

Metagenomic *De Novo* Assembly of an Aquatic Representative of the Verrucomicrobial Class *Spartobacteria*

Daniel P. R. Herlemann,^a Daniel Lundin,^b Matthias Labrenz,^a Klaus Jürgens,^a Zongli Zheng,^c Henrik Aspeborg,^d Anders F. Andersson^b

Leibniz Institute for Baltic Sea Research Warnemünde (IOW), Rostock, Germany^a; KTH Royal Institute of Technology, Science for Life Laboratory, School of Biotechnology, Division of Gene Technology, Solna, Sweden^b; Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden^c; KTH Royal Institute of Technology, AlbaNova University Center, School of Biotechnology, Division of Glycoscience, Stockholm, Sweden^d

ABSTRACT The verrucomicrobial subdivision 2 class *Spartobacteria* is one of the most abundant bacterial lineages in soil and has recently also been found to be ubiquitous in aquatic environments. A 16S rRNA gene study from samples spanning the entire salinity range of the Baltic Sea indicated that, in the pelagic brackish water, a phylotype of the *Spartobacteria* is one of the dominating bacteria during summer. Phylogenetic analyses of related 16S rRNA genes indicate that a purely aquatic lineage within the *Spartobacteria* exists. Since no aquatic representative from the *Spartobacteria* has been cultured or sequenced, the metabolic capacity and ecological role of this lineage are yet unknown. In this study, we reconstructed the genome and metabolic potential of the abundant Baltic Sea *Spartobacteria* phylotype by metagenomics. Binning of genome fragments by nucleotide composition and a self-organizing map recovered the near-complete genome of the organism, the gene content of which suggests an aerobic heterotrophic metabolism. Notably, we found 23 glycoside hydrolases that likely allow the use of a variety of carbohydrates, like cellulose, mannan, xylan, chitin, and starch, as carbon sources. In addition, a complete pathway for sulfate utilization was found, indicating catabolic processing of sulfated polysaccharides, commonly found in aquatic phytoplankton. The high frequency of glycoside hydrolase genes implies an important role of this organism in the aquatic carbon cycle. Spatiotemporal data of the phylotype's distribution within the Baltic Sea indicate a connection to *Cyanobacteria* that may be the main source of the polysaccharide substrates.

IMPORTANCE The ecosystem roles of many phylogenetic lineages are not yet well understood. One such lineage is the class *Spartobacteria* within the *Verrucomicrobia* that, despite being abundant in soil and aquatic systems, is relatively poorly studied. Here we circumvented the difficulties of growing aquatic *Verrucomicrobia* by applying shotgun metagenomic sequencing on a water sample from the Baltic Sea. By using a method based on sequence signatures, we were able to *in silico* isolate genome fragments belonging to a phylotype of the *Spartobacteria*. The genome, which represents the first aquatic representative of this clade, encodes a diversity of glycoside hydrolases that likely allow degradation of various complex carbohydrates. Since the phylotype cooccurs with *Cyanobacteria*, these may be the primary producers of the carbohydrate substrates. The phylotype, which is highly abundant in the Baltic Sea during summer, may thus play an important role in the carbon cycle of this ecosystem.

Received 4 December 2012 Accepted 26 April 2013 Published 28 May 2013

Citation Herlemann DPR, Lundin D, Labrenz M, Jürgens K, Zheng Z, Aspeborg H, Andersson AF. 2013. Metagenomic *de novo* assembly of an aquatic representative of the verrucomicrobial class *Spartobacteria*. *mBio* 4(3):e00569-12. doi:10.1128/mBio.00569-12.

Editor Farooq Azam, Scripps Institution of Oceanography, UCSD **Invited Editor** Meinhard Simon, University of Oldenburg

Copyright © 2013 Herlemann et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Noncommercial-ShareAlike 3.0 Unported license](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Anders F. Andersson, anders.andersson@scilifelab.se.

Representatives of the bacterial phylum *Verrucomicrobia* are morphologically diverse and are present in various terrestrial and aquatic habitats, including oligotrophic, eutrophic, extreme, polluted, and manmade ones (1). The few existing isolates have been collected from diverse ecological niches, including soil, freshwater, marine habitats, and feces, displaying aerobic, facultative anaerobic, or obligate anaerobic heterotrophic lifestyles. These *Verrucomicrobia* cultivars have the capacity to utilize various carbon compounds, such as plant polymers, for example, cellulose, xylan, and pectin (2, 3), and sugars or methane (4, 5). In addition, *Verrucomicrobia* have been found to be involved in nitrogen fixation in termites (6) and soil (7).

A recent 16S rRNA study of 181 soils, sampled across different soil types and continents, found *Verrucomicrobia* to be highly

abundant, averaging 23% of the 16S rRNA gene sequences per sample (8). This by far exceeded previous estimates, which was attributed to primer mismatches in commonly used 16S rRNA primers (8). In most of these samples, the *Verrucomicrobia* were dominated (92% in total) by sequences belonging to subdivision 2 class *Spartobacteria* (8). This class comprises one of the primary lineages in the phylum *Verrucomicrobia* (9) but has currently only one cultivated representative, *Chthoniobacter flavus*, an aerobic heterotrophic bacterium isolated from pasture soil that is able to grow on carbohydrate components of plant biomass (10). Although the majority of the *Spartobacteria* appear to be soil inhabiting (9), they have also been detected as endosymbionts of nematode worms (“Xiphinematobacter” [11]) and in aquatic environments (e.g., see references 12 to 15). In a global investiga-

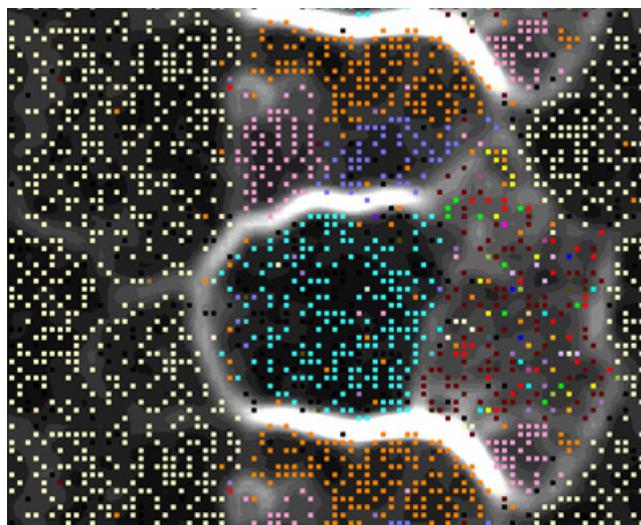
tion of the distribution and diversity of marine *Verrucomicrobia*, *Spartobacteria* were negatively correlated with salinity and were the dominant group in most areas with low salinities (16). The authors speculated that aquatic *Verrucomicrobia* play a significant role in aquatic ecosystems. Although it is well known that bacteria are the main consumers of organic matter in the aquatic environment, the specific roles of individual organisms are not well understood (17). Key enzymes for degradation of polysaccharides derived from plant/algae biomass are glycoside hydrolases (GHs) containing single or multiple catalytic modules frequently attached to one or more accessory noncatalytic carbohydrate binding modules (CBMs) (18, 19). This class of enzymes is well represented in soil-isolated *Verrucomicrobia*, including *Chthoniobacter flavus* (20).

We recently performed a 16S rRNA gene sequence analysis of 213 samples collected in summer that spanned the entire salinity range of the Baltic Sea (21). This revealed pronounced shifts in the bacterial community at different phylogenetic levels along the salinity gradient. The pyrosequencing reads in the brackish water environment (between salinities 5 and 8) were dominated (>10% of reads in many samples) by an operational taxonomic unit (OTU) belonging to the *Spartobacteria*. Despite elaborate attempts, no aquatic *Spartobacteria* have been isolated (e.g., see reference 22) and no genome has been sequenced so far. To get a better understanding of the physiology and ecological role of this organism, and of aquatic *Spartobacteria* in general, we applied shotgun metagenomics to a sample with high abundance of the *Spartobacteria* OTU, resulting in the first reconstruction of an aquatic *Spartobacteria* genome. The metagenome analysis revealed a rich repertoire of glycoside hydrolases, and the spatiotemporal distribution of the OTU suggests a connection to phytoplankton-derived polysaccharides.

RESULTS AND DISCUSSION

Metagenome of “*Spartobacteria baltica*” bin. Metagenomic sequencing of the surface water sample yielded 37,658,923 bp of 454 pyrosequencing data that was assembled into 58,176 contigs. In order to isolate contigs belonging to the target genome, we used an emerging self-organizing map (ESOM) approach that clusters genome fragments into phylogenetic groups based on tetranucleotide frequency distributions (23, 24). Since our previous 16S rRNA gene sequencing indicated that *Verrucomicrobia* are highly dominated by a single operational taxonomic unit (OTU) in this sample (21), the risk for extensive clustering of other verrucomicrobial genomes was considered low. The resulting ESOM contained a distinct region highly enriched in contigs with best BLAST matches to *Verrucomicrobia* (Fig. 1). The region was separated by a ridge, indicating large differences in tetranucleotide frequencies, from surrounding areas containing *Actinobacteria*, *Bacteroidetes*, *Proteobacteria*, and *Cyanobacteria* and a heterogeneous area containing a mixture of clades. The contigs within the region were assigned to a “*Spartobacteria baltica*” bin.

The “*Spartobacteria baltica*” metagenome bin contained 334 contigs with an average length of 5.4 kb (7.0× mean coverage) and a total of 1,811,214 bp (Table 1). The contigs had a relatively high GC content (62%), which is in the range of currently sequenced *Verrucomicrobia* genomes (25), with the exception of the methanotrophic thermophile “*Candidatus* Methylocidiphilum infernum.” The “*Spartobacteria baltica*” bin carries 2,226 predicted genes, of which 2,189 (98%) encode proteins. Only one contig



■ *Verrucomicrobia* ■ *Actinobacteria* ■ *Cyanobacteria*
 ■ *Proteobacteria* ■ *Bacteroidetes*

FIG 1 Emerging self-organizing map of the metagenome contigs. Pixels are colored according to the taxonomic annotation of the contig(s) that occupies the pixel. Background color represents the distance in data space between the pixels in the neighborhood; hence the white ridges represent borders between regions of highly dissimilar tetranucleotide frequency distributions.

with a 16S rRNA gene was found. Since the sequence coverage of the 16S rRNA gene (9×) was not significantly above the average of the “*Spartobacteria baltica*” bin (7×), the genome is likely to encode a single ribosomal RNA operon. The V3-V4 region of the 16S rRNA gene was identical to the sequence of the *Spartobacteria* OTU that we previously detected in high abundance in this sample (21). The 16S rRNA gene also had high similarity to cloned 16S sequences (97 to 99% identity) obtained earlier in the central Baltic Sea (22). Thirty-three genes encoding tRNAs for 17 standard amino acids were found (see Table S1 in the supplemental material). Of the protein-coding genes, 1,533 (69%) were functionally predicted, 621 (28%) were assigned to KEGG maps, 1,404 (63%) were assigned to clusters of orthologous groups (COGs), and 1,443 (65%) were assigned to specific domains in the Pfam database. In comparison with *Chthoniobacter flavus*, “*Spartobacteria baltica*” has a high overall functional assignment of the predicted proteins (Table 1).

To assess the completeness and purity of the “*Spartobacteria baltica*” bin, we used a set of 40 housekeeping genes known to normally occur in single copies (see Table S2 in the supplemental material) (26, 27). Thirty-eight of the 40 genes were found, distributed over 17 contigs. Three were found in multiple copies (duplicates), but two of these duplicates were found adjacent on the contig, indicating that sequencing errors had resulted in split genes. Placing the above-described 17 contigs in a reference phylogenetic tree based on their housekeeping genes (26) shows that all contigs are inferred to belong to the *Verrucomicrobia* (Fig. S1). Given that we recapture all but two of the housekeeping genes and that these, with few exceptions, occur in single copies, the “*Spartobacteria baltica*” bin likely represents a substantial fraction of a single genome (28), although some fragments are likely missing due to incomplete coverage (see Materials and Methods). The

TABLE 1 Comparison of metagenome/genome properties from the “*Spartobacteria baltica*” bin and *Chthoniobacter flavus* (56), the closest relative that has been genome sequenced

Property	“ <i>Spartobacteria baltica</i> ”	<i>Chthoniobacter flavus</i>
No. of genes	2,226	6,778
No. of bases	1,811,214	7,848,700
No. of coding bases	1,657,652	6,925,094
GC (%)	62	61
No. of DNA scaffolds	334	62
No. of RNAs	37	62
No. of rRNAs	3	4
5S	1	2
16S	1	1
23S	1	1
No. of tRNAs	34	58
No. of genes with function prediction (%)	1,533 (69)	3,584 (53)
No. of genes with Pfam assignment (%)	1,443 (65)	4,074 (60)
No. of genes with COG assignment (%)	1,541 (69)	3,658 (54)
Total no. of COG IDs ^a	964	1,426
No. of shared COG IDs ^b	831	831
No. of unique COG IDs ^b	133	595
No. of contigs	334	
Avg coverage (×)	7.0	
N_{50}/L_{50}	904,350/6,763	
N_{75}/L_{75}	1,358,441/3,933	

^a IDs, identifications.

^b Comparing *Chthoniobacter flavus* and “*Spartobacteria baltica*.”

binning approach may also fail to group contigs having markedly different tetranucleotide compositions, such as recently acquired genome fragments of distant phylogenetic origin (29). Redoing the binning after spiking the metagenome with artificial contigs of the closest sequenced relative (*Chthoniobacter flavus* Ellin428) did, however, generate a cohesive cluster for this genome (Fig. S2), suggesting that the method should be accurate also for “*Spartobacteria baltica*” in this metagenomic context.

An advantage with metagenomics compared to isolate sequencing is that it gives direct insight into population heterogeneity (e.g., see references 30 and 31). Manual inspection of patterns of single nucleotide polymorphisms on aligned reads in a random selection of “*Spartobacteria baltica*” contigs suggested that the metagenome bin represents a population of closely related strains (similar enough to coassemble) that undergo recombination (see Fig. S3 in the supplemental material for an example). Deeper genomic coverage is needed to assess the population structure in detail.

Phylogenetic analysis of “*Spartobacteria baltica*.” A phylogenetic analysis of a concatenation of 31 of the single-copy genes confirmed our previous 16S rRNA-based placement of “*Spartobacteria baltica*” within the class *Spartobacteria* (21) in the superphylum *Planctomycetes*, *Verrucomicrobia*, and *Chlamydia* (PVC) (Fig. 2). The class *Spartobacteria* currently comprises one validly described order (*Chthoniobacteriales*) and family (*Chthoniobacteriaceae*) (9). The pasture soil isolate *Chthoniobacter flavus* Ellin428 (10, 20) is the closest cultured and sequenced relative (Fig. 2) but is still phylogenetically distant from “*Spartobacteria baltica*” (0.48 branch length distance in Fig. 2). Interestingly, *C. flavus* has a nearly three-times-larger genome (Table 1), which corroborates an earlier estimate of genome sizes based on metagenome data, which indicated considerably larger bacterial genomes in soil than in marine environments (32). Based on 16S rRNA gene “*Spartobacteria baltica*” phylogeny of the Silva 111 SSU Ref NR tree (33),

“*Spartobacteria baltica*” belongs to the lineage “LD29.” Detailed phylogenetic analysis of 16S rRNA genes from this lineage shows that solely environmental sequences derived from brackish, freshwater, and wastewater environments are found in this lineage (Fig. 3; see also Fig. S4 in the supplemental material). The most similar (99% identity) 16S rRNA sequences are from a freshwater lake in the Netherlands (GenBank accession number AF009975) and short sequences from the Baltic Sea (GenBank accession number EF627955). The lineage “LD29” was among the first 16S rRNA sequences of the *Spartobacteria* found by Zwart et al. (15).

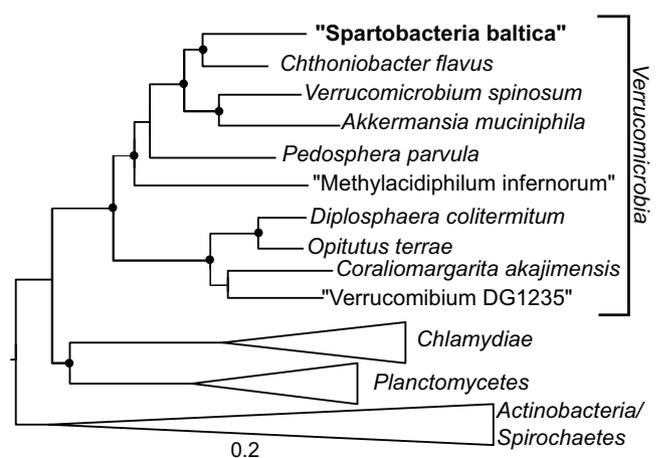


FIG 2 Maximum likelihood tree based on a concatenated alignment of 31 conserved genes of “*Spartobacteria baltica*” and representative genomes of *Verrucomicrobia* (10 representatives), *Chlamydia* (5 representatives), *Planctomycetes* (8 representatives), and *Spirochaetes/Actinobacteria* (21 representatives). The tree was rooted using the *Spirochaetes/Actinobacteria* as an outgroup. All groups except the *Verrucomicrobia* have been grouped into wedges for clarity. Dots indicate bootstrap values of >98%.

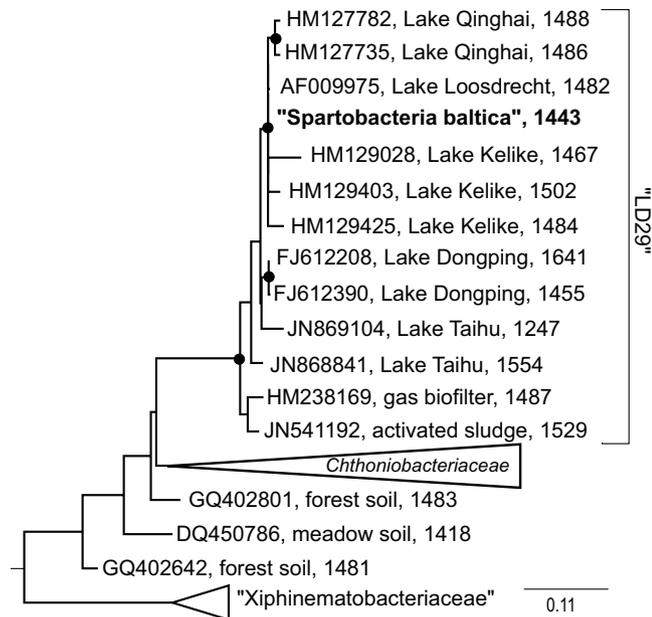


FIG 3 Phylogenetic tree of nonredundant sequences of >1,200 bp in the *Spartobacteria* class obtained from the Silva database 111 SSU Ref NR. "*Candidatus Methylophilum inferorum*" was used to root the tree. The tree was calculated using the RAxML algorithm with rapid bootstrap analysis (1,000 bootstraps). Only nodes supported by high bootstrap values are marked (filled circles, >95%). The origins of the sequences are indicated by the accession number, the isolation source, and the length of the sequence in bp.

Therefore, we named the first genomically characterized phylo-type of this lineage "*Spartobacteria baltica*."

Metabolism of "*Spartobacteria baltica*." A reconstruction of the energy metabolism by manual annotation of the metagenome bin revealed that "*Spartobacteria baltica*" uses a set of pathways typical of many aerobic heterotrophic organisms (see Table S3 in the supplemental material). Glucose can be converted to glucose 6-phosphate and degraded to pyruvate via the typical Embden-Meyerhof pathway (EMP). Pyruvate is further oxidized to acetyl coenzyme A (acetyl-CoA) that is used in the tricarboxylic acid cycle (TCA). The presence of fructose 1,6-bisphosphatase indicates the possibility for gluconeogenesis via the EMP, and the presence of genes coding for 2-oxoglutarate dehydrogenase, succinate dehydrogenase, and succinyl-CoA synthetase indicates a complete tricarboxylic acid cycle. The products of the TCA cycle and Embden-Meyerhof pathway are precursors of several amino acids. The pathways for the formation of L-alanine, L-valine, L-leucine, L-isoleucine, L-serine, and L-glycine, starting with intermediates of the EMP, are fully represented by the corresponding genes (Table S3). Also, biosynthetic pathways for the formation of L-aspartate, L-glutamate, L-glutamine, L-proline, L-threonine, L-lysine, and L-histidine from precursors of the TCA cycle were found. The biosynthetic pathways for L-arginine, L-methionine, and L-cysteine are not complete. Although we found the genes for a complete pentose phosphate pathway, which is involved in the regeneration of NADPH but also generates precursors for L-tryptophan, L-phenylalanine, and L-tyrosine biosynthesis, a few genes for a complete pathway of these three amino acids are missing. However, these more-complex pathways may miss single enzymes due to incomplete genome coverage. The same also accounts for the biosynthesis of purine and pyrimidine nucleotides

and the genes coding for lipopolysaccharides and peptidoglycan biosynthesis (Table S3). However, although *Verrucomicrobia* are described to have a Gram-negative staining cell wall, the class *Opitutae* has been reported to lack peptidoglycan (34), suggesting that *Spartobacteria* may also miss the corresponding genes.

Although the organism seems to be capable of biosynthesis of amino acids, essential prerequisites for the use of N_2 , NO_3^- or NO_2^- for the generation of nitrogen precursors were not found. While their absence may reflect incomplete genomics coverage, many ABC-type transporters involved in spermidine/putrescine (*potABCD*), peptide (*oppABCFD*), and branched-chain amino acid (*livKHMGF*) uptake may indicate the uptake of organic nitrogen and recycling of the acquired ammonia groups were found in the genome (see Table S3 in the supplemental material). Moreover, chitin—a putative substrate of "*Spartobacteria baltica*"—can support the nitrogen requirements of the organism (see below) (35). The metagenome "*Spartobacteria baltica*" has a complete *pstABSC* transporter system putatively involved in the uptake of phosphate and ABC transporters for iron (*fhuDBC*) and zinc (*znuABC*).

The sulfur metabolism is almost complete in the "*Spartobacteria baltica*" metagenome. Genes involved in the reduction of sulfate to hydrogen sulfide (via adenylylsulfate, 3'-phosphoadenylylsulfate, and sulfite), which is a precursor for the biosynthesis of L-cysteine by a cysteine synthase, were predicted. The sulfur metabolism plays an important role in the degradation of phytoplankton-derived polysaccharides since sulfated polysaccharides are frequently found in algae and *Cyanobacteria* (36). The metagenome bin also contains genes for a sulfate permease that may facilitate uptake of sulfate.

Polysaccharide-degrading enzymes. The aerobic heterotrophic metabolism described above requires a carbon source for the generation of energy and as a substrate for anabolism. Interestingly, "*Spartobacteria baltica*" contains several genes encoding glycoside hydrolases (GHs), key enzymes to degrade polysaccharide compounds. In total, 23 GHs representing 13 different GH families, as defined in the carbohydrate-active enzyme database (CAZy), were detected (18), suggesting the use of several different substrates, like cellulose, mannan, xylan, chitin, and starch (Table 2).

(i) Genes relevant for cellulose degradation. Three identified GH-encoding genes may be relevant for cellulose degradation; two are GH5 members, and one belongs to the family GH9. One of the predicted GH5 proteins has a single GH5 catalytic module, whereas the other GH5 member is supplemented with additional modules, including a family 6 carbohydrate binding module (CBM6) (Fig. 4A). The first GH5 protein sequence cannot be assigned to any subfamily, although it is distantly related to subfamilies GH5_7 and GH5_41 (Fig. 4B). The closest relatives to this sequence are GH5 enzymes from *Stackebrandtia nassauensis*, a cellulolytic member of the *Actinobacteria* (37), and *Lentisphaera araneosa*, an exopolymer-producing bacterium (38). The second modular GH5 member can be assigned to the recently described subfamily GH5_46 (39) (Fig. 4C), a poorly biochemically characterized GH5 subfamily. Carboxymethyl cellulose (CMC) activity has been described for a GH5_46 subfamily isolated from cow rumen (40), which currently is the only characterized enzyme in this subfamily (see Fig. S5 in the supplemental material). Moreover, the appended CBM6 module is known to bind to various β -glycans (41). Notably, the genome of *Chthoniobacter flavus* El-

TABLE 2 Comparison of CAZyme distributions in “*Spartobacteria baltica*” and the aquatic subdivision 1 *Verrucomicrobia* “*Verrucomicrobium* AAA168-F10” (46)

CAZy family	No. of “ <i>Spartobacteria baltica</i> ”	No. of “ <i>Verrucomicrobium</i> AAA168 F10”
GH1	4	2
GH2	3	
GH3	3	5
GH5	2	3
GH9	1	3
GH10	1	3
GH13		8
GH16	1	1
GH17	1	1
GH18	2	
GH26		1
GH30	2	
GH31		1
GH43	1	3
GH57	1	
GH77		1
GH78		2
GH81		5
GH109		19
GH119	1	

lin42 also contains two genes coding for GH5 proteins. However, these two, and a second GH5 representative from the “*Verrucomicrobium* AAA168-F10” genome (42), do not cluster together with the “*Spartobacteria baltica*” genes in the phylogenetic analysis, and none of them can currently be assigned to any GH5 subfamily (data not shown).

(ii) **Genes relevant for chitin degradation.** In addition, the “*Spartobacteria baltica*” metagenome bin also reveals two genes encoding candidate chitinolytic proteins belonging to the family GH18. One of the GH18 proteins contains two CBM2 modules and a CBM33 module. Another identified CBM33 module is independent, i.e., not connected to any catalytic module. Interestingly, members of CBM33 were recently shown to have enzymatic activity on insoluble substrates like chitin and cellulose (43, 44). Via a mechanism involving hydrolysis and oxidation, CBM33 enzymes boost degradation of chitin and cellulose by making crystalline polysaccharide regions accessible to enzymatic cleavage of GHs. The gene products from the three identified GH3 genes may harbor the chitinase activity required for complete hydrolysis of chitin. A bacterial GH3 protein with *N*-acetylhexosaminidase activity has previously been reported to have a function in the chitin utilization system (45).

(iii) **Other polysaccharide-degrading enzymes.** The genes encoding members of the GH families GH1, GH2, GH3, GH10, GH30, and GH43 represent candidates for the hydrolysis of non-cellulosic poly- and oligosaccharides and the side branches of hemicelluloses and pectins. For instance, endo-1,4- β -xylanase activity has been described in the families GH10, GH30, and GH43. However, the “*Spartobacteria baltica*” GH30 sequences cannot be classified into any of the defined GH30 subfamilies, and the top BLAST hit for the “*Spartobacteria baltica*” GH43 protein is a β -xylosidase (GenBank accession number ACE82692) from *Cellvibrio japonicus*. Since almost all characterized enzymes in GH10 are endo-1,4- β -xylanases, it is plausible that the sequence assigned to GH10 in our study can exhibit this activity. The identified GH16 and GH17 sequences are most likely involved in deg-

radation of laminarin, a β -1,3-glucan found mainly in brown algae (46). Of note, we discovered a family GH119 gene in the “*Spartobacteria baltica*” bin. Currently, GH119 contains only six members in the CAZy database, and the only biochemically characterized enzyme is an α -amylase (47).

Representatives from *Verrucomicrobia* have been shown to degrade polysaccharides in soil (subdivision 4 [2], subdivision 2 [10], and termite subdivision 4 [6]). Recently, Martinez-Garcia et al. (42) were able to identify coastal and freshwater *Verrucomicrobia* as polysaccharide-degraders using fluorescently labeled laminarin and xylan in combination with single-cell genomics. The coastal “*Verrucomicrobium* AAA168-F10” from the family *Verrucomicrobiaceae* (subdivision 1) contains 58 glycoside hydrolases putatively involved in the degradation of mucopolysaccharides, glycoproteins, peptidoglycan, celluloses, hemicelluloses, and glycogen. One of the three GH5 sequences identified in AAA168-F10 also falls within subfamily GH5_46 and, although truncated, shows high similarity to the GH5_46 member of “*Spartobacteria baltica*” (see Fig. S5 in the supplemental material).

Ecological role of “*Spartobacteria baltica*.” Utilization of polysaccharides by bacteria has been demonstrated in aquatic environments (17), but the identity and specific roles of the microbes performing this process are still elusive (48). “*Spartobacteria baltica*” has the genetic potential to use a variety of polysaccharides as carbon, nitrogen, and sulfur sources. In the marine environment, phytoplankton is a major source of such substrates, and a multitude of hydrolytic enzymes and sulfatases have been shown to be expressed during the decay of a phytoplankton bloom in associated bacteria (49). Previous studies reported a link between the dynamics of phytoplankton biomass and *Spartobacteria* in freshwater lakes (50, 51); moreover, Arnds et al. (12) found *Spartobacteria* cells attached to filamentous algae in a humic freshwater lake.

In the central Baltic Sea, pronounced phytoplankton blooms occur seasonally, with spring blooms being dominated by eukaryotic phytoplankton and summer blooms being dominated by *Cy-*

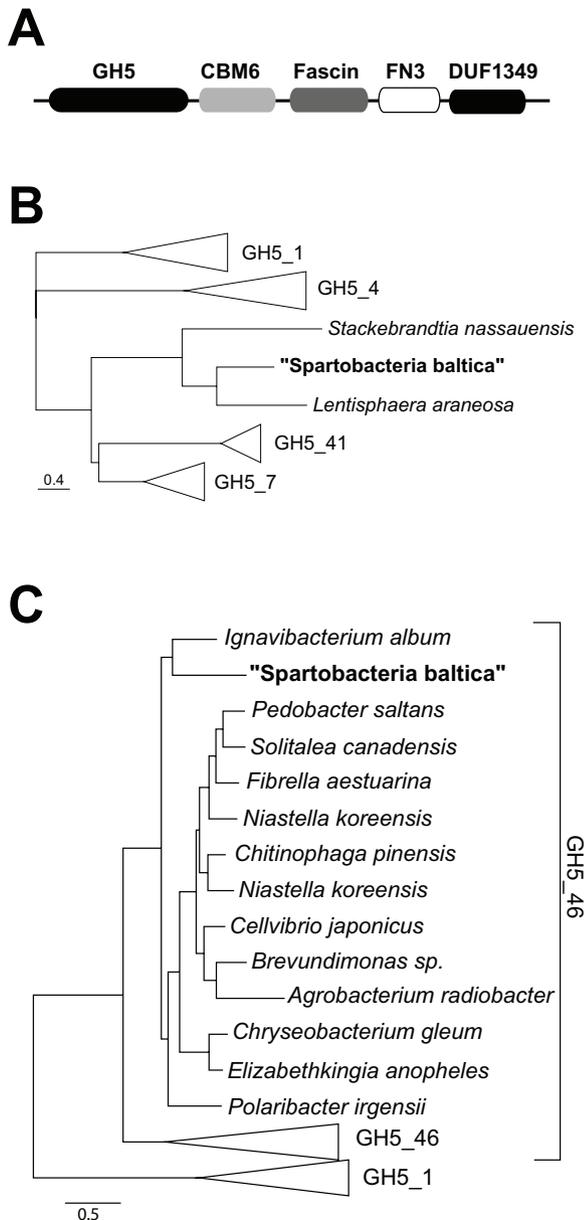


FIG 4 Analyses of “*Spartobacteria baltica*” GH5 sequences. (A) Modular structure of the GH5 protein sequence (gene id 2119805716 in Table S3). (B) A maximum likelihood tree of selected bacterial GH5 catalytic module sequences, including “*Spartobacteria baltica*” (gene id 2119806690 in Table S3). (C) A maximum likelihood tree of the subfamily GH5_46, including “*Spartobacteria baltica*” (gene id 2119805716 in Table S3), and selected bacterial sequences from related GH5 subfamilies. The phylogenetic analysis was restricted to the catalytic module.

anobacteria (52). In a previous study, the seasonal dynamics of surface water microbial communities in the central Baltic Sea (at the Landsort Deep) was investigated by 454 sequencing of amplicons of the V6 region of 16S rRNA genes (53). One of the most abundant OTUs in this data set is identical to the V6 region of the “*Spartobacteria baltica*” 16S rRNA gene. In the temporal study, the OTU displayed pronounced seasonal dynamics and peaked in July (with 5% of the reads) (see Fig. S6 in the supplemental material). This coincided roughly with blooms of filamentous *Cyano-*

bacteria (22), but the limited numbers of samples ($n = 8$) do not allow meaningful statistics. Instead, we used the 16S data from 213 samples of the Baltic Sea transect study (21) to search for spatial correlations between the “*Spartobacteria baltica*” OTU and other OTUs. Interestingly, the most highly correlated OTU was a picocyanobacterium (identical to *Synechococcus/Cyanobium* sequences from freshwater [54] and from the Baltic Sea [55], displaying a Spearman rank abundance correlation of 0.80 [P value of $<10^{-16}$]) (Fig. S7). Hence, the spatial data indicate a connection to picocyanobacteria, but it should be noted that filamentous *Cyanobacteria* were not accurately quantified in the study, and correlations to these may therefore have been missed. Moreover, the genomic findings indicate that substrates may additionally originate from eukaryotic phytoplankton, such as chitin. Besides crustaceans and copepods, phytoplankton blooms of *Thalassiosira* and *Skeletomena* are considered important sources of chitin since they produce chitin strands to increase their buoyancy (56). These species are highly abundant during spring phytoplankton blooms in the Baltic Sea (57) and may therefore provide the substrate during this period.

In summary, we have performed genomic analysis of the first aquatic representative of the *Spartobacteria*, one of the most abundant heterotrophic bacteria in the brackish Baltic Sea and other aquatic environments. The genome reveals a rich repertoire of polysaccharide-degrading genes, and the spatiotemporal data indicate ecological connections to phytoplankton. Further studies investigating seasonality and local distribution of microorganisms in the Baltic Sea will give more details on the interaction between aquatic *Spartobacteria* and phytoplankton; moreover, the enzymatic characterization of the glycoside hydrolases can give insight into their mode of action and substrate specificity.

MATERIALS AND METHODS

Sampling, DNA preparation, and sequencing. The water sample was obtained on a research cruise (MSM0803) of the RV Maria S. Merian in June and July 2008 at 59°47.88'N, 24°46.75'E (see Herlemann et al. [21] for details). Water samples for DNA analysis were filtered (0.22- μ m-pore-size white polycarbonate filters), and DNA was extracted according to Weinbauer et al. (58). The sample was sequenced at the Swedish Institute for Communicable Disease Control using 454 pyrosequencing (Roche) and a protocol for library preparation that allows minute amounts of sample DNA (59).

Metagenome assembly, binning, and annotation. 454 pyrosequencing reads were assembled using the Newbler assembler (Roche) with default parameter settings except that the “large” flag was used. Contigs with a size of ≥ 2 kb were subjected to phylogenetic binning by an emerging self-organizing map using the ESOM analyzer (60) based on tetranucleotide frequency distributions of contigs (23). The same parameter settings and initial data normalization as those used in Dick et al. (23) were applied, but a 50- by 80-pixel grid was used. Projecting contigs in the size range of 1 to 2 kb on the ESOM map that had already been generated with the longer contigs resulted in an additional 168 contigs (231,753 bp in total) falling in the “*Spartobacteria baltica*” region, indicating that a fraction of the genome was missing among the >2 -kb contigs. However, since the approach is unreliable for contigs <2 kb (23, 24), to minimize the risk for assigning external contigs to the genome, we restricted the analysis to contigs with a size of >2 kb. For making the spiked metagenome, the draft genome of *Chthoniobacter flavus* Ellin428 was downloaded from NCBI and split into 5-kb-long “contigs” and added to the metagenome. When running the ESOM analyzer on this, an 80- by 110-pixel grid was used, with other settings as described above. A Perl program for generating input to the ESOM analyzer can be downloaded at <https://github.com/tetramerFreqs/Binning>. For coloring the contigs in the map according to

(probable) phylum affiliation, contigs were BLASTx (61) searched against the NCBI nr database, and based on the BLAST output, MEGAN (62) was used to extract phylum-level annotations.

All contig sequences were annotated with the IMG/M metagenome analysis pipeline (see Table S3 in the supplemental material) (63). Automatic annotations with functional predictions were also improved manually with the annotation platform provided by Integrated Microbial Genomes (64). Metabolic pathways were reconstructed using MetaCyc (65) as a reference data set. Detailed information about the automatic genome annotation can be obtained from the JGI IMG website (http://img.jgi.doe.gov/w/doc/about_index.html).

Construction of the 16S rRNA gene tree. The metagenome revealed the complete 16S rRNA gene which was used for phylogenetic analysis. The phylogenetic 16S rRNA tree was constructed using the ARB program suite (66). All 16S rRNA spartobacterial sequences available in the Silva release 111 NR (33) were downloaded from the Silva browser (total of 631 sequences), the full-length sequence of “*Spartobacteria baltica*” was added, and “*Candidatus Methylocidiphilum infernorum*” was used as an outgroup. A core tree was estimated from 1,012 unambiguously aligned sequence positions of all nearly full-length (>1,200 bp) sequences (633 sequences), using maximum-likelihood analysis (RAxML) with rapid bootstrapping (1,000 replicates) and the GTRMIXI rate distribution model provided in the ARB package (Fig. 3). A total of 435 short sequences (>300 bp), positionally filtered by base frequency (50%), were added without changing the global tree topology by using the ARB parsimony tool (data not shown). Based on these results, a phylogenetic tree containing all sequences of >300 bp from the “LD29” lineage, including *Chthoniobacter flavus* as a reference and “Xiphinematobacteraceae” as an outgroup, was extracted (total of 168 sequences) (see Fig. S4 in the supplemental material). Phylogenetic trees were graphically processed using Fig tree (<http://tree.bio.ed.ac.uk/software/figtree/>).

Glycoside hydrolase identification and annotation. The domain structures of automatically annotated glycoside hydrolases were manually curated using SMART (67), Pfam (68), and the Conserved Domain Database (69). Glycoside hydrolase family annotations were revised by comparison to the carbohydrate-active enzymes database (<http://www.cazy.org>) (18). For the phylogenetic analysis, sequences of GH5 catalytic domains were aligned using MUSCLE (70), and the phylogenetic trees were generated using PhyML (71). Bootstrap support was calculated using 100 replicates. Subfamilies GH5_1 and GH5_4 were used as outgroups in the phylogenetic analysis.

Phylogenomic analysis. A phylogeny was estimated using a set of 31 conserved single-copy phylogenetic marker protein sequences, downloaded as HMMER3 HMM models (<http://hmmer.janelia.org>) from Pfam 26.0 (68) (PF00163, PF00203, PF00281, PF00347, PF00416, PF00828, PF03118, PF11987, PF00164, PF00237, PF00318, PF00366, PF00572, PF01000, PF03588, PF13393, PF00181, PF00238, PF00333, PF00410, PF00573, PF01193, PF03947, PF13603, PF00189, PF00252, PF00344, PF00411, PF00750, PF02403, PF10458). “*Spartobacteria baltica*” contigs were six-frame translated and searched with Pfam hmm profiles, as were the protein sequence complements of reference genomes. Marker proteins were identified in “*Spartobacteria baltica*” bin contigs and 44 microbial reference genomes based on the selection in the 2009 GEBA tree (72). The “*Spartobacteria baltica*” bin marker proteins were identified in twelve different contigs after six-frame translation. The sequences were aligned with Probcons (73) and analyzed with Zorro (74). Positions with a Zorro score of ≥ 6 were selected, and individual alignments were concatenated, producing an alignment with 7,597 well-aligned sites. A maximum likelihood tree was calculated with RAxML 7.2.8 using the LG substitution matrix (75) and a gamma model of rate heterogeneity (PROTGAMMALGF).

Nucleotide sequence accession number. The complete metagenome (all sequence reads) of the sample has been deposited in the European Nucleotide Archive under accession number ERP002583.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00569-12/-DCSupplemental>.

FIG S1, PDF file, 0.1 MB.
 FIG S2, PDF file, 0.3 MB.
 FIG S3, PDF file, 0.1 MB.
 FIG S4, PDF file, 0.4 MB.
 FIG S5, PDF file, 0.2 MB.
 FIG S6, PDF file, 0.1 MB.
 FIG S7, PDF file, 0.2 MB.
 TABLE S1, PDF file, 0.1 MB.
 TABLE S2, PDF file, 0.1 MB.
 TABLE S3, PDF file, 0.5 MB.

ACKNOWLEDGMENTS

The study was funded by the Leibniz Institute for Baltic Sea Research, the SAW Grant of the Leibniz Association (grants to D.P.R.H.), and the Swedish Research Councils VR and FORMAS (grants to A.F.A.).

REFERENCES

- Wagner M, Horn M. 2006. The *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* and sister phyla comprise a superphylum with biotechnological and medical relevance. *Curr. Opin. Biotechnol.* 17:241–249.
- Chin KJ, 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.
- Janssen PH, Schuhmann A, Mörschel E, Rainey FA. 1997. Novel anaerobic ultramicrobacteria belonging to the *Verrucomicrobiales* lineage of bacterial descent isolated by dilution culture from anoxic rice paddy soil. *Appl. Environ. Microbiol.* 63:1382–1388.
- Dunfield PF, Yuryev A, Senin P, Smirnova AV, Stott MB, Hou S, Ly B, Saw JH, Zhou Z, Ren Y, Wang J, Mountain BW, Crowe MA, Weatherby TM, Bodelier PL, Liesack W, Feng L, Wang L, Alam M. 2007. Methane oxidation by an extremely acidophilic bacterium of the phylum *Verrucomicrobia*. *Nature* 450:879–882.
- Pol A, Heijmans K, Harhangi HR, Tedesco D, Jetten MS, Op den Camp HJ. 2007. Methanotrophy below pH 1 by a new *Verrucomicrobia* species. *Nature* 450:874–878.
- Wertz JT, Kim E, Breznak JA, Schmidt TM, Rodrigues JL. 2012. Genomic and physiological characterization of the *Verrucomicrobia* isolate *Diplophaera colitermitum* gen. nov., sp. nov., reveals microaerophily and nitrogen fixation genes. *Appl. Environ. Microbiol.* 78:1544–1555.
- Khadem AF, Pol A, Jetten MS, Op den Camp HJ. 2010. Nitrogen fixation by the *verrucomicrobial* methanotroph “*Methylocidiphilum fumariolicum*” SolV. *Microbiology* 156:1052–1059.
- Bergmann GT, Bates ST, Eilers KG, Lauber CL, Caporaso JG, Walters WA, Knight R, Fierer N. 2011. The under-recognized dominance of *Verrucomicrobia* in soil bacterial communities. *Soil Biol. Biochem.* 43:1450–1455.
- Janssen PH, Costa KC, Hedlund BP. 2011. Class III. *Spartobacteria*, p 834–841. In Krieg NR, Ludwig W, Whitman WB, Hedlund BP, Paster BJ, Staley JT, Ward N, Brown D, Oarte A (ed), *Bergey’s manual of systematic bacteriology*, 2nd ed, vol 4. Springer Verlag, New York, NY.
- Sangwan P, Chen X, Hugenholtz P, Janssen PH. 2004. *Chthoniobacter flavus* gen. nov., sp. nov., the first pure-culture representative of subdivision two, *Spartobacteria* classis nov., of the phylum *Verrucomicrobia*. *Appl. Environ. Microbiol.* 70:5875–5881.
- Vandekerckhove TT, Willems A, Gillis M, Coomans A. 2000. Occurrence of novel *verrucomicrobial* species, endosymbiotic and associated with parthenogenesis in *Xiphinema americanum*-group species (*Nematoda*, *Longidoridae*). *Int. J. Syst. Evol. Microbiol.* 50:2197–2205.
- 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.
- Dojka MA, Hugenholtz P, Haack SK, Pace NR. 1998. Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. *Appl. Environ. Microbiol.* 64:3869–3877.
- Zwart G, Hiorns WD, Methé BA, van Agterveld MP, Huismans R, Nold SC, Zehr JP, Laanbroek HJ. 1998. Nearly identical 16S rRNA sequences

- recovered from lakes in North America and Europe indicate the existence of clades of globally distributed freshwater bacteria. *Syst. Appl. Microbiol.* 21:546–556.
15. Zwart G, Huismans R, van Agterveld MP, Van de Peer Y, De Rijk P, Eenhoorn H, Muyzer G, van Hannen EJ, Gons HJ, Laanbroek HJ. 1998. Divergent members of the bacterial division *Verrucomicrobiales* in a temperate freshwater lake. *FEMS Microbiol. Ecol.* 25:159–169.
 16. Freitas S, Hatosy S, Fuhrman JA, Huse SM, Welch DB, Sogin ML, Martiny AC. 2012. Global distribution and diversity of marine *Verrucomicrobia*. *ISME J.* 6:1499–1505.
 17. Arnosti C. 2011. Microbial extracellular enzymes and the marine carbon cycle. *Ann. Rev. Mar. Sci.* 3:401–425.
 18. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The carbohydrate-active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* 37:D233–D238.
 19. Henrissat B, Davies G. 1997. Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* 7:637–644.
 20. Kant R, van Passel MW, Palva A, Lucas S, Lapidus A, Glavina del Rio T, Dalin E, Tice H, Bruce D, Goodwin L, Pitluck S, Larimer FW, Land ML, Hauser L, Sangwan P, de Vos WM, Janssen PH, Smidt H. 2011. Genome sequence of *Chthoniobacter flavus* Ellin428, an aerobic heterotrophic soil bacterium. *J. Bacteriol.* 193:2902–2903.
 21. Herlemann DP, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF. 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* 5:1571–1579.
 22. Riemann L, Leitet C, Pommier T, Simu K, Holmfeldt K, Larsson U, Hagström A. 2008. The native bacterioplankton community in the central Baltic Sea is influenced by freshwater bacterial species. *Appl. Environ. Microbiol.* 74:503–515.
 23. Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, Banfield JF. 2009. Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* 10:R85.
 24. Weber M, Teeling H, Huang S, Waldmann J, Kassabgy M, Fuchs BM, Klindworth A, Klockow C, Wichels A, Gerdts G, Amann R, Glöckner FO. 2011. Practical application of self-organizing maps to interrelate biodiversity and functional data in NGS-based metagenomics. *ISME J.* 5:918–928.
 25. Kielak A, Rodrigues JL, Kuramae EE, Chain PS, van Veen JA, Kowalchuk GA. 2010. Phylogenetic and metagenomic analysis of *Verrucomicrobia* in former agricultural grassland soil. *FEMS Microbiol. Ecol.* 71:23–33.
 26. Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science* 311:1283–1287.
 27. Stark M, Berger SA, Stamatakis A, von Mering C. 2010. MLTreeMap—accurate maximum likelihood placement of environmental DNA sequences into taxonomic and functional reference phylogenies. *BMC Genomics* 11:461.
 28. von Mering C, Hugenholtz P, Raes J, Tringe SG, Doerks T, Jensen LJ, Ward N, Bork P. 2007. Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* 315:1126–1130.
 29. Sandberg R, Winberg G, Bränden CI, Kaske A, Ernberg I, Cöster J. 2001. Capturing whole-genome characteristics in short sequences using a naive Bayesian classifier. *Genome Res.* 11:1404–1409.
 30. Andersson AF, Banfield JF. 2008. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* 320:1047–1050.
 31. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyyev VV, Rubin EM, Rokhsar DS, Banfield JF. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428:37–43.
 32. Raes J, Korb J, Lercher MJ, von Mering C, Bork P. 2007. Prediction of effective genome size in metagenomic samples. *Genome Biol.* 8:R10.
 33. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO. 2007. Silva: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35:7188–7196.
 34. Cho JC, Janssen PH, Costa KC, Hedlund BP. 2011. Class II. *Opiritae*, p 834–841. In Krieg NR, Ludwig W, Whitman WB, Hedlund BP, Paster BJ, Staley JT, Ward N, Brown D, Oarte A (ed), Bergey's manual of systematic bacteriology, 2nd ed, vol 4. Springer Verlag, New York, NY.
 35. Kirchman DL, White J. 1999. Hydrolysis and mineralization of chitin in the Delaware estuary. *Aquat. Microb. Ecol.* 18:187–196.
 36. De Philippis R, Vincenzini M. 1998. Exocellular polysaccharides from *Cyanobacteria* and their possible applications. *FEMS Microbiol. Rev.* 22:151–175.
 37. Anderson I, Abt B, Lykidis A, Klenk HP, Kyrpides N, Ivanova N. 2012. Genomics of aerobic cellulose utilization systems in *Actinobacteria*. *PLoS One* 7:e39331. <http://dx.doi.org/10.1371/journal.pone.0039331>.
 38. Cho JC, Vergin KL, Morris RM, Giovannoni SJ. 2004. *Lentisphaera araneosa* gen. nov., sp. nov, a transparent exopolymer producing marine bacterium, and the description of a novel bacterial phylum, *Lentisphaerae*. *Environ. Microbiol.* 6:611–621.
 39. Aspeborg H, Coutinho PM, Wang Y, Brumer H, III, Henrissat B. 2012. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol. Biol.* 12:186.
 40. Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, Rubin EM. 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* 331:463–467.
 41. Michel G, Barbeyron T, Kloareg B, Czjzek M. 2009. The family 6 carbohydrate-binding modules have coevolved with their appended catalytic modules toward similar substrate specificity. *Glycobiology* 19:615–623.
 42. Martinez-Garcia M, Brazel DM, Swan BK, Arnosti C, Chain PS, Reitenga KG, Xie G, Poulton NJ, Lluesma Gomez M, Masland DE, Thompson B, Bellows WK, Ziervogel K, Lo CC, Ahmed S, Gleasner CD, Detter CJ, Stepanauskas R. 2012. Capturing single cell genomes of active polysaccharide degraders: an unexpected contribution of *Verrucomicrobia*. *PLoS One* 7:e35314. <http://dx.doi.org/10.1371/journal.pone.0035314>.
 43. Forsberg Z, Vaaje-Kolstad G, Westereng B, Bunæs AC, Stenstrom Y, MacKenzie A, Sørli M, Horn SJ, Eijsink VG. 2011. Cleavage of cellulose by a CBM33 protein. *Protein Sci.* 20:1479–1483.
 44. Vaaje-Kolstad G, Westereng B, Horn SJ, Liu Z, Zhai H, Sørli M, Eijsink VG. 2010. An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science* 330:219–222.
 45. Li H, Morimoto K, Katagiri N, Kimura T, Sakka K, Lun S, Ohmiya K. 2002. A novel beta-N-acetylglucosaminidase of *Clostridium paraputrificum* M-21 with high activity on chitobiose. *Appl. Microbiol. Biotechnol.* 60:420–427.
 46. Cock JM, Coelho SM, Brownlee C, Taylor AR. 2010. The *Ectocarpus* genome sequence: insights into brown algal biology and the evolutionary diversity of the eukaryotes. *New Phytol.* 188:1–4.
 47. Watanabe H, Nishimoto T, Kubota M, Chaen H, Fukuda S. 2006. Cloning, sequencing, and expression of the genes encoding an isocyclomaltooligosaccharide glucanotransferase and an alpha-amylase from a *Bacillus circulans* strain. *Biosci. Biotechnol. Biochem.* 70:2690–2702.
 48. 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.
 49. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennis CM, Kassabgy M, Huang S, Mann AJ, Waldmann J, Weber M, Klindworth A, Otto A, Lange J, Bernhardt J, Reinsch C, Hecker M, Peplies J, Bockelmann FD, Callies U, Gerdts G, Wichels A, Wiltshire KH, Glöckner FO, Schweder T, Amann R. 2012. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* 336:608–611.
 50. Ouellette AJ, Handy SM, Wilhelm SW. 2006. Toxic *Microcystis* is widespread in Lake Erie: PCR detection of toxin genes and molecular characterization of associated cyanobacterial communities. *Microb. Ecol.* 51:154–165.
 51. Parveen B, Mary I, Vellet A, Ravet V, Debros D. 2013. Temporal dynamics and phylogenetic diversity of free-living and particle-associated *Verrucomicrobia* communities in relation to environmental variables in a mesotrophic lake. *FEMS Microbiol. Ecol.* 83:189–201.
 52. Feistel R, Nausch G, Wasmund N (ed). 2008. State and evolution of the Baltic sea 1952–2005: a detailed 50-year survey of meteorology and climate, physics, chemistry, biology, and marine environment. John Wiley & Sons Inc., Hoboken, NJ.
 53. Andersson AF, Riemann L, Bertilsson S. 2010. Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J.* 4:171–181.
 54. Crosbie ND, Pöckl M, Weisse T. 2003. Dispersal and phylogenetic diversity of nonmarine picocyanobacteria, inferred from 16S rRNA gene

- and *cpcBA*-intergenic spacer sequence analyses. *Appl. Environ. Microbiol.* 69:5716–5721.
55. Sjöstedt J, Koch-Schmidt P, Pontarp M, Canbäck B, Tunlid A, Lundberg P, Hagström A, Riemann L. 2012. Recruitment of members from the rare biosphere of marine bacterioplankton communities after an environmental disturbance. *Appl. Environ. Microbiol.* 78:1361–1369.
 56. Durkin CA, Mock T, Armbrust EV. 2009. Chitin in diatoms and its association with the cell wall. *Eukaryot. Cell* 8:1038–1050.
 57. Wasmund N, Tuimala J, Suikkanen S, Vandepitte L, Kraberg A. 2011. Long-term trends in phytoplankton composition in the western and central Baltic sea. *J. Mar. Syst.* 87:145–159.
 58. Weinbauer MG, Fritz I, Wenderoth DF, Höfle MG. 2002. Simultaneous extraction from bacterioplankton of total RNA and DNA suitable for quantitative structure and function analyses. *Appl. Environ. Microbiol.* 68:1082–1087.
 59. Zheng Z, Advani A, Melefors O, Glavas S, Nordström H, Ye W, Engstrand L, Andersson AF. 2010. Titration-free massively parallel pyrosequencing using trace amounts of starting material. *Nucleic Acids Res.* 38:e137. <http://dx.doi.org/10.1093/nar/gkq332>.
 60. Ultsch A, Mörchen F. 2005. ESOM-Maps: tools for clustering, visualization, and classification with emergent SOM. Technical report, p 46. Department of Mathematics and Computer Science, University of Marburg, Marburg, Germany.
 61. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
 62. Huson DH, Auch AF, Qi J, Schuster SC. 2007. Megan analysis of metagenomic data. *Genome Res.* 17:377–386.
 63. Hauser L, Larimer F, Land M, Shah M, Uberbacher E. 2004. Analysis and annotation of microbial genome sequences. *Genet. Eng.* 26:225–238.
 64. Markowitz VM, Szeto E, Palaniappan K, Grechkin Y, Chu K, Chen IM, Dubchak I, Anderson I, Lykidis A, Mavromatis K, Ivanova NN, Kyrpides NC. 2008. The integrated microbial genomes (IMG) system in 2007: data content and analysis tool extensions. *Nucleic Acids Res.* 36:D528–D533.
 65. Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Pujar A, Shearer AG, Travers M, Weerasinghe D, Zhang P, Karp PD. 2012. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 40:D742–D753.
 66. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar A, Buchner T, Steppi S, Jobb G, Forster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, König A, Liss T, Lussmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, Schleifer KH, Schleifer KH. 2004. ARB: a software environment for sequence data. *Nucleic Acids Res.* 32:1363–1371.
 67. Letunic I, Doerks T, Bork P. 2012. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* 40:D302–D305.
 68. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. 2012. The Pfam protein families database. *Nucleic Acids Res.* 40:D290–D301.
 69. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH. 2011. CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* 39:D225–D229.
 70. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
 71. Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52:696–704.
 72. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, Hooper SD, Pati A, Lykidis A, Spring S, Anderson IJ, D’Haeseleer P, Zemla A, Singer M, Lapidus A, Nolan M, Copeland A, Han C, Chen F, Cheng JF, Lucas S, Kerfeld C, Lang E, Gronow S, Chain P, Bruce D, Rubin EM, Kyrpides NC, Klenk HP, Eisen JA. 2009. A phylogeny-driven genomic encyclopedia of *Bacteria* and *Archaea*. *Nature* 462:1056–1060.
 73. Do CB, Mahabhashyam MS, Brudno M, Batzoglou S. 2005. ProbCons: probabilistic consistency-based multiple sequence alignment. *Genome Res.* 15:330–340.
 74. Wu M, Chatterji S, Eisen JA. 2012. Accounting for alignment uncertainty in phylogenomics. *PLoS One* 7:e30288. <http://dx.doi.org/10.1371/journal.pone.0030288>.
 75. Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25:1307–1320.