bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ecophysiology of freshwater Verrucomicrobia inferred from genomes recovered through time-series metagenomics

Shaomei He[1,2], Sarah LR Stevens[1], Leong-Keat Chan[4], Stefan Bertilsson[3], Tijana Glavina del Rio[4], Susannah G Tringe[4], Rex R Malmstrom[4], and Katherine D McMahon[1,5,*]

[1]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA
[2]Department of Geoscience, University of Wisconsin-Madison, Madison, WI, USA
[3]Department of Ecology and Genetics, Limnology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden
[4]DOE Joint Genome Institute, Walnut Creek, CA, USA
[5]Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI, USA

[*]**Corresponding author**
Katherine D McMahon
1550 Linden Drive
5552 Microbial Sciences Building
Madison, WI 53706, USA
trina.mcmahon@wisc.edu

Running title: Freshwater Verrucomicrobia ecophysiology inferred from genomics

Keywords: Metagenome-assembled genome (MAG), Verrucomicrobia, Freshwater, Glycoside hydrolase, Nutrient limitation, Cytochrome *c*

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

Microbes are critical in carbon and nutrient cycling in freshwater ecosystems. Members of the Verrucomicrobia are ubiquitous in such systems, yet their roles and ecophysiology are not well understood. In this study, we recovered 19 Verrucomicrobia draft genomes by sequencing 184 time-series metagenomes from a eutrophic lake and a humic bog that differ in carbon source and nutrient availabilities. These genomes span four of the seven previously defined Verrucomicrobia subdivisions, and greatly expand the known genomic diversity of this freshwater lineage. Genome analysis revealed their role as (poly)saccharide-degraders in freshwater, uncovered interesting genomic features for this life style, and suggested their adaptation to nutrient availabilities in their environments. Between the two lakes, Verrucomicrobia populations differ significantly in glycoside hydrolase gene abundance and functional profiles, reflecting the autochthonous and terrestrially-derived allochthonous carbon sources of the two ecosystems respectively. Several bog populations exhibited nitrogen cost minimization in their proteomes and genomes, which is likely an adaptation to long-term nitrogen limitation in the bog. Interestingly, a number of genomes recovered from the bog contained gene clusters that potentially encode a novel porin-multiheme cytochrome *c* complex and might be involved in extracellular electron transfer in the anoxic humic-rich environment. Notably, most epilimnion genomes have large numbers of Planctomycete-specific cytochrome *c*-containing genes, which exhibited nearly opposite distribution patterns with glycoside hydrolase genes, probably associated with the different environmental oxygen availability and carbohydrate complexity between lakes/layers. Overall, the recovered genomes are a major step towards understanding the role, ecophysiology and distribution of Verrucomicrobia in freshwater.

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

INTRODUCTION

Microbes play important roles in mediating carbon (C) and nutrient cycling in freshwater ecosystems. Among them are the ubiquitous Verrucomicrobia, which exhibit a cosmopolitan distribution in freshwater lakes. For example, they were present in 90% of 81 studied lakes (Zwart et al 2003). Verrucomicrobia abundances often range between <1% to 6% of the total microbial community (Eiler and Bertilsson 2004, Newton et al 2011, Parveen et al 2013), and contributed up to 19% in a humic lake (Arnds et al 2010). Yet, in comparison to other freshwater bacterial groups, such as members of the Actinobacteria, Cyanobacteria and Proteobacteria phyla, Verrucomicrobia have received relatively less attention, and their functions and ecophysiology in freshwater are not well understood.

As a phylum, Verrucomicrobia (V) was first proposed relatively recently, in 1997 (Hedlund et al 1997). Together with Planctomycetes (P), Chlamydiae (C), and sister phyla such as Lentisphaerae, they comprise the PVC superphylum. In addition to being cosmopolitan in freshwater, Verrucomicrobia have been found in oceans (Yoon et al 2007, Yoon et al 2008), soil (Sangwan et al 2005), wetlands (Qiu et al 2014), rhizosphere (da Rocha et al 2010), and animal guts (Derrien et al 2004, Wertz et al 2012), as free-living organisms or symbionts of eukaryotes. Verrucomicrobia isolates are metabolically diverse, including aerobes, facultative anaerobes, and obligate anaerobes, and they are mostly heterotrophs, using various mono-, oligo-, and poly-saccharides for growth (Chin et al 2001, Derrien et al 2004, Hedlund et al 1996, Hedlund et al 1997, Otsuka et al 2013a, Otsuka et al 2013b, Qiu et al 2014, Sangwan et al 2004, Scheuermayer et al 2006, Yoon et al 2007). Not long ago an autotrophic verrucomicrobial methanotroph (*Methylacidiphilum*

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

fumariolicum SolV) was discovered in acidic thermophilic environments as the only non-proteobacterial aerobic methanotroph to date (Pol et al 2007).

In marine environments, Verrucomicrobia are also ubiquitous (Freitas et al 2012) and suggested to have a key role as polysaccharide degraders (Cardman et al 2014, Martinez-Garcia et al 2012a). Genomic insights gained through sequencing single cells (Martinez-Garcia et al 2012a) or extracting Verrucomicrobia bins from metagenomes (Herlemann et al 2013) have revealed high abundances of glycoside hydrolase genes, providing more evidence for their critical roles in C cycling in marine environments.

In freshwater, Verrucomicrobia have been suggested to degrade glycolate (Paver and Kent 2010) and polysaccharides (Martinez-Garcia et al 2012a). The abundance of some phylum members was favored by high nutrient availabilities (Haukka et al 2006, Lindström et al 2004), algal blooms (Kolmonen et al 2004), low pH, high temperature, high hydraulic retention time (Lindström et al 2005), and more labile DOC (Arnds et al 2010). To date, there are very few freshwater Verrucomicrobia isolates, including *Verrucomicrobium spinosum* (Schlesner 1987) and several *Prosthecobacter* spp. (Hedlund et al 1997). Physiological studies showed that they are aerobes, primarily using carbohydrates, but not amino acids, alcohols, or rarely organic acids for growth. However, these few cultured isolates only represent a single clade within subdivision 1. By contrast, 16S rRNA gene based studies discovered a much wider phylogenic range of freshwater Verrucomicrobia, including subdivisions 1, 2, 3, 4, 5, and 6 (Arnds et al 2010, Eiler and Bertilsson 2004, Martinez-Garcia et al 2012a, Parveen et al 2013, Zwart et al 1998). Due to the very few cultured representatives and few available genomes from this freshwater

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

lineage, the ecological functions of the vast uncultured freshwater Verrucomicrobia are largely unknown.

In this study, we sequenced a total of 184 metagenomes in a time-series study of two lakes with contrasting characteristics, particularly differing in C source, nutrient availabilities, and pH. We recovered a total of 19 Verrucomicrobia draft genomes spanning subdivision 1, 2, 3, and 4 of the seven previously defined Verrucomicrobia subdivisions. We inferred their metabolisms, revealed their adaptation to C and nutrient conditions, and uncovered some interesting and novel features, including a novel putative porin-multiheme cytochrome system that may be involved in extracellular electron transfer. The gained insights advanced our understanding of the ecophysiology, roles in C cycling, and ecological niches of this ubiquitous freshwater bacterial group.

MATERIALS AND METHODs

Study sites. Samples for metagenome sequencing were collected from two temperate lakes in Wisconsin, USA, Lake Mendota and Trout Bog Lake, during ice-off periods of each year (May to November). Mendota is an urban eutrophic lake with most of its C being autochthonous (in-lake produced), whereas Trout Bog is a small, acidic and nutrient-poor dystrophic lake with mostly terrestrially-derived (allochthonous) C. General lake characteristics are summarized in **Table 1**.

Sampling. For Mendota, we collected depth-integrated water samples from the surface 12 m (mostly consisting of the epilimnion layer) at 94 time points from 2008 to 2012, and samples were referred to as "ME" (Garcia et al 2016). For Trout Bog, we collected the

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

integrated hypolimnion layer at 45 time points from 2007 to 2009 and the integrated epilimnion layer at 45 time points from 2007 to 2009, and samples were referred to as "TH" and "TE", respectively (Bendall et al 2016). All samples were filtered through 0.22 μm polyethersulfone filters and stored at -80°C until extraction. DNA was extracted from the filters using the FastDNA kit (MP Biomedicals) according to manufacturer's instruction with some minor modifications as described previously (Shade et al 2008).

Metagenome sequencing, assembly, and draft genome recovery. Details of metagenome sequencing, assembly, and binning were described in Bendall *et al.* (2016) and Hamilton et al (preprint). Briefly, shotgun Illumina HiSeq 2500 metagenome libraries were constructed for each of the DNA samples. Three combined assemblies were generated by co-assembling reads from all metagenomes within the ME, TE, and TH groups, respectively. Binning was conducted on the three combined assemblies to recover "metagenome-assembled genomes" (MAGs) based on the combination of contig tetranucleotide frequency and differential coverage patterns across time points using MetaBAT (Kang et al 2015). Subsequent manual curation of MAGs was conducted to remove contigs that did not correlate well with the median temporal abundance pattern of all contigs within a MAG, as described in Bendall *et al.* (2016).

Genome annotation and completeness estimation. MAGs were submitted to the DOE Joint Genome Institute's Integrated Microbial Genome (IMG) database for gene prediction and function annotation (Markowitz et al 2013). The IMG Taxon Object IDs for Verrucomicrobia MAGs are listed in **Table 2**. The completeness of each MAG was estimated

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

based on the number of recovered single-copy essential genes compared to the total of 105 single-copy essential genes expected for complete Verrucomicrobia genomes (Albertsen et al 2013). MAGs with an estimated completeness lower than 50% were not included in this study.

Taxonomic and phylogenetic analysis. A total of 19 MAGs were classified to the Verrucomicrobia phylum based on taxonomic assignment by PhyloSift using 37 conserved phylogenetic marker genes (Darling et al 2014), as described in Bendall *et al.* (2016). A phylogenetic tree was reconstructed from the 19 Verrucomicrobia MAGs and 24 reference genomes using an alignment concatenated from individual protein alignments of five conserved essential single-copy genes (represented by TIGR01391, TIGR01011, TIGR00663, TIGR00460, and TIGR00362) that were recovered in all Verrucomicrobia MAGs. Individual alignments were first generated with MUSCLE (Edgar 2004), concatenated, and trimmed to exclude columns that contain gaps for more than 30% of all sequences. A maximum likelihood phylogenetic tree was constructed using PhyML 3.0 (Guindon et al 2010), with the LG substitution model and the gamma distribution parameter estimated by PhyML. Bootstrap values were calculated based on 100 replicates. *Kiritimatiella glycovorans* L21-Fru-AB was used as an outgroup in the phylogenetic tree. This bacterium was initially designated as the first (and so far the only) cultured representative of Verrucomicrobia subdivision 5. However, this subdivision was later proposed as a novel sister phylum associated with Verrucomicrobia (Spring et al 2016), making it an ideal outgroup for this analysis.

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Estimate of metabolic potential. IMG provides functional annotation based on KO (KEGG orthology) term, COG (cluster of orthologous group), pfam, and TIGRfam. To estimate metabolic potential, we primarily used KO terms due to their direct link to KEGG pathways. COG, pfam, and TIGRfam were also used when KO terms are not available for a function. Pathways are primarily reconstructed according to KEGG modules, and MetaCyc pathway is used if a KEGG module is not available for a pathway. As these MAGs are incomplete genomes, a fraction of genes in a pathway may be missing due to genome incompleteness. Therefore, we estimated the completeness of a pathway as the fraction of recovered enzymes in that pathway (e.g. a pathway is 100% complete if all enzymes in that pathway are encoded by genes recovered in a MAG). As some genes are shared by multiple pathways, signature genes specific for a pathway were used to indicate the presence of a pathway. If signature genes for a pathway were missing in all MAGs, that pathway was likely absent in all genomes. Based on this, we established criteria for estimating pathway completeness in each MAG. If a signature gene in a pathway was present, we report the percentage of genes in the pathway that we found. If a signature gene was absent in a MAG, but present in at least one third of all MAGs (i.e. >=7), we still report the pathway completeness for that MAG in order to account for genome incompleteness. Otherwise, we considered the pathway to be absent (i.e. completeness is 0%).

Glycoside hydrolase identification. Glycoside hydrolase (GH) genes were identified using the dbCAN annotation tool (http://csbl.bmb.uga.edu/dbCAN/annotate.php) (Yin et al 2012) using HMMER search against hidden Markov models (HMMs) built for all GHs, with an E-value cutoff of 1e-7, except GH109, for which we found that the HMM used by dbCAN

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

is not specific for this GH. To identify verrucomicrobial GH109, BLASTP was performed using GH109 sequences from verrucomicrobial *Akkermansia muciniphila* ATCC BAA-835 listed in the CAZy database (http://www.cazy.org), with E-value cutoff of 1e-6 and query sequence coverage cutoff of 50%.

Other bioinformatic analyses. Protein cellular location was predicted using CELLO v.2.5 (http://cello.life.nctu.edu.tw) (Yu et al 2006) and PSORTb v.3.0 (http://www.psort.org/psortb) (Yu et al 2010). The beta-barrel structure of outer membrane proteins was predicted by PRED-TMBB (http://bioinformatics.biol.uoa.gr//PRED-TMBB) (Bagos et al 2004).

RESULTS AND DISCUSSION

Comparison of the two lakes

The two studied lakes exhibited contrasting characteristics (**Table 1**). The most notable difference is the primary C source and nutrient availabilities. Mendota is an urban eutrophic lake with most of its C being autochthonous (in-lake produced through photosynthesis). By contrast, Trout Bog is a nutrient-poor dystrophic lake, surrounded by boreal forests and sphagnum mats, thus receiving large amounts of terrestrially-derived allochthonous C that is rich in humic and fulvic acids. Compared to Mendota, Trout Bog features higher DOC levels, but is more limited in nutrient availability, with much higher DOC:TN and DOC:TP ratios (**Table 1**). Nutrient limitation in Trout Bog is even more extreme than revealed by these ratios because much of the N and P is tied up in complex dissolved organic matter. In addition, Trout Bog has lower oxygenic photosynthesis due to

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

decreased photosynthetically active radiation (PAR) as a result of absorption by DOC (Read and Rose 2013). Together with the high consumption of dissolved oxygen by heterotrophic respiration, oxygen levels decrease quickly with depth in the water column in Trout Bog. Dissolved oxygen levels are below detection in the hypolimnion nearly year-round (Shade et al 2008). Due to these contrasts, we expected to observe differences in bacterial C and nutrient use, as well as differences reflecting the electron acceptor conditions between these two lakes. Hence, the retrieval of numerous Verrucomicrobia draft genomes in the two lakes not only allows the revelation of their general functions in freshwater, but also provides an opportunity to study their ecophysiological adaptation associated with the local environmental differences.

Verrucomicrobia draft genome retrieval and their distribution patterns

Using the binning facilitated by tetranucleotide frequency and relative abundance patterns over time, a total of 19 Verrucomicrobia MAGs were obtained, including eight from ME, three from TE, and eight from TH (**Table 2**). The 19 MAGs exhibited a clustering of their tetranucleotide frequency largely based on the two lakes (**Figure S1**), suggesting distinct overall genomic signatures associated with each system.

Genome completeness of the 19 MAGs ranged from 53% to 96%, as determined based on the 105 single-copy essential genes expected for complete Verrucomicrobia genomes (Albertsen et al 2013). We performed phylogenetic analysis of these MAGs using a concatenated alignment of their conserved genes, and found that they span a wide phylogenetic spectrum and distribute in subdivisions 1, 2, 3, and 4 of the seven previously defined Verrucomicrobia subdivisions (Arnds et al 2010, Pol et al 2007, Schlesner et al

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2006) (**Figure 1**), as well as three unclassified Verrucomicrobia MAGs. In fact, the phylogenetic diversity of Verrucomicrobia in these two lakes is higher than that represented by these 19 MAGs. For example, an un-binned contig from the Mendota metagenome is likely from subdivision 6 representing the freshwater clade LD19 (Zwart et al 2003), as its genes share an average amino acid identity of 89% with previously recovered MAGs from subdivision 6 (Hugerth et al 2015), Notably, this contig contains a gene encoding proteorhodopsin, a light-driven proton pump, consistent with a previously recovered rhodopsin gene from a freshwater Verrucomicrobia single cell genome (Martinez-Garcia et al 2012b). However, no rhodopsin gene was observed in any of the 19 other higher quality Verrucomicrobia MAGs that were analyzed extensively in this study. Nevertheless, we restricted our analysis to the 19 MAGs based on their high quality, while at the same time acknowledging that they do not cover all the genomic diversity of Verrucomicrobia populations in our study systems.

Presently available freshwater Verrucomicrobia isolates are restricted to subdivision 1. The recovered MAGs allow the inference of metabolisms and ecology of a considerable diversity within uncultured freshwater Verrucomicrobia. Notably, all MAGs from subdivision 3 were recovered from TH, and all MAGs from subdivision 1, except TH2746, were from the epilimnion (either ME or TE), indicating differences in phylogenetic distribution between lakes and between layers within a lake.

We used normalized fold coverage of MAGs within individual metagenomes to comparatively infer relative population abundance (see detailed coverage depth estimation in **Supplementary Text**). Briefly, we counted the number of reads mapping with a minimum identity of 95% and calculated a relative abundance for each MAG based on

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

coverage depth per contig and several normalization steps. Thus, we assume that each MAG represents a distinct population within the lake-layer from which it was recovered (Bendall et al 2016, Garcia et al 2016). This estimate does not directly indicate the actual relative abundance of these populations within the total community per se; rather it allows us to compare population abundance levels from different lakes and sampling occasions within the set of 19 MAGs. This analysis indicates that Verrucomicrobia populations in Trout Bog were proportionally more abundant and persistent over time compared to those in Mendota (**Figure 2**). Verrucomicrobia populations in Mendota were only transiently abundant, bloomed once to a few times during the sampling season and diminished to extremely low levels for the remainder of the year.

Saccharolytic life style and adaptation to different C sources

Verrucomicrobia isolates from different environments are known to grow on various mono-, oligo-, and poly-saccharides, but are unable to grow on amino acids, alcohols, or most organic acids (Chin et al 2001, Derrien et al 2004, Hedlund et al 1996, Hedlund et al 1997, Otsuka et al 2013a, Otsuka et al 2013b, Qiu et al 2014, Sangwan et al 2004, Scheuermayer et al 2006, Shieh and Jean 1998, Yoon et al 2007). Culture-independent research suggests marine Verrucomicrobia as candidate polysaccharide degraders with large number of genes involved in polysaccharide utilization (Cardman et al 2014, Herlemann et al 2013, Martinez-Garcia et al 2012a).

In the 19 Verrucomicrobia MAGs, we observed rich arrays of glycoside hydrolase (GH) genes, representing a total of 78 different GH families acting on diverse polysaccharides (**Figure S2**). As these genomes have different degrees of completeness, to

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

281 compare among them, we normalized GH occurrence frequencies by the total number of
282 genes in each MAG to estimate the percentage of genes annotated as GHs (i.e. GH coding
283 density), which ranged from 0.4% to 4.9% for these MAGs (**Figure 3a**). In general, GH
284 coding density was higher in Trout Bog MAGs than in Mendota MAGs. Notably, six TH MAGs
285 had extremely high (~4%) GH coding densities (**Figure 3a**), with each MAG harboring 119-
286 239 GH genes, representing 36-59 different GH families (**Figures 4 & S2**). Although GH
287 coding density in most ME genomes in subdivisions 1 and 2 was relatively low (0.4-1.6%),
288 it was still higher than in many other bacterial groups (Martinez-Garcia et al 2012a).

289 The GH abundance and diversity within a genome may determine the width of the
290 substrate spectrum and/or the complexity of carbohydrates used by that organism. For
291 example, there are 20 GH genes in the *Rubritalea marina* genome, and this marine
292 verrucomicrobial aerobe only uses a limited spectrum of carbohydrate monomers and
293 dimers, but not the majority of (poly)saccharides tested (Scheuermayer et al 2006). By
294 contrast, 164 GH genes are present in the *Opitutus terrae* genome, and this soil
295 verrucomicrobial anaerobe can thus grow on a wider range of mono-, di- and poly-
296 saccharides (Chin et al 2001). Therefore, it is plausible that the GH-rich Trout Bog
297 Verrucomicrobia populations may be able to use a wider range of more complex
298 polysaccharides than the Mendota populations.

299 The 10 most abundant GH families in these Verrucomicrobia MAGs include GH2, 29,
300 78, 95, and 106 (**Figure 4**). These specific GHs were absent or at very low abundances in
301 marine Verrucomicrobia genomes (Herlemann et al 2013, Martinez-Garcia et al 2012a),
302 suggesting a general difference in carbohydrate substrate use between freshwater- and
303 marine Verrucomicrobia. Hierarchical clustering of MAGs based on overall GH abundance

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

profiles indicated a grouping pattern largely separated by lake (**Figure S3**). Prominently over-represented GHs in most Trout Bog MAGs include GH29, 78, 95, and 106, all of which mainly function as α-L-fucosidases or α-L-rhamnosidases, as well as GH2, a β-galactosidase that also acts on other β-linked dimers. By contrast, over-represented GHs in the Mendota MAGs are GH13, 20, 33, 57, and 77. Among them, GH13 and 57 are α-amylases, and GH20 is β-hexosaminidases, clearly exhibiting different substrate spectra from GHs over-represented in the Trout Bog MAGs. Therefore, the patterns in GH functional profiles may suggest varied carbohydrate substrate preferences and ecological niches occupied by Verrucomicrobia, probably reflecting the different carbohydrate composition derived from different sources between Mendota and Trout Bog.

A particularly interesting contrast of GH genes was observed between ME3880 and TH2746. These two populations were phylogenetically close relatives in subdivision 1 (**Figure 1**), and had an average nucleotide identity of 74% among the 654 bi-directional best hits between the two genomes. Albeit that, their estimated genome sizes differ substantially (**Table 2**). Notably, TH2746 possesses a total of 239 GH genes, whereas ME3880 has only 17 GH genes, despite the two MAGs having comparably high levels of completeness. Based on the GH diversity and abundance profile (**Figures 3a, 4, and S3**), TH2746 is an outlier of subdivision 1, and instead shares more similarity to other TH genomes in subdivision 3. Therefore this subdivision 1 Verrucomicrobia population genome might have been expanded to adapt to the carbohydrate substrate composition in Trout Bog.

Overall, GH diversity and abundance profile may reflect the DOC availability, chemical variety and complexity, and may suggest microbial adaptation to different C

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sources in the two ecosystems. We speculate that the rich arrays of GH genes, and presumably broader substrate spectra of Trout Bog populations, partly contribute to their higher abundance and persistence over the sampling season (**Figure 2**), as they are less likely impacted by fluctuations of individual carbohydrates; whereas Mendota populations with fewer GHs and presumably more specific substrate spectra are relying on autochthonous C and therefore exhibit a bloom-and-bust abundance pattern (**Figure 2**) that might be associated with algal blooms as previous suggested (Kolmonen et al 2004). On the other hand, bogs also experience seasonal algal blooms (Kent et al 2004, Kent et al 2007) that introduce brief pulses of autochthonous C to these otherwise allochthonous-driven systems. Clearly, much remains to be learned about the routes through which C is metabolized by bacteria in such lakes, and comparative genomics is a novel way to use the organisms to tell us about C flow through the ecosystem.

Other genome features of the saccharide-degrading life style

Seven Verrucomicrobia MAGs spanning subdivisions 1, 2, 3, and 4 possess genes needed to construct bacterial microcompartments (BMCs), which are quite rare among studied bacterial lineages. Such BMC genes in Planctomycetes are involved in the degradation of plant and algal cell wall sugars, and are required for growth on L-fucose, L-rhamnose and fucoidans (Erbilgin et al 2014). Genes involved in L-fucose and L-rhamnose degradation cluster with BMC shell protein-coding genes in the seven Verrucomicrobia MAGs (**Figure 5a**). This is consistent with the high abundance of α-L-fucosidase or α-L-rhamnosidase GH genes in these MAGs (**Figure 4**), suggesting the importance of fucose- and rhamnose-containing polysaccharides for these Verrucomicrobia populations.

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TonB-dependent receptor (TBDR) genes were found in Verrucomicrobia MAGs, and are present at over 20 copies in TE1800 and TH2519. TBDRs are located on the outer cellular membrane of Gram-negative bacteria, usually mediating the transport of iron siderophore complex and vitamin B_{12} across the outer membrane through an active process. More recently, TBDRs were suggested to be involved in carbohydrate transport across the outer membrane by some bacteria that consume complex carbohydrates, and in their carbohydrate utilization (CUT) loci, TBDR genes usually cluster with genes encoding inner membrane transporters, GHs and regulators for efficient carbohydrate transportation and utilization (Blanvillain et al 2007). Such novel CUT loci are present in TE1800 and TH2519, with TBDR genes clustering with genes encoding inner membrane sugar transporters, monosaccharide utilization enzymes, and GHs involved in the degradation of pectin, xylan, and fucose-containing polymers (**Figure 5b**). Notably, most GHs in the CUT loci are predicted to be extracellular or outer membrane proteins (**Figure 5b**), catalyzing extracellular hydrolysis reactions to release mono- and oligo-saccharides, which are transported across the outer membrane by TBDR proteins. Therefore, such CUT loci may allow these Verrucomicrobial populations to coordinately and effectively scavenge the hydrolysis products before they diffuse away.

Genes encoding for inner membrane carbohydrate transporters are abundant in Verrucomicrobia MAGs (**Figure S4**). The Embden-Mayerhof pathway for glucose degradation, as well as pathways for degrading a variety of other sugar monomers, including galactose, rhamnose, fucose, xylose, and mannose, were recovered (complete or partly-complete) in most MAGs (**Figure 6**). As these sugars are abundant carbohydrate monomers in plankton and plant cell walls, the presence of these pathways together with

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

GH genes suggest that these Verrucomicrobia populations may use plankton- and plant-derived saccharides. Machinery for pyruvate degradation to acetyl-CoA and the TCA cycle are also present in most MAGs. These results are largely consistent with their hypothesized role in carbohydrate degradation and previous studies on Verrucomicrobia isolates.

Notably, a large number of genes encoding proteins belonging to a sulfatase family (pfam00884) are present in the majority of MAGs (**Figure 3b**), similar to the high representation of these genes in marine Verrucomicrobia genomes (Herlemann et al 2013, Martinez-Garcia et al 2012a). Sulfatases hydrolyze sulfate esters, which are rich in sulfated polysaccharides. In general, sulfated polysaccharides are abundant in marine algae and plants (mainly in seaweeds) (Jiao et al 2011), but have also been found in some freshwater cyanobacteria (Filali Mouhim et al 1993) and plant species (Dantas-Santos et al 2012). Sulfatase genes in our Verrucomicrobia MAGs were often located in the same neighborhood as genes encoding for extracellular proteins with a putative pectin lyase activity, proteins with a carbohydrate-binding module (pfam13385), GHs, and proteins with PSCyt domains (**Figure 3c** and discussed later). Their genome context lends support for the participation of these genes in C and sulfur cycling by degrading sulfated polysaccharides, which can serve as an abundant source of sulfur for cell biosynthesis as well as C for energy and growth.

Overall, the high abundance of GH, sulfatase, and carbohydrate transporter genes, metabolic pathways for degrading diverse carbohydrate monomers, and other genome features adapted to the saccharolytic life style suggest that Verrucomicrobia are primarily (poly)saccharide-degraders in freshwater, rather than degraders of the algal exudate

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

glycolate as previously suggested (Paver and Kent 2010) (See details of glycolate utilization by Verrucomicrobia MAGs in **Supplementary Text** and **Figure S5**).

Nitrogen (N) metabolism and adaptation to different N availabilities

Most Verrucomicrobia MAGs in our study do not appear to reduce nitrate or other nitrogenous compounds, and they seem to uptake and use ammonia (**Figure 6**), and occasionally amino acids (**Figure S4**), as an N-source. Further, some Trout Bog populations may have additional avenues to generate ammonia, including genetic machineries for assimilatory nitrate reduction in TH2746, nitrogenase genes for nitrogen fixation and urease genes in some of the Trout Bog MAGs (**Figure 6**), probably as adaptions to N-limited conditions in Trout Bog.

Although Mendota is a eutrophic lake, N can become temporarily limiting during the high-biomass period when N is consumed by large amounts of phytoplankton and bacterioplankton (Beversdorf et al 2013). For some bacteria, when N is temporarily limited while C is in excess, cells convert and store the extra C as biopolymers. For example, the verrucomicrobial methanotroph *M. fumariolicum* SolV accumulated a large amount of glycogen (up to 36% of the total dry weight of cells) when the culture was N-limited (Khadem et al 2012). Similar to this verrucomicrobial methanotroph, genes in glycogen biosynthesis are present in most MAGs from Mendota and Trout Bog (**Figure 6**). Indeed, a glycogen synthesis pathway is also present in most Verrucomicrobia genomes in the public database (data not shown), suggesting that glycogen accumulation might be a common feature for this phylum to cope with the changing pools of C and N in the environment and facilitate their survival when either is temporally limited.

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To cope with long-term sustained N-limitation, such as that which occurs in Trout Bog, some bacteria may have evolved features to minimize N cost in amino acid and nucleotide biosynthesis, thus having reduced proteome N-contents and low genome G+C contents (Bragg and Hyder 2004, Grzymski and Dussaq 2012). Grzymski and Dussaq (2012) used the numbers of N and C atoms per amino-acid residue side chain (ARSC) for each predicted protein to indicate proteome N-contents. Based on ARSC, we found that most, although not all, Trout Bog proteomes have lower N-contents than most Mendota proteomes, with TE1800, TH2519 and TH4093 having extremely low N-contents (**Figure S6a, c**). Interestingly, proteome C-content is reversely correlated with N-content (correlation coefficient r = -0.83), exhibiting an opposite trend in terms of N and C atom usage between the two lakes (**Figure S6b, d, e**), probably reflecting the extremely high DOC and low available N in Trout Bog as compared to Mendota. The genome G+C contents ranged from 42% to 68%, and are positively correlated to proteome N-contents (r = 0.84) (**Figure S6f**), consistent with the expectation that low G+C genomes are favored under long-term N limitation (Bragg and Hyder 2004). Therefore, genome G+C and proteome N contents both suggest that some of the Verrucomicrobia populations in Trout Bog have been adapted to long-term N-limitation.

Phosphorus (P) metabolism and other metabolic features

Verrucomicrobia populations represented by these MAGs may be able to survive under low P conditions, as suggested by the presence of genes responding to P limitation, such as the two-component regulator (*phoRB*), alkaline phosphatase (*phoA*), phosphonoacetate hydrolase (*phnA*), and high-affinity phosphate-specific transporter system (*pstABC*)

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(**Figure 6**). Detailed discussion in P acquisition and metabolism and other metabolic aspects, such as acetate metabolism, sulfur metabolism, oxygen tolerance, and the presence of the alternative complex III and cytochrome *c* oxidase genes in the oxidative phosphorylation pathway, are discussed in the Supplementary Text (**Figure S7**).

Anaerobic respiration and a putative porin-multiheme cytochrome *c* system

Respiration using alternative electron acceptors is important for overall lake metabolism in the DOC-rich humic Trout Bog, as the oxygen levels decrease quickly with depth in the water column. We therefore searched for genes involved in anaerobic respiration, and found that genes in the dissimilatory reduction of nitrate, nitrite, sulfate, sulfite, DMSO, and TMAO are largely absent in all MAGs (**Supplementary Text, Figure S7)**. Compared to those anaerobic processes, genes for dissimilatory metal reduction are less well understood. In more extensively studied cultured iron [Fe(III)] reducers, outer surface *c*-type cytochromes (cyt*c*), such as OmcE and OmcS in *Geobacter sulfurreducens* are involved in Fe(III) reduction at the cell outer surface (Mehta et al 2005). Further, a periplasmic multiheme cytochrome *c* (MHC, e.g. MtrA in *Shewanella oneidensis* and OmaB/OmaC in *G. sulfurreducens*) can be embedded into a porin (e.g. MtrB in *S. oneidensis* and OmbB/OmbC in *G. sulfurreducens*), forming a porin-MHC complex as an extracellular electron transfer (EET) conduit to reduce extracellular Fe(III) (Liu et al 2014, Shi et al 2014). Such outer surface cyt*c* and porin-MHC systems involved in Fe(III) reduction were also suggested to be important in reducing the quinone groups in humic substances (HS) at the cell surface (Bucking et al 2012, Shyu et al 2002, Voordeckers et al 2010). The reduced HS can be re-oxidized by Fe(III) or oxygen, thus HS can serve as electron shuttles to facilitate Fe(III)

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

464 reduction (Lovley et al 1996, Lovley and Blunt-Harris 1999) or as regenerateable electron
465 acceptors at the anoxic-oxic interface or over redox cycles (Klupfel et al 2014).

466 Outer surface cyt*c* or porin-MHC systems homologous to the ones in *G.*
467 *sulfurreducens* and *S. oneidensis* are not present in Verrucomicrobia MAGs. Instead, we
468 identified a novel porin-coding gene clustering with MHC genes in six MAGs (**Figure 5c**).
469 These porins were predicted to have at least 20 transmembrane motifs, and their adjacent
470 cyt*c* were predicted to be periplasmic proteins with eight conserved heme-binding sites. In
471 several cases, a gene encoding an extracellular MHC is also located in the same gene cluster.
472 As their gene organization is analogous to the porin-MHC gene clusters in *G. sulfurreducens*
473 and *S. oneidensis*, we hypothesize that these genes in Verrucomicrobia may encode a novel
474 porin-MHC complex involved in EET.

475 As these porin-MHC gene clusters are novel, we further confirmed that they are
476 indeed from Verrucomicrobia. Their containing contigs were indeed classified to
477 Verrucomicrobia based on the consensus of the best BLASTP hits for genes on these
478 contigs. Notably, the porin-MHC gene cluster was only observed in MAGs recovered from
479 the HS-rich Trout Bog, especially from the anoxic hypolimnion environment. Searching the
480 NCBI and IMG databases for the porin-MHC gene clusters homologous to those in Trout
481 Bog, we identified homologs in genomes within the Verrucomicrobia phylum, including
482 *Opitutus terrae* PB90-1 isolated from rice paddy soil, *Opitutus* sp. GAS368 isolated from
483 forest soil, "*Candidatus* Udaeobacter copiosus" recovered from prairie soil, Opititae-40 and
484 Opititae-129 recovered from freshwater sediment, and Verrucomicrobia bacterium
485 IMCC26134 recovered from freshwater; some of their residing environments are also rich
486 in HS. Therefore, based on the occurrence pattern of porin-MHC among Verrucomicrobia

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

genomes, we hypothesize that such porin-MHCs might participate in EET to HS in anoxic HS-rich environments, and HS may further shuttle electrons to poorly soluble metal oxides or be regenerated at the anoxic-oxic interface, thereby diverting more C flux to respiration instead of fermentation and methanogenesis, which could impact the overall energy metabolism and green-house gas emission in the bog environment.

Occurrence of Planctomycete-specific cytochrome *c* and domains

One of the interesting features of Verrucomicrobia and its sister phyla in the PVC superphylum is the presence of a number of novel protein domains in some of their member genomes (Kamneva et al 2012, Studholme et al 2004). These domains were initially identified in marine planctomycete *Rhodopirellula baltica* (Studholme et al 2004) and therefore, were referred to as "Planctomycete-specific", although some of them were later identified in other PVC members (Kamneva et al 2012). In our Verrucomicrobia MAGs, most genes containing Planctomycete-specific cytochrome *c* domains (PSCyt1 to PSCyt3) also contain other Planctomycete-specific domains (PSD1 through PSD5) with various combinations and arrangements (**Figure 7** and **S8a**). Further, PSCyt2-containing and PSCyt3-containing genes are usually next to two different families of unknown genes, respectively (**Figure S8b**). Such conserved domain architectures and gene organizations, as well as their high occurrence frequencies in some of the Verrucomicrobia MAGs are intriguing, yet nothing is known about their functions. However, some of the PSCyt-containing genes also contain protein domains identifiable as carbohydrate-binding modules (CBMs), suggesting a role in carbohydrate metabolism (see detailed discussion in **Supplementary Text**).

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The coding density of PSCyt-containing genes indicates that they tend to be more abundant in the epilimnion (either ME or TE) genomes (**Figure 3c**) and exhibit a reverse correlation with the GH coding density (r = -0.62). Interestingly, sulfatase-coding genes are often in the neighborhood of PSCyt-containing genes in ME and TE genomes, whereas sulfatase-coding genes often neighbor with GH genes in TH genomes. The genomic context suggests PSCyt-containing gene functions somewhat mirror those of GHs (although their reaction mechanisms likely differ fundamentally). However, these PSCyt-containing genes were predicted to be periplasmic or cytoplasmic proteins rather than extracellular or outer membrane proteins. Hence, if they are indeed involved in carbohydrate degradation, they likely act on mono- or oligomers that can be transported into the cell. Further, the distribution patterns of GH versus PSCyt-containing genes between the epilimnion and hypolimnion may reflect the difference in oxygen availability and their carbohydrate substrate complexity between the two layers, suggesting some niche differentiation within Verrucomicrobia in freshwater systems. Therefore, a plausible hypothesis is that these bacterial populations have evolved different genetic machineries to use carbohydrates of different complexities under different oxygen and C availabilities.

SUMMARY

Verrucomicrobia MAGs recovered from the two contrasting lakes greatly expanded the known genomic diversity of freshwater Verrucomicrobia, revealed the ecophysiology and some interesting adaptive features of this ubiquitous yet less understood freshwater lineage. The overrepresentation of GH, sulfatase, and carbohydrate transporter genes, the genetic potential to use various sugars, and the microcompartments for fucose and

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

rhamnose degradation suggest that they are primarily (poly)saccharide degraders in freshwater. Most of the MAGs encode machineries to cope with the changing availability of N and P and can survive nutrient limitation. Despite these generalities, these Verrucomicrobia differ significantly between lakes in the abundance and functional profiles of their GH genes, which may reflect different C sources of the two lakes. Further, several Trout Bog MAGs exhibit N cost minimization in their proteomes and genomes, likely an adaptation to long-term sustained N limitation. Interestingly, a number of MAGs in Trout Bog possess gene clusters potentially encoding a novel porin-multiheme cytochrome *c* complex, and might be involved in extracellular electron transfer in the anoxic humic-rich environment. Intriguingly, large numbers of Planctomycete-specific cytochrome *c*-containing genes are present in MAGs from the epilimnion, exhibiting nearly opposite distribution patterns with GH genes. Future studies are needed to elucidate the functions of these novel and fascinating genomic features.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the North Temperate Lakes Microbial Observatory 2007-2012 field crews, UW-Trout Lake Station, the UW Center for Limnology, and the Global Lakes Ecological Observatory Network for field and logistical support. We give special thanks to past McMahon lab graduate students Ashley Shade, Ryan Newton, Emily Read, and Lucas Beversdorf. We acknowledge efforts by many McMahon Lab undergrads and technicians

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

related to sample collection and DNA extraction, particularly Georgia Wolfe. KDM acknowledges funding from the United States National Science Foundation Microbial Observatories program (MCB-0702395), the Long Term Ecological Research program (NTL-LTER DEB-1440297) and an INSPIRE award (DEB-1344254). This material is also based upon work that supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture (Hatch Project 1002996). Finally, we personally thank the individual program directors and leadership at the National Science Foundation for their commitment to continued support of long term ecological research.

REFERENCE

Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat Biotech 31: 533-538.

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. Systematic and applied microbiology 33: 139-148.

Bagos PG, Liakopoulos TD, Spyropoulos IC, Hamodrakas SJ (2004). PRED-TMBB: a web server for predicting the topology of beta-barrel outer membrane proteins. Nucleic Acids Res 32: W400-404.

Bendall ML, Stevens SLR, Chan L-K, Malfatti S, Schwientek P, Tremblay J, Schackwitz W, Martin J, Pati A, Bushnell B, Froula J, Kang D, Tringe SG, Bertilsson S, Moran MA, Shade A, Newton RJ, McMahon KD, Malmstrom RR (2016). Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. The ISME journal.

Beversdorf LJ, Miller TR, McMahon KD (2013). The role of nitrogen fixation in cyanobacterial bloom toxicity in a temperate, eutrophic lake. PloS one 8: e56103.

Blanvillain S, Meyer D, Boulanger A, Lautier M, Guynet C, Denance N, Vasse J, Lauber E, Arlat M (2007). Plant carbohydrate scavenging through tonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. PloS one 2: e224.

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bragg JG, Hyder CL (2004). Nitrogen versus carbon use in prokaryotic genomes and proteomes. Proceedings Biological sciences / The Royal Society 271 Suppl 5: S374-377.

Bucking C, Piepenbrock A, Kappler A, Gescher J (2012). Outer-membrane cytochrome-independent reduction of extracellular electron acceptors in Shewanella oneidensis. Microbiology (Reading, England) 158: 2144-2157.

Cardman Z, Arnosti C, Durbin A, Ziervogel K, Cox C, Steen AD, Teske A (2014). Verrucomicrobia are candidates for polysaccharide-degrading bacterioplankton in an arctic fjord of Svalbard. Applied and environmental microbiology 80: 3749-3756.

Chin KJ, Liesack W, Janssen PH (2001). Opitutus terrae gen. nov., sp. nov., to accommodate novel strains of the division 'Verrucomicrobia' isolated from rice paddy soil. International journal of systematic and evolutionary microbiology 51: 1965-1968.

da Rocha UN, van Elsas JD, van Overbeek LS (2010). Real-time PCR detection of Holophagae (Acidobacteria) and Verrucomicrobia subdivision 1 groups in bulk and leek (Allium porrum) rhizosphere soils. Journal of Microbiological Methods 83: 141-148.

Dantas-Santos N, Gomes DL, Costa LS, Cordeiro SL, Costa MS, Trindade ES, Franco CR, Scortecci KC, Leite EL, Rocha HA (2012). Freshwater plants synthesize sulfated polysaccharides: heterogalactans from Water Hyacinth (Eicchornia crassipes). International journal of molecular sciences 13: 961-976.

Darling AE, Jospin G, Lowe E, Matsen IV FA, Bik HM, Eisen JA (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ 2: e243.

Derrien M, Vaughan EE, Plugge CM, de Vos WM (2004). Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. International journal of systematic and evolutionary microbiology 54: 1469-1476.

Edgar RC (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113-113.

Eiler A, Bertilsson S (2004). Composition of freshwater bacterial communities associated with cyanobacterial blooms in four Swedish lakes. Environmental microbiology 6: 1228-1243.

Erbilgin O, McDonald KL, Kerfeld CA (2014). Characterization of a planctomycetal organelle: a novel bacterial microcompartment for the aerobic degradation of plant saccharides. Applied and environmental microbiology 80: 2193-2205.

Filali Mouhim R, Cornet J-F, Fontane T, Fournet B, Dubertret G (1993). Production, isolation and preliminary characterization of the exopolysaccharide of the cyanobacterium Spirulina platensis. Biotechnology Letters 15: 567-572.

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Freitas S, Hatosy S, Fuhrman JA, Huse SM, Welch DB, Sogin ML, Martiny AC (2012). Global distribution and diversity of marine Verrucomicrobia. The ISME journal 6: 1499-1505.

Garcia SL, Stevens SLR, Crary B, Martinez-Garcia M, Stepanauskas R, Woyke T, Tringe SG, Andersson S, Bertilsson S, Malmstrom RR, McMahon KD (2016). Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations. bioRxiv.

Grzymski JJ, Dussaq AM (2012). The significance of nitrogen cost minimization in proteomes of marine microorganisms. The ISME journal 6: 71-80.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology 59: 307-321.

Haukka K, Kolmonen E, Hyder R, Hietala J, Vakkilainen K, Kairesalo T, Haario H, Sivonen K (2006). Effect of nutrient loading on bacterioplankton community composition in lake mesocosms. Microbial ecology 51: 137-146.

Hedlund BP, Gosink JJ, Staley JT (1996). Phylogeny of Prosthecobacter, the fusiform caulobacters: members of a recently discovered division of the bacteria. International journal of systematic bacteriology 46: 960-966.

Hedlund BP, Gosink JJ, Staley JT (1997). Verrucomicrobia div. nov., a new division of the bacteria containing three new species of Prosthecobacter. Antonie van Leeuwenhoek 72: 29-38.

Herlemann DP, Lundin D, Labrenz M, Jurgens K, Zheng Z, Aspeborg H, Andersson AF (2013). Metagenomic de novo assembly of an aquatic representative of the verrucomicrobial class Spartobacteria. mBio 4: e00569-00512.

Hugerth L, Larsson J, Alneberg J, Lindh M, Legrand C, Pinhassi J, Andersson A (2015). Metagenome-assembled genomes uncover a global brackish microbiome. Genome Biology 16: 279.

Jiao G, Yu G, Zhang J, Ewart HS (2011). Chemical structures and bioactivities of sulfated polysaccharides from marine algae. Marine drugs 9: 196-223.

Kamneva OK, Knight SJ, Liberles DA, Ward NL (2012). Analysis of genome content evolution in pvc bacterial super-phylum: assessment of candidate genes associated with cellular organization and lifestyle. Genome biology and evolution 4: 1375-1390.

Kang DD, Froula J, Egan R, Wang Z (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3: e1165.

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Kent AD, Jones SE, Yannarell AC, Graham JM, Lauster GH, Kratz TK, Triplett EW (2004). Annual Patterns in Bacterioplankton Community Variability in a Humic Lake. Microbial ecology 48: 550-560.

Kent AD, Yannarell AC, Rusak JA, Triplett EW, McMahon KD (2007). Synchrony in aquatic microbial community dynamics. The ISME journal 1: 38-47.

Khadem AF, van Teeseling MC, van Niftrik L, Jetten MS, Op den Camp HJ, Pol A (2012). Genomic and Physiological Analysis of Carbon Storage in the Verrucomicrobial Methanotroph "Ca. Methylacidiphilum Fumariolicum" SolV. Frontiers in microbiology 3: 345.

Klupfel L, Piepenbrock A, Kappler A, Sander M (2014). Humic substances as fully regenerable electron acceptors in recurrently anoxic environments. Nature Geosci 7: 195-200.

Kolmonen E, Sivonen K, Rapala J, Haukka K (2004). Diversity of cyanobacteria and heterotrophic bacteria in cyanobacterial blooms in Lake Joutikas, Finland. Aquatic Microbial Ecology 36: 201-211.

Lindström ES, Vrede K, Leskinen E (2004). Response of a member of the Verrucomicrobia, among the dominating bacteria in a hypolimnion, to increased phosphorus availability. Journal of Plankton Research 26: 241-246.

Lindström ES, Kamst-Van Agterveld MP, Zwart G (2005). Distribution of Typical Freshwater Bacterial Groups Is Associated with pH, Temperature, and Lake Water Retention Time. Applied and environmental microbiology 71: 8201-8206.

Liu Y, Wang Z, Liu J, Levar C, Edwards MJ, Babauta JT, Kennedy DW, Shi Z, Beyenal H, Bond DR, Clarke TA, Butt JN, Richardson DJ, Rosso KM, Zachara JM, Fredrickson JK, Shi L (2014). A trans-outer membrane porin-cytochrome protein complex for extracellular electron transfer by Geobacter sulfurreducens PCA. Environmental Microbiology Reports 6: 776-785.

Lovley DR, Coates JD, Blunt-Harris EL, Phillips EJP, Woodward JC (1996). Humic substances as electron acceptors for microbial respiration. Nature 382: 445-448.

Lovley DR, Blunt-Harris EL (1999). Role of humic-bound iron as an electron transfer agent in dissimilatory Fe(III) reduction. Applied and environmental microbiology 65: 4252-4254.

Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J, Woyke T, Huntemann M, Anderson I, Billis K, Varghese N, Mavromatis K, Pati A, Ivanova NN, Kyrpides NC (2013). IMG 4 version of the integrated microbial genomes comparative analysis system. Nucleic Acids Research.

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (2012a). Capturing single cell genomes of active polysaccharide degraders: an unexpected contribution of Verrucomicrobia. PloS one 7: e35314.

Martinez-Garcia M, Swan BK, Poulton NJ, Gomez ML, Masland D, Sieracki ME, Stepanauskas R (2012b). High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. The ISME journal 6: 113-123.

Mehta T, Coppi MV, Childers SE, Lovley DR (2005). Outer membrane c-type cytochromes required for Fe(III) and Mn(IV) oxide reduction in Geobacter sulfurreducens. Applied and environmental microbiology 71: 8634-8641.

Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S (2011). A Guide to the Natural History of Freshwater Lake Bacteria. Microbiology and Molecular Biology Reviews : MMBR 75: 14-49.

Otsuka S, Suenaga T, Vu HT, Ueda H, Yokota A, Senoo K (2013a). Brevifollis gellanilyticus gen. nov., sp. nov., a gellan-gum-degrading bacterium of the phylum Verrucomicrobia. International journal of systematic and evolutionary microbiology 63: 3075-3078.

Otsuka S, Ueda H, Suenaga T, Uchino Y, Hamada M, Yokota A, Senoo K (2013b). Roseimicrobium gellanilyticum gen. nov., sp. nov., a new member of the class Verrucomicrobiae. International journal of systematic and evolutionary microbiology 63: 1982-1986.

Parveen B, Mary I, Vellet A, Ravet V, Debroas 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 microbiology ecology 83: 189-201.

Paver SF, Kent AD (2010). Temporal patterns in glycolate-utilizing bacterial community composition correlate with phytoplankton population dynamics in humic lakes. Microbial ecology 60: 406-418.

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.

Qiu YL, Kuang XZ, Shi XS, Yuan XZ, Guo RB (2014). Terrimicrobium sacchariphilum gen. nov., sp. nov., an anaerobic bacterium of the class 'Spartobacteria' in the phylum Verrucomicrobia, isolated from a rice paddy field. International journal of systematic and evolutionary microbiology 64: 1718-1723.

Read JS, Rose KC (2013). Physical responses of small temperate lakes to variation in dissolved organic carbon concentrations. Limnology and Oceanography 58: 921-931.

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. Applied and environmental microbiology 70: 5875-5881.

Sangwan P, Kovac S, Davis KE, Sait M, Janssen PH (2005). Detection and cultivation of soil verrucomicrobia. Applied and environmental microbiology 71: 8402-8410.

Scheuermayer M, Gulder TA, Bringmann G, Hentschel U (2006). Rubritalea marina gen. nov., sp. nov., a marine representative of the phylum 'Verrucomicrobia', isolated from a sponge (Porifera). International journal of systematic and evolutionary microbiology 56: 2119-2124.

Schlesner H (1987). Verrucomicrobium spinosum gen. nov., sp. nov.: a fimbriated prosthecate bacterium. Systematic and applied microbiology 10: 54-56.

Schlesner H, Jenkins C, Staley J (2006). The Phylum Verrucomicrobia: A Phylogenetically Heterogeneous Bacterial Group. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (eds). The Prokaryotes. Springer New York. pp 881-896.

Shade A, Jones SE, McMahon KD (2008). The influence of habitat heterogeneity on freshwater bacterial community composition and dynamics. Environmental microbiology 10: 1057-1067.

Shi L, Fredrickson JK, Zachara JM (2014). Genomic analyses of bacterial porin-cytochrome gene clusters. Frontiers in microbiology 5: 657.

Shieh WY, Jean WD (1998). Alterococcus agarolyticus, gen.nov., sp.nov., a halophilic thermophilic bacterium capable of agar degradation. Canadian journal of microbiology 44: 637-645.

Shyu JBH, Lies DP, Newman DK (2002). Protective Role of tolC in Efflux of the Electron Shuttle Anthraquinone-2,6-Disulfonate. Journal of bacteriology 184: 1806-1810.

Spring S, Bunk B, Sproer C, Schumann P, Rohde M, Tindall BJ, Klenk HP (2016). Characterization of the first cultured representative of Verrucomicrobia subdivision 5 indicates the proposal of a novel phylum. The ISME journal 10: 2801-2816.

Studholme DJ, Fuerst JA, Bateman A (2004). Novel protein domains and motifs in the marine planctomycete Rhodopirellula baltica. FEMS microbiology letters 236: 333-340.

Voordeckers JW, Kim BC, Izallalen M, Lovley DR (2010). Role of Geobacter sulfurreducens outer surface c-type cytochromes in reduction of soil humic acid and anthraquinone-2,6-disulfonate. Applied and environmental microbiology 76: 2371-2375.

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Wertz JT, Kim E, Breznak JA, Schmidt TM, Rodrigues JLM (2012). Genomic and Physiological Characterization of the Verrucomicrobia Isolate Diplosphaera colitermitum gen. nov., sp. nov., Reveals Microaerophily and Nitrogen Fixation Genes. Applied and environmental microbiology 78: 1544-1555.

Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 40: W445-451.

Yoon J, Yasumoto-Hirose M, Katsuta A, Sekiguchi H, Matsuda S, Kasai H, Yokota A (2007). Coraliomargarita akajimensis gen. nov., sp. nov., a novel member of the phylum 'Verrucomicrobia' isolated from seawater in Japan. International journal of systematic and evolutionary microbiology 57: 959-963.

Yoon J, Matsuo Y, Katsuta A, Jang JH, Matsuda S, Adachi K, Kasai H, Yokota A (2008). Haloferula rosea gen. nov., sp. nov., Haloferula harenae sp. nov., Haloferula phyci sp. nov., Haloferula helveola sp. nov. and Haloferula sargassicola sp. nov., five marine representatives of the family Verrucomicrobiaceae within the phylum 'Verrucomicrobia'. International journal of systematic and evolutionary microbiology 58: 2491-2500.

Yu CS, Chen YC, Lu CH, Hwang JK (2006). Prediction of protein subcellular localization. Proteins 64: 643-651.

Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FSL (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26: 1608-1615.

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 microbiology ecology 25: 159-169.

Zwart G, van Hannen EJ, Kamst-van Agterveld MP, Van der Gucht K, Lindström ES, Van Wichelen J, Lauridsen T, Crump BC, Han S-K, Declerck S (2003). Rapid Screening for Freshwater Bacterial Groups by Using Reverse Line Blot Hybridization. Applied and environmental microbiology 69: 5875-5883.

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TABLES

Table 1. Lakes included in this study[a]

Lake	Mendota	Trout Bog
GPS location	43.100°N, 89.405°W	46.041°N, 89.686°W
Lake type	Drainage lake	Seepage lake
Surface area (ha)	3938	1.1
Mean depth (m)	12.8	5.6
Max depth (m)	25.3	7.9
pH	8.3	5.2
Primary carbon source	Phytoplankton	Terrestrial subsidies
DOC (mg/L)	5.6	19.9
Total N (mg/L)	1.5	1.3
Total P (µg/L)	131	71
DOC/N	3.7	15.5
DOC/P	42.6	281.0
Trophic state	Eutrophic	Dystrophic

[a]Data from NTL-LTER https://lter.limnology.wisc.edu, averaged from the study years. DOC = Dissolved organic carbon. N = Nitrogen. P = Phosphorus.

Table 2. Characteristics of Verrucomicrobia MAGs

Genome	IMG Taxon OID	Subdivision	Genome Completeness (%)	Recovered Genome Size (Mbp)	GC content (%)	Coding Base (%)	Gene Count
ME3880	2582580573	1	76	1.6	58	90.89	1585
TH2746	2582580664	1	87	6.5	62	86.67	5430
ME12612	2582580523	1	88	2.2	59	88.98	2335
ME12173	2582580521	1	56	2.1	52	91.29	2070
TE4605	2582580638	1	89	4.7	59	91.07	4380
ME6381	2582580593	1	68	2.4	57	92.45	2221
ME8366	2582580607	2	74	3.6	63	87.36	3450
TH2747	2582580665	3	89	5.2	58	89.64	4846
TH3004	2582580668	3	88	4.5	57	91.37	3798
TH0989	2556921153	3	95	7.2	62	90.33	5583
TH2519	2593339181	4	84	1.8	42	94.34	1654
TE1800	2593339189	4	92	2.2	42	94.29	1998
TH4590	2582580688	4	92	3.3	65	90.73	3132
ME2014	2582580546	4	53	1.9	66	93.72	1700
ME12657	2582580524	4	82	1.9	68	94.01	1838
TE1301	2582580616	4	96	2.0	54	94.74	1943
TH4093	2582580682	Unclassified	70	4.7	48	86.52	3982
ME30509	2582580559	Unclassified	78	1.2	63	92.25	1160
TH4820	2582580691	Unclassified	64	3.0	63	86.71	2794

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURES

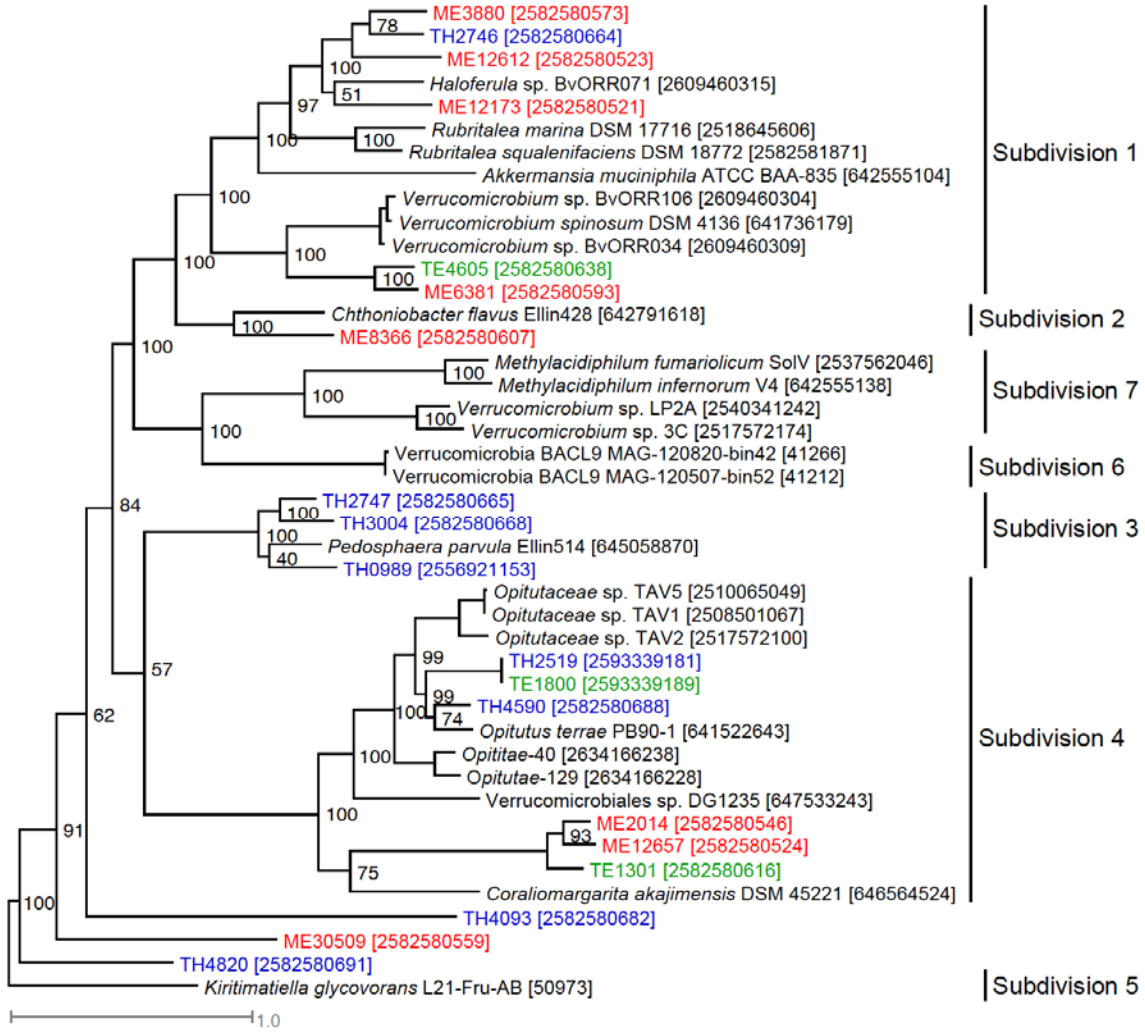

Figure 1. Phylogenetic tree constructed with a concatenated alignment of protein sequences from five conserved essential single-copy genes, which were found in all MAGs. ME, TE and TH MAGs are labeled with red, green and blue, respectively. Genome ID in IMG or NCBI is indicated in the bracket. The outgroup is *Kiritimatiella glycovorans* L21-Fru-AB, which was initially assigned to subdivision 5, but this subdivision was recently proposed as a novel sister phylum to Verrucomicrobia (Spring et al 2016).

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

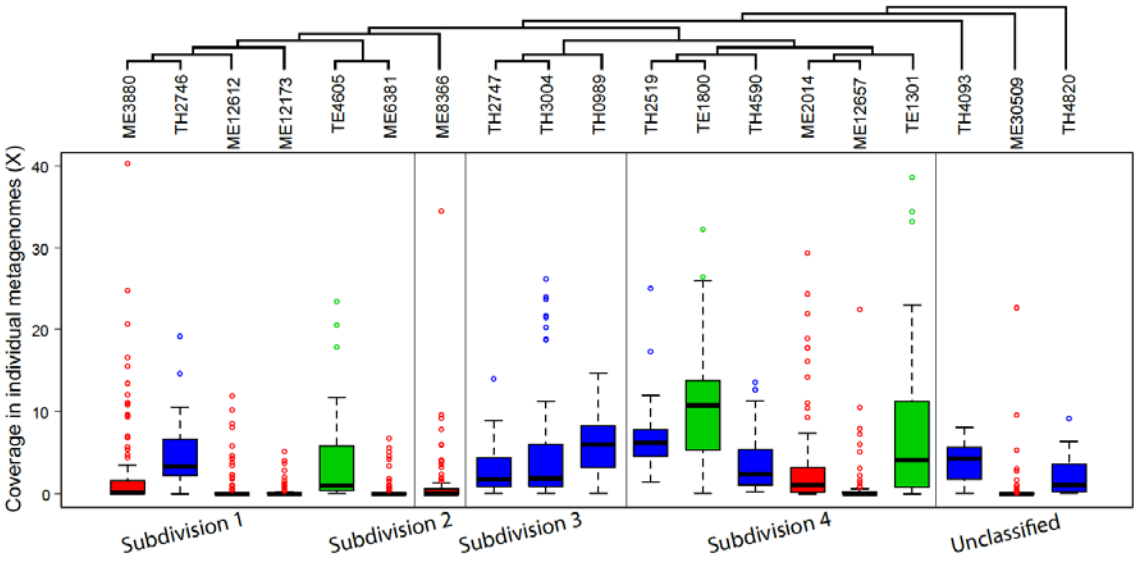

Figure 2. A boxplot showing the fold coverage of the 19 MAGs in individual metagenomes. The boxplot whiskers are defined by 1.5 times the interquartile range above the upper quartile and below the lower quartile, respectively. ME, TE and TH MAGs are labeled with red, green and blue, respectively. Genome clustering on the top is based on a subtree extracted from the phylogenetic tree in Figure 1 to indicate the phylogenetic relatedness of the 19 MAGs.

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

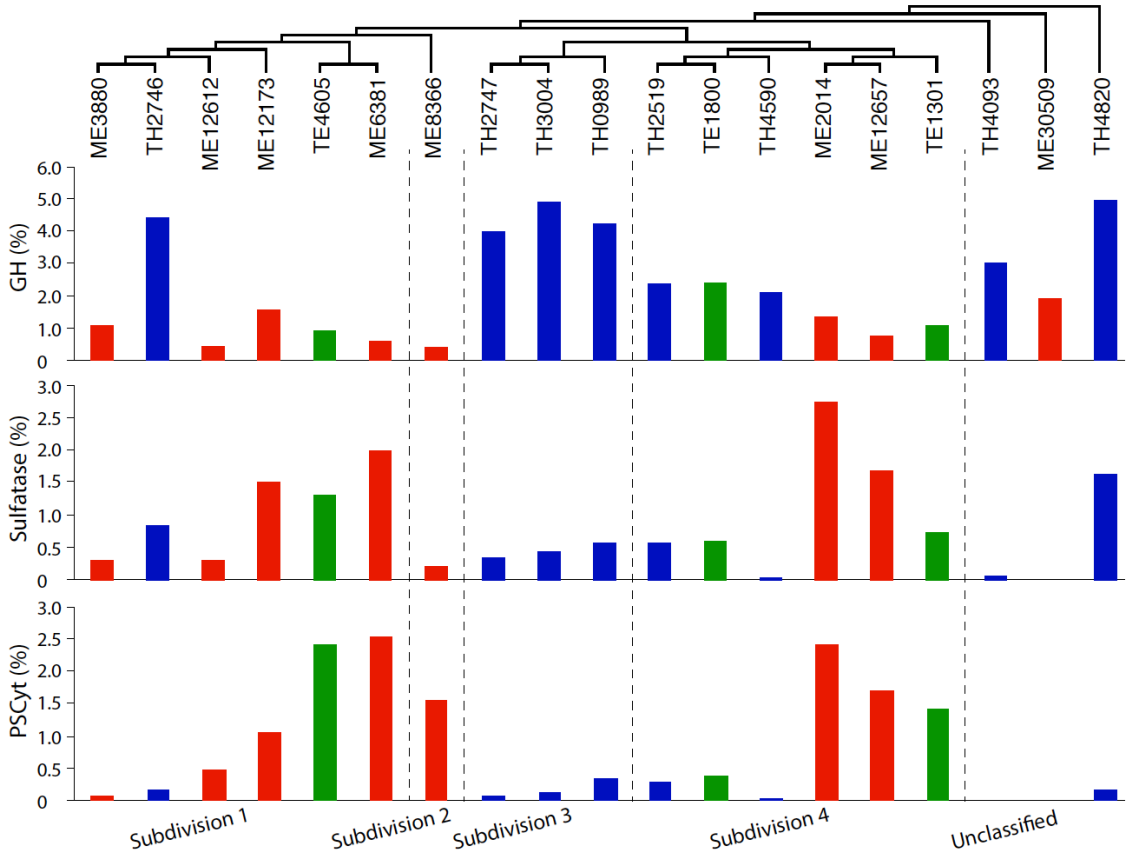

Figure 3. Coding densities of glycoside hydrolase genes (**a**), sulfatase genes (**b**) and Planctomycete-specific cytochrome *c* (PSCyt)-containing genes (**c**). Data from ME, TE and TH MAGs are labeled with red, green and blue, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

GH families	Main Activities	ME3880	TH2746	ME12612	ME12173	TE4605	ME6381	ME8366	TH2747	TH3004	TH0989	TH2519	TE1800	TH4590	ME2014	ME12657	TE1301	TH4093	ME30509	TH4820
GH29	α-fucosidases	2	22	0	1	0	0	0	17	19	24	4	5	6	0	0	0	14	3	13
GH2	β-galactosidases and other β-linked dimers	1	24	0	1	2	0	0	18	14	20	5	5	5	0	0	0	6	2	7
GH78	α-L-rhamnosidases	1	31	0	2	0	0	0	11	11	13	3	5	9	0	0	0	5	2	7
GH95	1,2-α-L-fucosidase	0	29	0	1	0	0	0	7	9	17	1	2	3	0	0	0	4	0	7
GH106	α-L-rhamnosidase	0	21	1	0	0	0	0	4	9	12	1	2	1	0	0	0	3	1	10
GH13	α-amylase	3	3	2	2	3	2	1	7	6	7	2	3	2	2	2	2	3	0	3
GH20	β-hexosaminidase	0	14	0	4	0	0	0	3	5	2	2	2	3	1	2	4	10	1	3
GH5	endoglucanase, endomannanase, β-glucosidase, β-mannosidase	0	4	0	1	4	1	3	13	4	10	0	0	0	0	0	1	6	1	6
GH28	polygalacturonases, related to pectin degradation	0	1	0	0	0	0	0	7	11	2	8	8	3	1	0	1	0	1	3
GH43	α-L-arabinofuranosidases, endo-α-L-arabinanases, β-D-xylosidases	1	4	0	2	2	0	0	11	15	9	1	1	2	0	0	0	0	0	2
Counts of other GH genes		9	86	7	18	29	10	10	95	83	119	12	15	32	19	10	13	68	11	77
Total counts of all GH genes		**17**	**239**	**10**	**32**	**40**	**13**	**14**	**193**	**186**	**235**	**39**	**48**	**66**	**23**	**14**	**21**	**119**	**22**	**138**
Total number of GH families represented		**12**	**48**	**7**	**23**	**23**	**7**	**10**	**53**	**49**	**58**	**19**	**21**	**26**	**10**	**10**	**13**	**35**	**15**	**45**

Figure 4. Gene counts for the top 10 most abundant GH families, total gene counts for all GH families, and the number of GH families represented by these genes. MAGs are ordered as in the clustering in Figure 2.

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

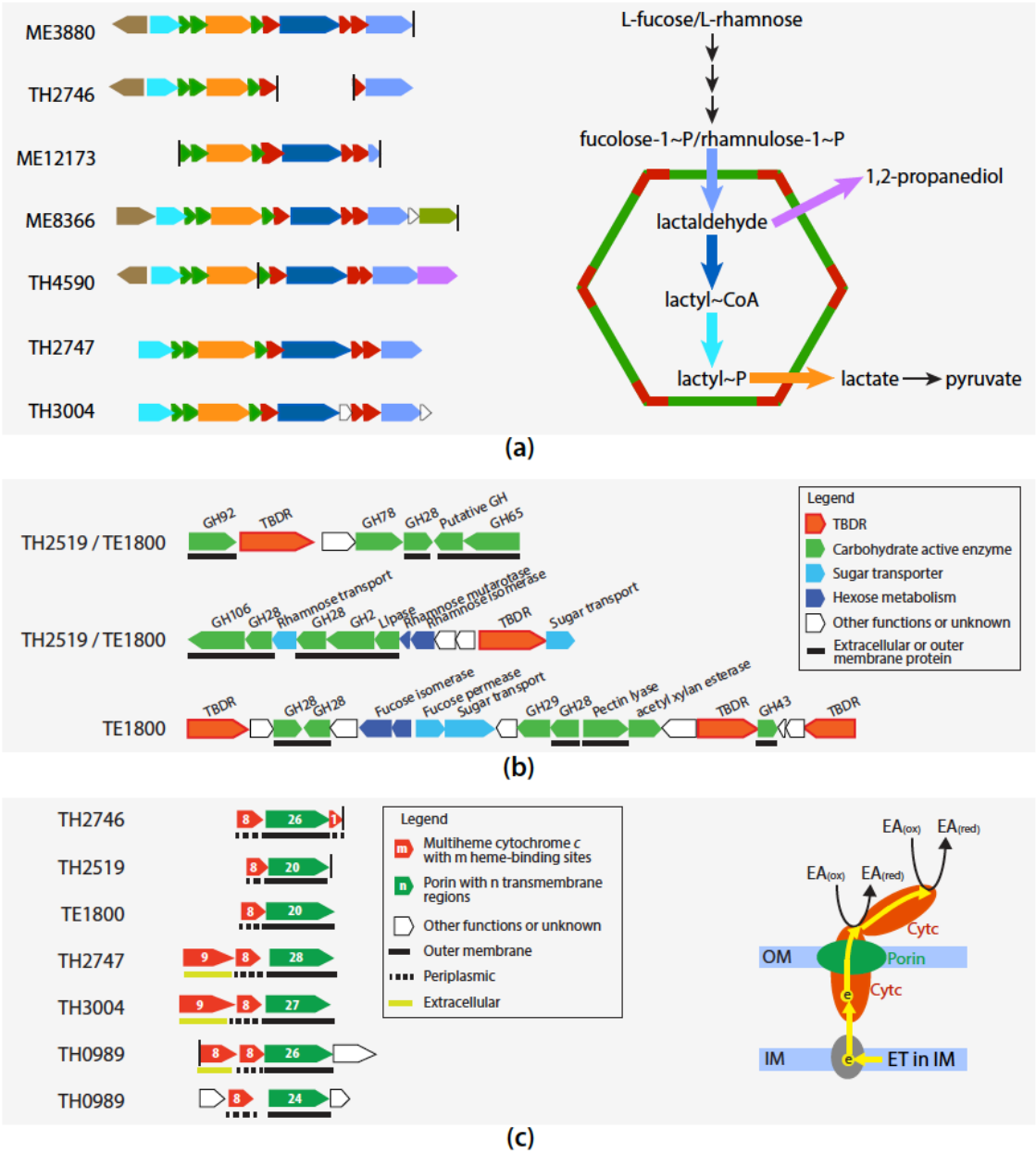

Figure 5. Gene clusters encoding BMCs involved in L-fucose and L-rhamnose degradation (**a**), putative tonB-dependent carbohydrate utilization (CUT) loci (**b**), and putative porin-multiheme cytochrome *c* complex (**c**). The vertical line indicates the end of a contig, and horizontal lines below genes indicate predicted cellular locations of their encoded proteins. In (**a**), the two building blocks of the BMC and reactions inside the BMC are colored according to their encoding genes' color labels on the left side. In (**c**), putative EET genes are shown, and these clusters are in 18.1, 9.0, 6.1, 18.4, 70.0, 10.6 and 10.8 kbp long contigs, respectively. In (**c**), a hypothesized model of extracellular electron transfer is shown on the right with yellow arrows indicating electron flows. "ET in IM", electron transfer in the inner membrane, "$EA_{(ox)}$" and "$EA_{(red)}$", oxidized and reduced forms of the electron acceptor, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genes and Pathways	ME3880	TH2746	ME12612	ME12173	TE4605	ME6381	ME8366	TH2747	TH3004	TH0908	TH2519	TE1800	TH4590	ME2014	ME12657	TE1301	TH4093	ME30509	TH4820
Central carbon metabolism																			
Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate	0.78	0.89	0.89	0.67	1	0.89	0.78	1	0.89	1	0.67	0.89	1	0.89	0.56	1	0.89	0.56	0.33
Glycolysis (Entner-Doudoroff pathway)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Pentose phosphate pathway (Pentose phosphate cycle)	0.86	0.71	0.71	0.71	0.86	0.57	0.71	1	0.86	1	0.71	0.86	1	0.86	0.57	1	0.71	0.29	0.43
Pyruvate oxidation, pyruvate => acetyl-CoA	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Citrate cycle (TCA cycle, Krebs cycle)	0.63	0.88	1	0.38	0.75	0.63	0.88	0.88	0.88	0.75	1	1	0.88	0.63	0.75	1	0.5	0.25	0.25
Other carbohydrate metabolism																			
Galactose degradation to glycerate-3P	0.5	0.75	0.75	0.25	0.75	0.5	0.5	1	0.75	1	0.5	0.5	0.75	0.25	0.25	0.5	0.75	0.25	0.5
Rhamnose degradation	1	1	0.75	0.5	0.5	0.5	0.5	1	1	1	0.75	1	0.75	0.5	0.5	0.5	1	0.5	0.75
Fucose degradation	1	0.75	0.75	0.25	0.75	0.5	0.75	1	0.75	0.75	0.5	0.75	0.75	0.75	0.5	0.75	1	0.75	0.75
L-Arabinose degradation to xylulose-P for pentose pathway	1	1	0	0	0.67	0.33	0.33	1	1	1	0	0	0.33	0	0	0	0.67	0.33	0.33
Xylose degradation	1	1	1	1	1	0.5	1	1	1	1	0.5	1	1	1	1	1	0.5	1	1
D-Galacturonate degradation to pyruvate & D-glyceraldehyde 3P	(0)	(0)	(0)	(0)	0.8	0.2	(0)	0.6	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
D-Glucuronate degradation to pyruvate and D-glyceraldehyde 3P	0	0.6	0.4	0.4	0.6	0	0.2	0.6	0.8	0.6	0.6	0.8	0.6	0.2	0.4	0.4	0.6	0.4	0.8
Mannose degradation to glucose-P	0.5	1	1	1	1	1	1	1	1	1	1	1	1	1	0.5	0.5	1	0.5	1
Lactaldehyde degradation to pyruvate (Aerobic)	1	1	0.33	1	0.33	0.33	1	1	1	0.67	0.33	0.33	1	0.33	0.33	0.33	0.67	0.67	0.67
Glycogen biosynthesis from alpha-D-glucose-6P via ADP-D glucose	0.75	0.25	0.75	0.75	0.75	0.5	1	1	0.75	1	0.5	0.5	1	1	1	1	0.5	0	0.75
Fermentation																			
Pyruvate to acetate via acetyl-coA	1	1	0.67	0.67	0.33	0.33	1	1	1	0.67	0.33	0.33	1	0.67	0.67	0.67	1	1	1
Pyruvate to propanoate	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Pyruvate to succinate	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Pyruvate to butanoate	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Pyruvate to butanol	0.71	0.71	0.71	0.29	1	0.71	0.57	0.43	0.43	0.43	0.57	0.57	1	0.57	0.57	0.57	0.43	0.43	0.43
Pyruvate to ethanol	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Pyruvate to lactate	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
Pyruvate to acetone	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Nitrogen related																			
Dissimilatory nitrate reduction, nitrate => ammonia	0	0	0	0	0	0	0	0	0	0	0	0	(0)	0	0	0	0	0	(0)
Denitrification, nitrate => nitrogen gas	0	0	0	0	0	0	0	0.25	0.25	0	0	0	0	0	0	0	0	0	0
TMAO (trimethylamine-N-oxide) reductioin	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nitrification, ammonia => nitrite	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nitrogen fixation, nitrogen => ammonia	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0
Assimilatory nitrate reduction, nitrate => ammonia	0	1	0	0	(0)	0	0	0	0	0	0	0	0	0	0	0	(0)	0	(0)
ABC-type urea transporter	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
Urease (Urea ==> CO2 + NH3)	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0
Ammonia permease	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	1	0	1
Phosphorus related																			
Alkaline phosphatase (PhoA)	1	1	1	1	0	0	1	0	0	0	0	0	0	1	1	1	1	0	0
ABC-type phosphate-specific transport (Pst) system, high-affinity	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0
PiT inorganic phosphate transporter (PitA), permease, low-affinity	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1
Polyphosphate storage and utilization (PPK)	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
ABC-type phosphonate transport system	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Phosphonoacetate degradation	0	1	0	1	1	0	1	1	1	0	0	0	1	0	0	0	0	0	0

Figure 6. Completeness estimates of key metabolic pathways. MAGs are ordered as in the clustering in Figure 2. Completeness value of "1" indicates a pathway is complete; "0" indicates no genes were found in that pathway; and "(0)" indicates that although some genes in a pathway are present, the pathway is likely absent because signature genes for that pathway were not found in that draft genome AND signature genes are missing in more than two thirds of all draft genomes.

bioRxiv preprint doi: https://doi.org/10.1101/150078; this version posted June 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

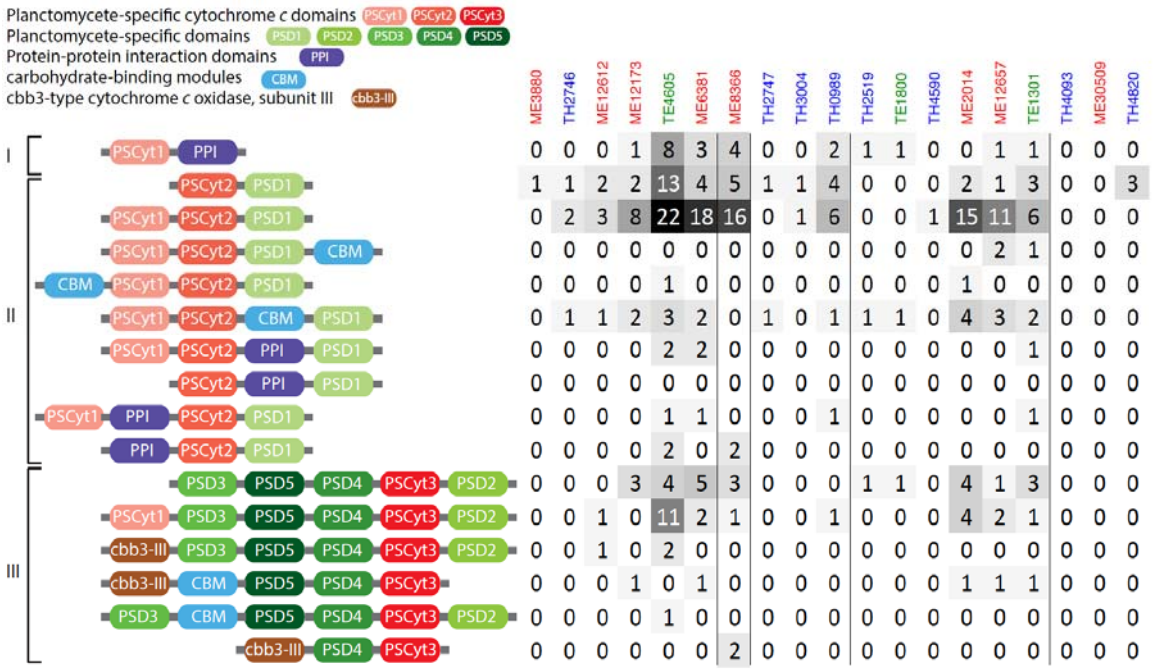

Figure 7. Domain architecture and occurrence of PSCyt-containing genes. Based on the combination of specific PSCyt and PSD domains, these domain structures can be classified into three groups (I, II, and III). "CBM" refers to carbohydrate-binding modules, which include pfam13385 (Laminin_G_3), pfam08531 (Bac_rhamnosid_N), pfam08305 (NPCBM), pfam03422 (CBM_6), and pfam07691 (PA14). "PPI" refers to protein-protein interaction domains, which include pfam02368 (Big_2), pfam00400 (WD40), and pfam00754 (F5_F8_type_C).