sci USA* 97: 1813–1817.
- Chin K-J, hahn D, Hengstmann U, Liesack W, Janssen PH (1999) Characterization and identification of numerically abundant culturable bacteria from the anoxic bulk soil of rice paddy microcosms. *Appl Environ Microbiol* 65: 5042–5049.
- Boraston AB, Bolam DN, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochemical Journal* 382: 769–781.
- Bayer EA, Belaich JP, Shoham Y, Lamed R (2004) The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides. *Ann Rev Microbiol* 58: 521–554.
- Arnosti C (1995) Measurement of depth- and site-related differences in polysaccharide hydrolysis rates in marine sediments. *Geochim Cosmochim Acta* 59: 4247–4257.
- Arnosti C (1996) A new method for measuring polysaccharide hydrolysis rates in marine environments. *Org Geochem* 25: 105–115.
- Sieracki M, Poulton N, Crosbie N (2005) Automated isolation techniques for microalgae. In: Andersen R, ed. *Algal Culturing Techniques*. New York: Elsevier Academic. pp 101–116.
- Dean FB, Hosono S, Fang LH, Wu XH, Faruqi AF, et al. (2002) Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci USA* 99: 5261–5266.
- Woyke T, Sczyrba A, Lee J, Rinke C, Tighe D, et al. (2011) Decontamination of MDA reagents for single cell whole genome amplification. *PLoS ONE* 10: e26161.
- Zhang K, Martiny AC, Reppas NB, Barry KW, Malek J, et al. (2006) Sequencing genomes from single cells by polymerase cloning. *Nat Biotechnol* 24: 680–686.
- Casamayor EO, Schaefer H, Baneras L, Pedros-Alio C, Muyzer G (2000) Identification of and spatio-temporal differences between microbial assemblages

- from two neighboring sulfurous lakes: Comparison by microscopy and denaturing gradient gel electrophoresis. *Appl Environ Microbiol* 66: 499–508.
47. Lane DJ (1991) 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M, eds. *Nucleic acid techniques in bacterial systematics*. Chichester, UK: John Wiley.
 48. Woyke T, Xie G, Copeland A, Gonzalez JM, Han C, et al. (2009) Assembling the marine metagenome, one cell at a time. *PLoS ONE* 4: e5299.
 49. Heywood JL, Sieracki ME, Bellows W, Poulton NJ, Stepanauskas R (2011) Capturing diversity of marine heterotrophic protists: one cell at a time. *ISME J* 5: 674–684.
 50. Fleming EJ, Langdon AE, Martinez-Garcia M, Stepanauskas R, Poulton NJ, et al. (2011) What's New Is Old: Resolving the identity of *Leptothrix ochracea* using single cell genomics, pyrosequencing and FISH. *PLoS ONE* 6: e17769.
 51. Yoon HS, Price DC, Stepanauskas R, Rajah VD, Sieracki ME, et al. (2011) Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* 332: 714–717.
 52. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acid Res* 35: 7188–7196.
 53. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.
 54. Ludwig W, Strunk O, Westram R, Richter L, Meier H, et al. (2004) ARB: a software environment for sequence data. *Nucleic Acid Res* 32: 1363–1371.
 55. Hamady M, Lozupone C, Knight R (2009) Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J* 4: 17–27.
 56. Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18: 821–829.
 57. Besemer J, Borodovsky M (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acid Res* 33: W451–W454.
 58. Park BH, Karpinets TV, Syed MH, Leuze MR, Uberbacher EC (2010) CAZymes Analysis Toolkit (CAT): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZY database. *Glycobiology* 20: 1574–1584.
 59. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, et al. (2010) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acid Res* 39: D225–D229.
 60. Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009) The SWISS-MODEL Repository and associated resources. *Nucleic Acid Res* 37: D387–D392.
 61. Sigrist CJA, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, et al. (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acid Res* 38: D161–D166.
 62. Markowitz VM, Chen IA, Palaniappan K, Chu K, Szeto E, et al. (2010) The integrated microbial genomes system: An expanding comparative analysis resource. *Nucleic Acids Research* 38: D382–D390.
 63. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, et al. (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *IJSEM* 57: 81–91.
 64. Teeling H, Meyerdierks A, Bauer M, Amann R, Glöckner FO (2004) Application of tetranucleotide frequencies for the assignment of genomic fragments. *Environ Microbiol* 6: 938–947.
 65. Richter M, Rossello-Mora R (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* 106: 19126–19131.
 66. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, et al. (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5.
 67. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410.

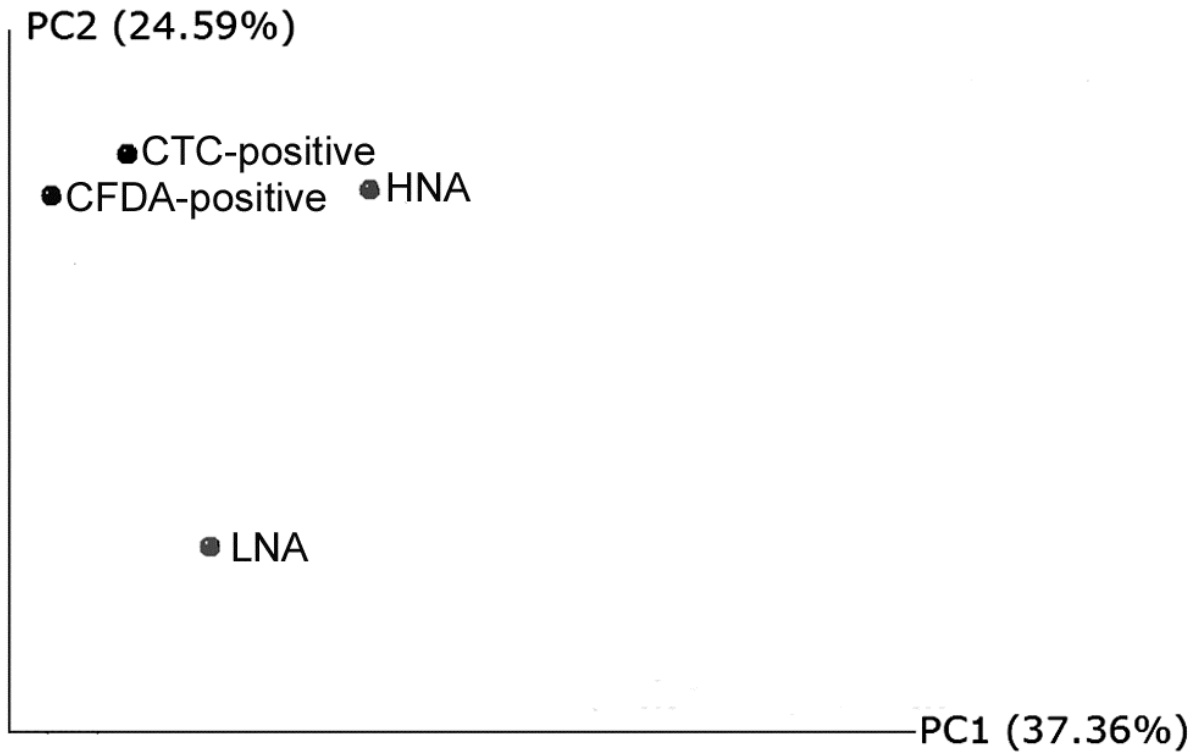
Lake Damariscotta



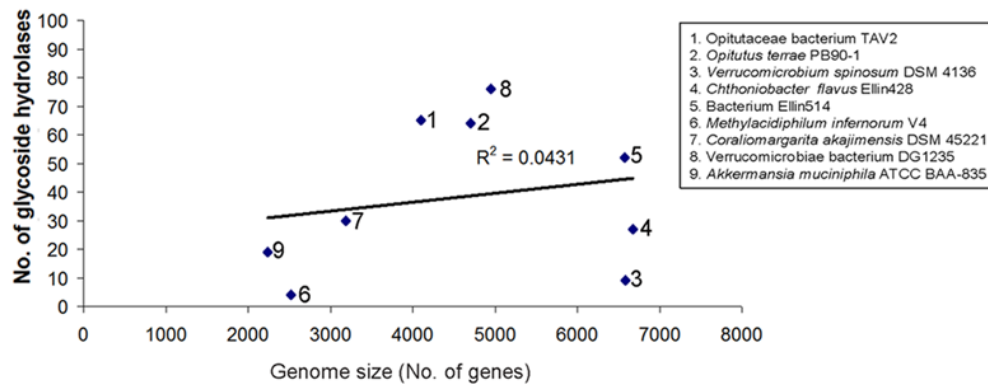
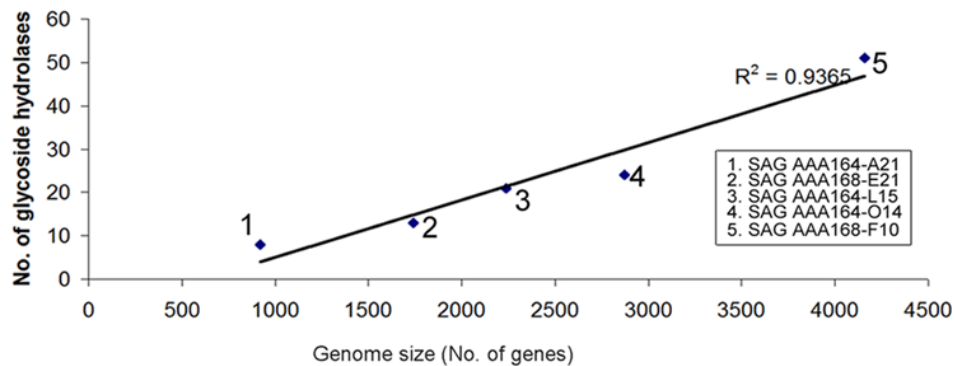
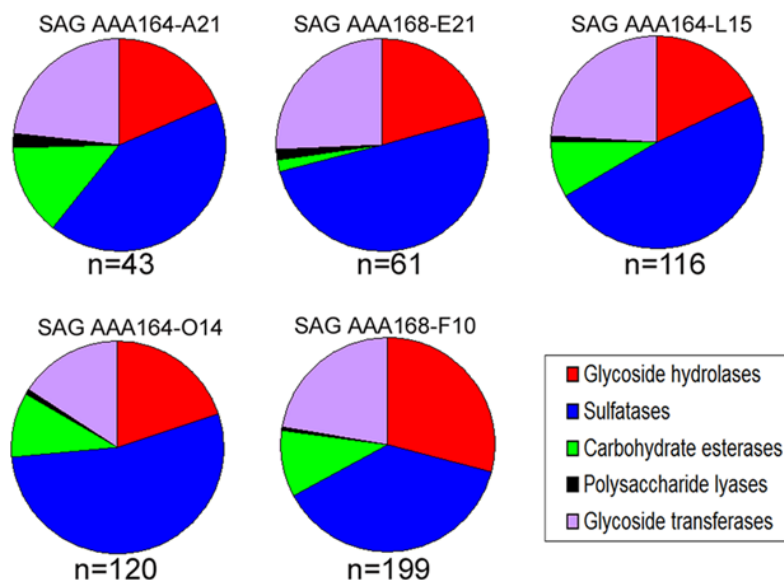
Taxonomic composition of freshwater single amplified genomes (SAGs). Bacterioplankton were probed with the nucleic acid stain SYTO-9, representing a random subset of the total microbial assemblage, and with fluoresceinamine-labeled polysaccharides laminarin and xylan.



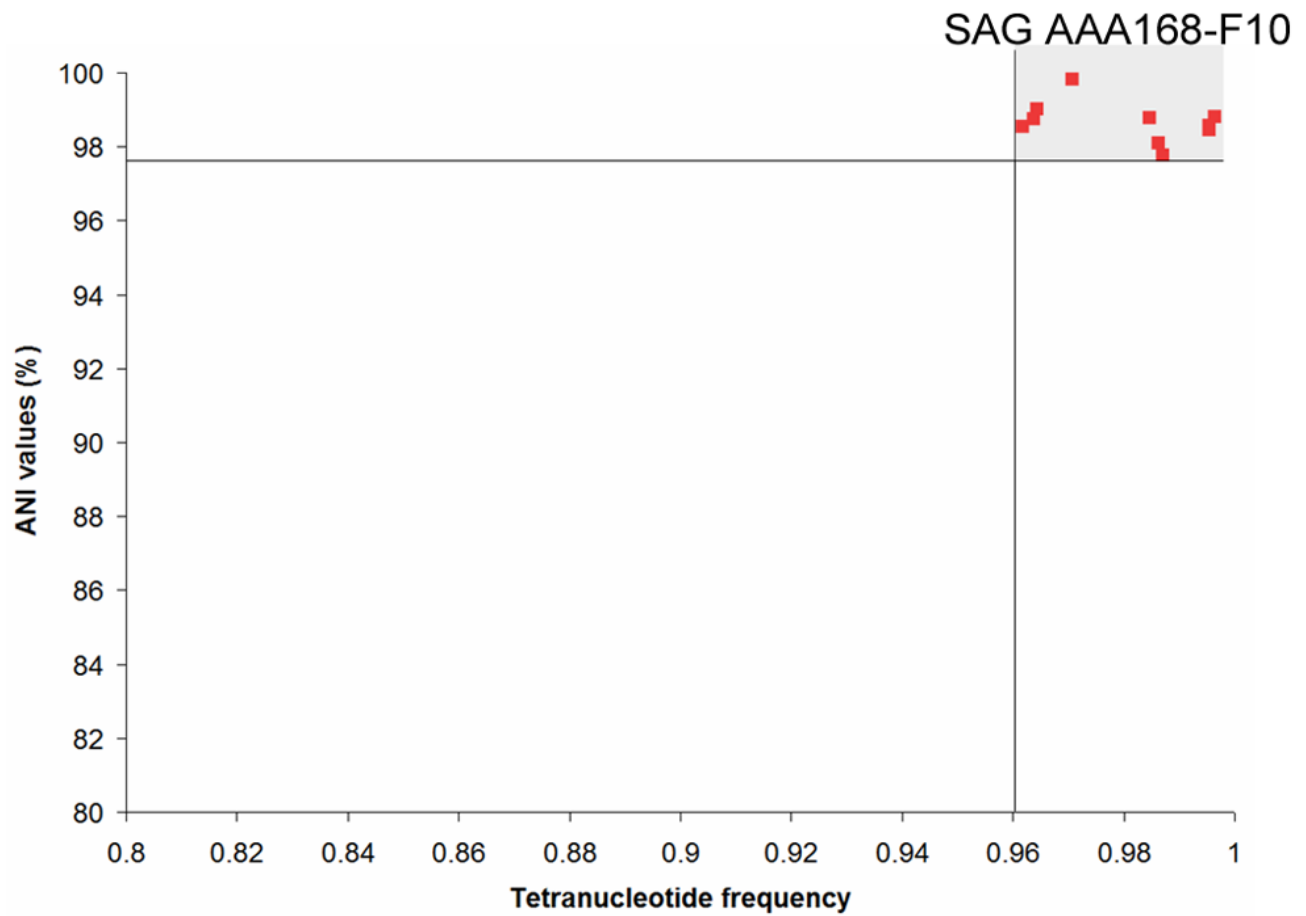
Phylogenetic composition of polysaccharide-positive *Gammaproteobacteria*, *Acidobacteria*, *Bacteroidetes*, *Planctomycetes*, *OP10*, and *OP11*. (A) Maximum likelihood phylogenetic analysis of the SSU rRNA gene sequences. Bootstrap (1000 replicates) values ≥ 50 are displayed. Each phylotype, indicated in red (coastal) or blue (freshwater) is formed by SAGs with $\geq 99\%$ SSU rRNA gene sequence similarity. (B) Phylotype relative abundances in SAG libraries generated using various fluorescent probes. (nd) = not detected in a SAG library.



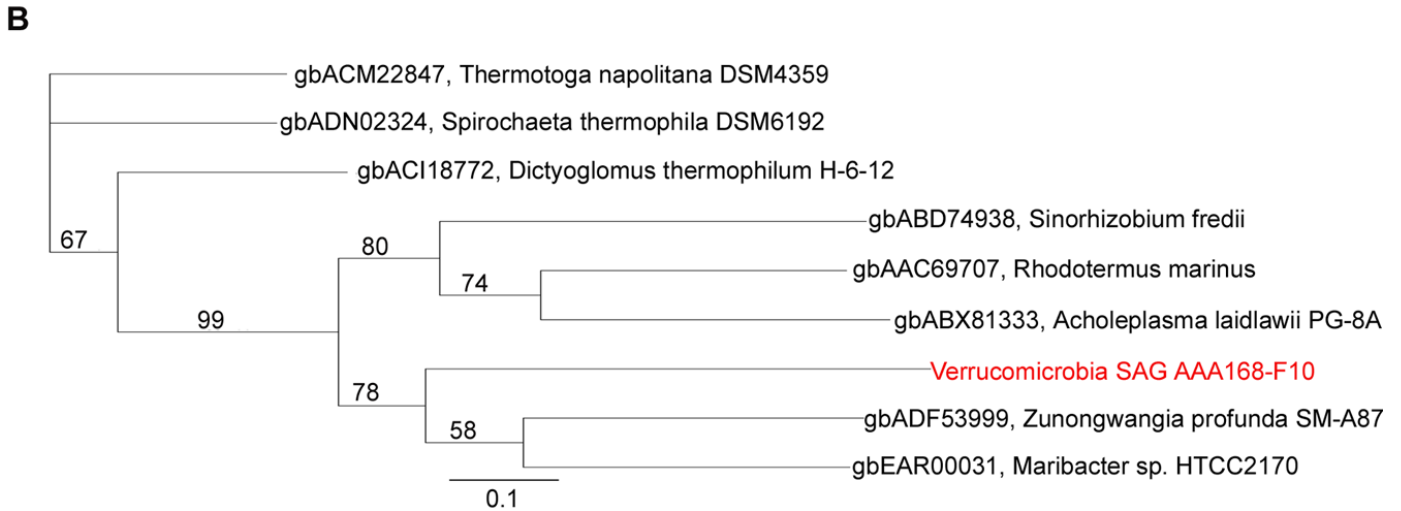
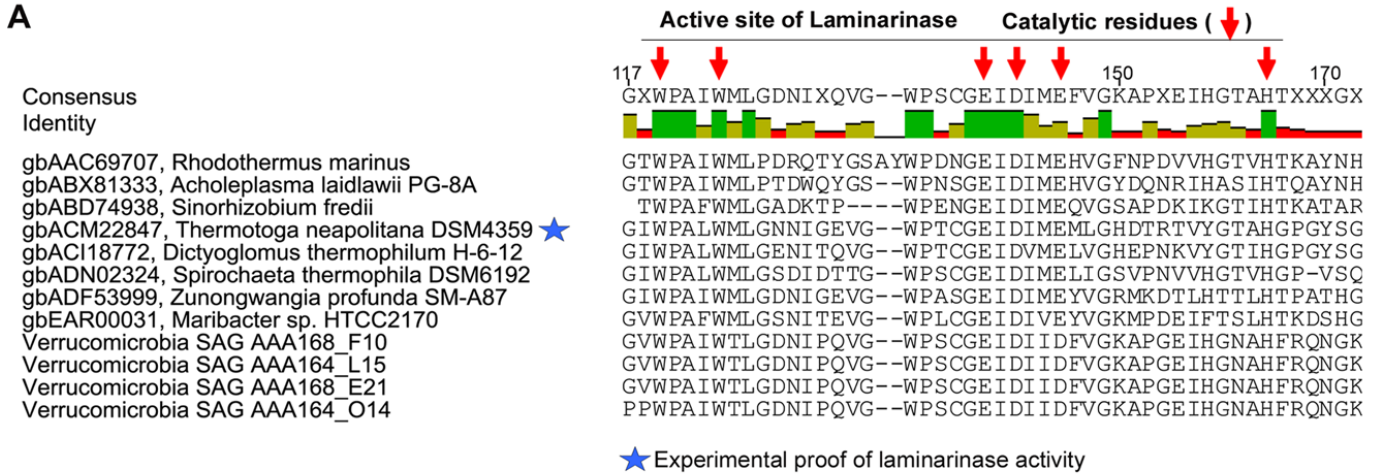
Principal coordinate analysis (PCoA) of weighted pairwise UniFrac distances between SSU rRNA gene sequences from the various coastal bacterial fractions. Included are high nucleic acid content (HNA), low nucleic acid content (LNA), ETS-active and esterase-active cells. A Neighbor-Joining tree employing Jukes-Cantor substitution model served as the input data.

A**B****C**

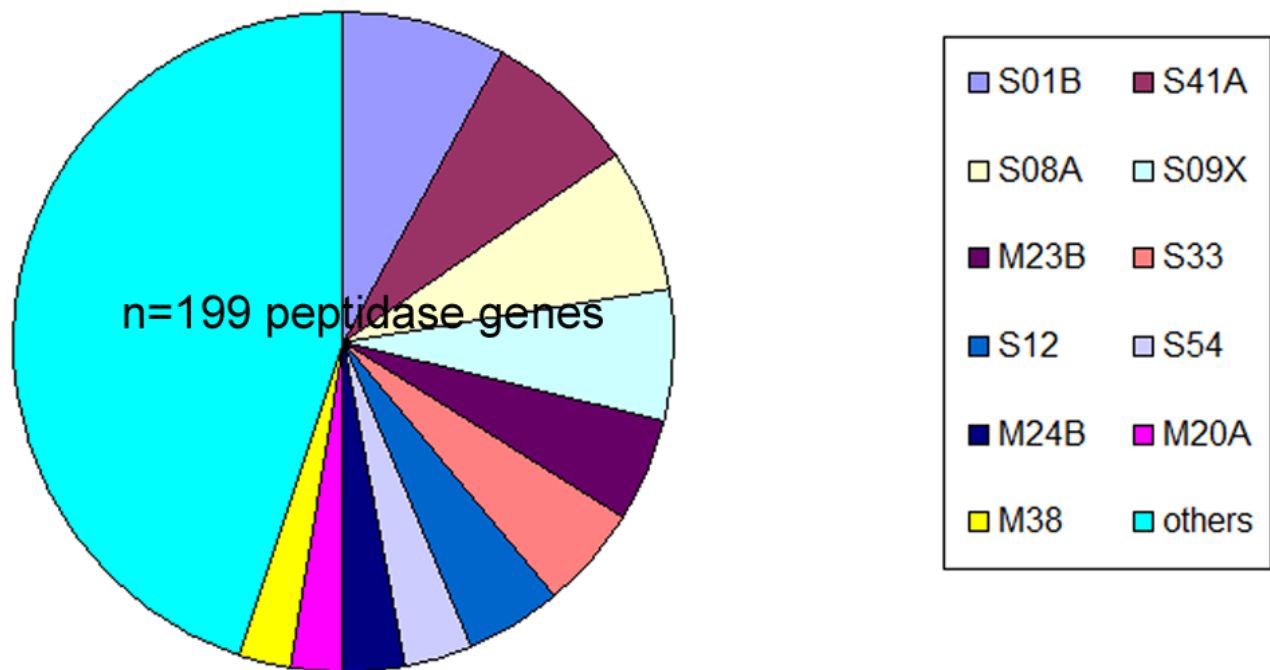
Frequency of glycoside hydrolase genes in *Verrucomicrobia* genomes. (A) Relationship between the abundance of glycoside hydrolase genes and genome size in the publicly available *Verrucomicrobia* genomes. (B) Relationship between the abundance of glycoside hydrolase genes and the genome size in the five sequenced *Verrucomicrobia* SAGs. (C) The number of genes encoding carbohydrate-active enzymes, glycoside transferases and sulfatases in the five sequenced SAGs of the phylotype AAA168-F10.



Comparative genome analysis of the five sequenced single amplified genomes (SAGs) of the *Verrucomicrobia* phylotype AAA168-F10. Plotted are values of the average nucleotide identity (ANI) and the tetranucleotide frequency signature for each of the pairwise genome comparisons.



Evidence for the xylanase gene in the single amplified genome AAA168-F10. (A) Active site, including the catalytic residues responsible for xylan hydrolysis, derived from Conserved Domain Protein, SWISS-MODEL, and PROSITE databases. (B) Neighbor-joining phylogenetic tree of amino acid sequences, applying Kimura evolutionary model and indicating bootstrap values above 50.



Peptidase genes encoded by the single amplified genome AAA168-F10. A total of 67 peptidase families were found. Peptidases acting on polypeptides (e.g., family M1) and oligopeptides (e.g., S9), carboxy/aminopeptidases (e.g., M14/M42), dipeptidyl-peptidases (e.g., S15) and endopeptidases (e.g., S01B) are encoded on the AAA168-F10 genome. Annotation was performed using the MEROPS peptidase database.

Environment	SAG libraries	SAG label	SAGs ^a
Gulf of Maine	Laminarin-positive	(AAA164)	121 (38%)
	Xylan-positive	(AAA168)	11 (4%)
	HNA (Total bacterioplankton)	(AAA160)	106 (34%)
	LNA (Total bacterioplankton)	(AAA163)	50 (16%)
	ETS-positive	(AAA158)	47 (15%)
	Esterase-positive	(AAA076)	68 (21%)
	Total coastal		414 (22%)
Lake Damariscotta	Laminarin-positive	(AAA204)	33 (10%)
	Xylan-positive	(AAA202)	5 (2%)
	HNA (Total bacterioplankton)	(AAA206)	16 (5%)
	LNA (Total bacterioplankton)	(AAA208)	14 (4%)
	Total lake		68 (4%)

^aSAGs yielding SSU rRNA gene sequence, fraction of all single cell MDA products in parentheses.

SAG name	ETS-positive SAG	esterase-positive SAGs
SAG phylotype-1 (AAA168-F10)	6%	nd ^a
SAG phylotype-2 (AAA164-L23)	nd	nd
SAG phylotype-3 (AAA164-J09)	nd	nd
SAG phylotype-4 (AAA164-L05)	2%	nd
SAG phylotype-5 (AAA164-E20)	nd	nd
Total	8%	nd

^aNot detected in the analyzed SAG library

	AAA164-A21	AAA168-E21	AAA164-L15	AAA164-O14	AAA168-F10
Assembly size (bp)	1,011,436	2,060,559	2,641,408	3,346,464	4,850,044
Estimated genome recovery (%)	31.9	55.3	61.5	72.9	87.6
Number of contigs	754	1084	710	1141	1397
Largest contig (bp)	34216	31946	116022	57159	92178
GC content (%)	48	49	49	49	49
Illumina sequencing effort (Mbp)	3276	2582	2187	2705	2554
PacBio sequencing effort (Mbp)	328	476	429	170	430
Total predicted genes	920	1741	2237	2870	4157

CAZy family	No. of genes	Enzyme	Putative substrates	Best BLASTp hit (bit score/ evalue)
GH2	2	(2) β -glucuronidase (EC 3.2.1.31)	β -D-glucuronic acid (glycosaminoglycans/mucopolysaccharides)	<i>Bacteroides</i> sp. ZP_06201167 (114 4e-024)
		(2) β -N-acetylhexosaminidase (EC 3.2.1.52)	Hexosamines	<i>Chlorobium phaeobacteroides</i> YP_912990 (294 / 1e-077)
GH3	5	(2) β -glucosidase (EC 3.2.1.21)	beta-D-glucosides (β -D-galactosides, α -L-arabinosides, β -D-xylosides, β -D-fucosides)	<i>Zunongwangia profunda</i> YP_003585551 (380 / e-103)
		(1) glucan 1,4- β -glucosidase (EC 3.2.1.74)	1,4- β -D-glucans and related oligosaccharides	<i>Maricaulis maris</i> P_757950 (936 / 0.0)
GH5	3	Cellulase family A (including endo-1,4- β -xylanase EC 3.2.1.8)	Cellulose and hemicellulose	<i>Polaribacter irgensii</i> ZP_01118913 (177 / 7e-043)
GH9	3	cellobiohydrolase (EC 3.2.1.91)	cellulose and cellotetraose	uncultured organism ACY24809 (458 / e-126)
GH10	3	endo-1,4- β -xylanase (EC 3.2.1.8)	xylan	Candidatus <i>Solibacter usitatus</i> YP_823955 (217 / 9e-055)
		(1) cyclomaltodextrinase (EC 3.2.1.54)	cyclomaltodextrin	uncultured soil bacterium ABA39290 (474 / e-131)
GH13	8	(7) α -amylase (EC 3.2.1.1)	starch, glycogen and related polysaccharides and oligosaccharides	<i>Croceibacter atlanticus</i> 003715784 (595 / e-168)
GH16	1	glucan endo-1,3- β -glucosidase or laminarinase (EC 3.2.1.39)	laminarin	<i>Zunongwangia profunda</i> SM-A87 YP_003584827.1 (147 / 5e-034)
GH17	1	glucan endo-1,3- β -glucosidase (EC 3.2.1.x)	(1 \rightarrow 3)- β -D-glucans Lichenin	Verrucomicrobiae bacterium ZP_05057481 (165 / 2e-039)
GH26	1	β -mannanase (EC 3.2.1.78)	mannans, galactomannans and glucomannans	<i>Rhodothermus marinus</i> YP_003289757 (260 / 6e-067)
GH31	1	α -glucosidase (EC	Polysaccharides with	<i>Lacinutrix</i> sp.

		3.2.1.20)	(1→4)- α -glucosidic linkages	AEH02520 (836 / 0.0)
GH43	3	xylosidase/arabinosidase (EC 3.2.1.37, EC 3.2.1.55)	xylan , arabinans	<i>Bacteroides plebeius</i> ZP_03207375 (674 / 0.0)
GH77	1	Amylomaltase or 4-alpha-glucanotransferase (EC 2.4.1.25)	maltose	<i>Geobacter sulfurreducens</i> NP_952235 (182 / 1e-044)
GH78	2	α -L-rhamnosidase (EC 3.2.1.40)	α -L-rhamnose	<i>Rhodopirellula baltica</i> EGF24665 (79 / 3e-013)
GH81	5	endo- β -1,3-glucanase (EC 3.2.1.39)	laminarin	<i>Micromonospora aurantiaca</i> (YP_003836446) (198 / 5e-049)
GH109	19	α -N-acetylgalactosaminidase (EC 3.2.1.49)	N-acetylgalactosamine (glycoproteins from cellular surface and cell wall)	<i>Akkermansia muciniphila</i> YP_001877822 (313 / 3e-83)