Ecophysiology of freshwater Verrucomicrobia inferred from metagenome-assembled genomes

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 (trina.mcmahon@wisc.edu)

Running title: Freshwater Verrucomicrobia genomes

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

Word count for the Abstract: 226

Word count for the text (excluding references and figure legends): 5850

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 freshwater Verrucomicrobia. Genome analysis revealed their potential 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. 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 so-called "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.

IMPORTANCE

Freshwater Verrucomicrobia are cosmopolitan in lakes and rivers, yet their roles and ecophysiology are not well understood, as cultured freshwater Verrucomicrobia are restricted to one subdivision of this phylum. Here, we greatly expand the known genomic diversity of this freshwater lineage by recovering 19 Verrucomicrobia draft genomes from 184 metagenomes collected from a eutrophic lake and a humic bog across multiple years. Most of these genomes represent first freshwater representatives of several Verrucomicrobia subdivisions. Genomic analysis revealed Verrucomicrobia as potential (poly)saccharide-degraders, and suggested their adaptation to carbon source of different origins in the two contrasting ecosystems. We identified putative extracellular electron transfer genes and so-called "Planctomycete-specific" cytochrome *c*-containing genes, and found their distinct distribution patterns between the lakes/layers. Overall, our analysis greatly advances the understanding of the function, ecophysiology and distribution of freshwater Verrucomicrobia, while highlighting their potential role in freshwater carbon cycling.

INTRODUCTION

Verrucomicrobia are ubiquitous in freshwater and exhibit a cosmopolitan distribution in lakes and rivers. They are present in up to 90% of lakes (1), with abundances typically between <1% and 6% of total microbial community (2-4), but as high as 19% in a humic lake (5). 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.

73 As a phylum, Verrucomicrobia (V) was first proposed relatively recently, in 1997
74 (6). Together with Planctomycetes (P), Chlamydiae (C), and sister phyla such as
75 Lentisphaerae, they comprise the PVC superphylum. In addition to being cosmopolitan in
76 freshwater, Verrucomicrobia have been found in oceans (7, 8), soil (9, 10), wetlands (11),
77 rhizosphere (12), and animal guts (13, 14), as free-living organisms or symbionts of
78 eukaryotes. Verrucomicrobia isolates are metabolically diverse, including aerobes,
79 facultative anaerobes, and obligate anaerobes, and they are mostly heterotrophs, using
80 various mono-, oligo-, and poly-saccharides for growth (6, 7, 11, 14-20). Not long ago an
81 autotrophic verrucomicrobial methanotroph (*Methylacidiphilum fumariolicum* SolV) was
82 discovered in acidic thermophilic environments (21).

83 In marine environments, Verrucomicrobia are also ubiquitous (22) and suggested to
84 have a key role as polysaccharide degraders (23, 24). Genomic insights gained through
85 sequencing single cells (24) or extracting Verrucomicrobia bins from metagenomes (25)
86 have revealed high abundances of glycoside hydrolase genes, providing more evidence for
87 their critical roles in C cycling in marine environments.

88 In freshwater, Verrucomicrobia have been suggested to degrade glycolate (26) and
89 polysaccharides (24). The abundance of some phylum members was favored by high
90 nutrient availabilities (27, 28), cyanobacterial blooms (29), low pH, high temperature, high
91 hydraulic retention time (30), and more labile DOC (5). To date, there are very few
92 freshwater Verrucomicrobia isolates, including *Verrucomicrobium spinosum* (31) and
93 several *Prosthecobacter* spp. (6). Physiological studies showed that they are aerobes,
94 primarily using carbohydrates, but not amino acids, alcohols, or rarely organic acids for
95 growth. However, these few cultured isolates only represent a single clade within

96 subdivision 1. By contrast, 16S rRNA gene based studies discovered a much wider
97 phylogenic range of freshwater Verrucomicrobia, including subdivisions 1, 2, 3, 4, 5, and 6
98 (3-5, 24, 32). Due to the very few cultured representatives and few available genomes from
99 this freshwater lineage, the ecological functions of the vast uncultured freshwater
100 Verrucomicrobia are largely unknown.

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

111

112 **RESULTS AND DISCUSSION**

113 **Comparison of the two lakes**

114 The two studied lakes exhibited contrasting characteristics (**Table 1**). The most notable
115 difference is the primary C source and nutrient availabilities. Mendota is an urban
116 eutrophic lake with most of its C being autochthonous (in-lake produced through
117 photosynthesis). By contrast, Trout Bog is a nutrient-poor dystrophic lake, surrounded by
118 temperate forests and sphagnum mats, thus receiving large amounts of terrestrially-

119 derived allochthonous C that is rich in humic and fulvic acids. Compared to Mendota, Trout
120 Bog features higher DOC levels, but is more limited in nutrient availability, with much
121 higher DOC:TN and DOC:TP ratios (**Table 1**). Nutrient limitation in Trout Bog is even more
122 extreme than revealed by these ratios because much of the N and P is tied up in complex
123 dissolved organic matter. In addition, Trout Bog has lower oxygenic photosynthesis due to
124 decreased photosynthetically active radiation (PAR) as a result of absorption by DOC (33).
125 Together with the consumption of dissolved oxygen by heterotrophic respiration, oxygen
126 levels decrease quickly with depth in the water column in Trout Bog. Dissolved oxygen
127 levels are below detection in the hypolimnion nearly year-round (34). Due to these
128 contrasts, we expected to observe differences in bacterial C and nutrient use, as well as
129 differences reflecting the electron acceptor conditions between these two lakes. Hence, the
130 retrieval of numerous Verrucomicrobia draft genomes in the two lakes not only allows the
131 prediction of their general functions in freshwater, but also provides an opportunity to
132 study their ecophysiological adaptation to the local environmental differences.
133

134 **Verrucomicrobia draft genome retrieval and their distribution patterns**

135 A total of 184 metagenomes were generated from samples collected across multiple years,
136 including 94 from the top 12 m of Mendota (mostly consisting of the epilimnion layer,
137 therefor referred to as "ME"), 45 from Trout Bog epilimnion ("TE"), and 45 from Trout Bog
138 hypolimnion ("TH"). Three combined assemblies were generated by co-assembling reads
139 from all metagenomes within the ME, TE, and TH groups, respectively. Using the binning
140 facilitated by tetranucleotide frequency and relative abundance patterns over time, a total
141 of 19 Verrucomicrobia metagenome-assembled genomes (MAGs) were obtained, including

142 eight from the combined assembly of ME, three from the combined assembly of TE, and
143 eight from the combined assembly of TH (**Table 2**). The 19 MAGs exhibited a clustering of
144 their tetranucleotide frequency largely based on the two lakes (**Fig. S1**), suggesting distinct
145 overall genomic signatures associated with each system.

146 Genome completeness of the 19 MAGs ranged from 51% to 95%, as determined by
147 checkM (35). Phylogenetic analysis of these MAGs using a concatenated alignment of their
148 conserved genes indicates that they span a wide phylogenetic spectrum and distribute in
149 subdivisions 1, 2, 3, and 4 of the seven previously defined Verrucomicrobia subdivisions (5,
150 21, 36) (**Fig. 1**), as well as three unclassified Verrucomicrobia MAGs.

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

157 We used normalized coverage depth of MAGs within individual metagenomes
158 collected at different sampling time points and different lakes/layers to comparatively infer
159 relative population abundance across time and space (see detailed coverage depth
160 estimation in **Supplementary Text**). Briefly, we mapped reads from each metagenome to
161 MAGs with a minimum identity of 95%, and used the number of mapped reads to calculate
162 the relative abundance for each MAG based on coverage depth per contig and several
163 normalization steps. Thus, we assume that each MAG represents a distinct population
164 within the lake-layer from which it was recovered (37, 38). This estimate does not directly

165 indicate the actual relative abundance of these populations within the total community per
166 se; rather it allows us to compare population abundance levels from different lakes and
167 sampling occasions within the set of 19 MAGs. This analysis indicates that Verrucomicrobia
168 populations in Trout Bog were proportionally more abundant and persistent over time
169 compared to those in Mendota in general (**Table 2**). Verrucomicrobia populations in
170 Mendota boosted their abundances once to a few times during the sampling season and
171 diminished to extremely low levels for the remainder of the sampling season (generally
172 May to November), as reflected by the low median coverage depth of Mendota MAGs and
173 their large coefficient of variation (**Table 2**)
174
175 **Saccharolytic life style and adaptation to different C sources**
176 Verrucomicrobia isolates from different environments are known to grow on various
177 mono-, oligo-, and poly-saccharides, but are unable to grow on amino acids, alcohols, or
178 most organic acids (6, 7, 11, 14-20, 39). Culture-independent research suggests marine
179 Verrucomicrobia as candidate polysaccharide degraders with large number of genes
180 involved in polysaccharide utilization (23-25).
181 In the 19 Verrucomicrobia MAGs, we observed rich arrays of glycoside hydrolase
182 (GH) genes, representing a total of 78 different GH families acting on diverse
183 polysaccharides (**Fig. S2**). Although these genomes have different degrees of completeness,
184 genome completeness was not correlated with the number of GH genes recovered
185 (correlation coefficient = 0.312, p-value = 0.194), or the number of GH families represented
186 in each MAG (i.e. GH diversity, correlation coefficient = 0.278, p-value = 0.250). To compare
187 GH abundance among MAGs, we normalized GH occurrence frequencies by the total

188 number of genes in each MAG to estimate the percentage of genes annotated as GHs (i.e. GH
189 coding density) to account for the different genome size and completeness. This
190 normalization assumes GH genes are randomly distributed between the recovered and the
191 missing parts of the genome, and it allows us to make some general comparison among
192 these MAGs. GH coding density ranged from 0.4% to 4.9% for these MAGs (**Fig. 2a**), and in
193 general, was higher in Trout Bog MAGs than in Mendota MAGs. Notably, six TH MAGs had
194 extremely high (~4%) GH coding densities (**Fig. 2a**), with each MAG harboring 119-239 GH
195 genes, representing 36-59 different GH families (**Fig. 3 and S2**). Although GH coding
196 density in most ME genomes in subdivisions 1 and 2 was relatively low (0.4-1.6%), it was
197 still higher than in many other bacterial groups (24).

198 The GH abundance and diversity within a genome may determine the width of the
199 substrate spectrum and/or the complexity of carbohydrates used by that organism. For
200 example, there are 20 GH genes in the *Rubritalea marina* genome, and this marine
201 verrucomicrobial aerobe only uses a limited spectrum of carbohydrate monomers and
202 dimers, but not the majority of (poly)saccharides tested (15). By contrast, 164 GH genes
203 are present in the *Opitutus terrae* genome, and this soil verrucomicrobial anaerobe can thus
204 grow on a wider range of mono-, di- and poly-saccharides (16). Therefore, it is plausible
205 that the GH-rich Trout Bog Verrucomicrobia populations may be able to use a wider range
206 of more complex polysaccharides than the Mendota populations.

207 The 10 most abundant GH families in these Verrucomicrobia MAGs include GH2, 29,
208 78, 95, and 106 (**Fig. 3**). These specific GHs were absent or at very low abundances in
209 marine Verrucomicrobia genomes (24, 25), suggesting a general difference in carbohydrate
210 substrate use between freshwater- and marine Verrucomicrobia. Hierarchical clustering of

211 MAGs based on overall GH abundance profiles indicated a grouping pattern largely
212 separated by lake (**Fig. S3**). Prominently over-represented GHs in most Trout Bog MAGs
213 include GH2, GH29, 78, 95, and 106. By contrast, over-represented GHs in the Mendota
214 MAGs are GH13, 20, 33, 57, and 77, which have different substrate spectra from GHs over-
215 represented in the Trout Bog MAGs. Therefore, the patterns in GH functional profiles may
216 suggest varied carbohydrate substrate preferences and ecological niches occupied by
217 Verrucomicrobia, probably reflecting the different carbohydrate composition derived from
218 different sources between Mendota and Trout Bog.

219 Overall, GH diversity and abundance profile may reflect the DOC availability,
220 chemical variety and complexity, and suggest microbial adaptation to different C sources in
221 the two ecosystems. We speculate that the rich arrays of GH genes, and presumably
222 broader substrate spectra of Trout Bog populations, partly contribute to their higher
223 abundance and persistence over the sampling season (**Table 2**), as they are less likely
224 impacted by fluctuations of individual carbohydrates. By contrast, Mendota populations
225 with fewer GHs and presumably more specific substrate spectra are relying on
226 autochthonous C and therefore exhibit a "bloom-and-bust" abundance pattern (**Table 2**)
227 that might be associated with cyanobacterial blooms as previous suggested (29). On the
228 other hand, bogs experience seasonal phytoplankton blooms (40, 41) that introduce brief
229 pulses of autochthonous C to these otherwise allochthonous-driven systems. Clearly, much
230 remains to be learned about the routes through which C is metabolized by bacteria in such
231 lakes, and comparative genomics is a novel way to use the organisms to tell us about C flow
232 through the ecosystem.

233

234 **Other genome features of the saccharide-degrading life style**

235 Seven Verrucomicrobia MAGs spanning subdivisions 1, 2, 3, and 4 possess genes needed to
236 construct bacterial microcompartments (BMCs), which are quite rare among studied
237 bacterial lineages. Such BMC genes in Planctomycetes are involved in the degradation of
238 plant and algal cell wall sugars, and are required for growth on L-fucose, L-rhamnose and
239 fucoidans (42). Genes involved in L-fucose and L-rhamnose degradation cluster with BMC
240 shell protein-coding genes in the seven Verrucomicrobia MAGs (**Fig. 4**). This is consistent
241 with the high abundance of α-L-fucosidase or α-L-rhamnosidase GH genes (represented
242 by GH29, 78, 95, 106) in most of these MAGs (**Fig. 3**), suggesting the importance of fucose-
243 and rhamnose-containing polysaccharides for these Verrucomicrobia populations.

244 TonB-dependent receptor (TBDR) genes were found in Verrucomicrobia MAGs, and
245 are present at over 20 copies in TE1800 and TH2519. TBDRs are located on the outer
246 cellular membrane of Gram-negative bacteria, usually mediating the transport of iron
247 siderophore complex and vitamin B_{12} across the outer membrane through an active
248 process. More recently, TBDRs were suggested to be involved in carbohydrate transport
249 across the outer membrane by some bacteria that consume complex carbohydrates, and in
250 their carbohydrate utilization (CUT) loci, TBDR genes usually cluster with genes encoding
251 inner membrane transporters, GHs and regulators for efficient carbohydrate
252 transportation and utilization (43). Such novel CUT loci are present in TE1800 and
253 TH2519, with TBDR genes clustering with genes encoding inner membrane sugar
254 transporters, monosaccharide utilization enzymes, and GHs involved in the degradation of
255 pectin, xylan, and fucose-containing polymers (**Fig. 5**). Notably, most GHs in the CUT loci
256 are predicted to be extracellular or outer membrane proteins (**Fig. 5**), catalyzing

257 extracellular hydrolysis reactions to release mono- and oligo-saccharides, which are
258 transported across the outer membrane by TBDR proteins. Therefore, such CUT loci may
259 allow these verrucomicrobial populations to coordinately and effectively scavenge the
260 hydrolysis products before they diffuse away.

261 Genes encoding for inner membrane carbohydrate transporters are abundant in
262 Verrucomicrobia MAGs (**Fig. S4**). The Embden-Meyerhof pathway for glucose degradation,
263 as well as pathways for degrading a variety of other sugar monomers, including galactose,
264 rhamnose, fucose, xylose, and mannose, were recovered (complete or partly-complete) in
265 most MAGs (**Fig. 6**). As these sugars are abundant carbohydrate monomers in plankton and
266 plant cell walls, the presence of these pathways together with GH genes suggest that these
267 Verrucomicrobia populations may use plankton- and plant-derived saccharides. Machinery
268 for pyruvate degradation to acetyl-CoA and the TCA cycle are also present in most MAGs.
269 These results are largely consistent with their hypothesized role in carbohydrate
270 degradation and previous studies on Verrucomicrobia isolates.

271 Notably, a large number of genes encoding proteins belonging to a sulfatase family
272 (pfam00884) are present in the majority of MAGs (**Fig. 2b**), similar to the high
273 representation of these genes in marine Verrucomicrobia genomes (24, 25). Sulfatases
274 hydrolyze sulfate esters, which are rich in sulfated polysaccharides. In general, sulfated
275 polysaccharides are abundant in marine algae and plants (mainly in seaweeds) (44), but
276 have also been found in some freshwater cyanobacteria (45) and plant species (46).
277 Sulfatase genes in our Verrucomicrobia MAGs were often located in the same neighborhood
278 as genes encoding for extracellular proteins with a putative pectin lyase activity, proteins
279 with a carbohydrate-binding module (pfam13385), GHs, and proteins with PSCyt domains

280 (**Fig. 2c** and discussed later). Their genome context lends support for the participation of
281 these genes in C and sulfur cycling by degrading sulfated polysaccharides, which can serve
282 as an abundant source of sulfur for cell biosynthesis as well as C for energy and growth.

283 Previously, freshwater Verrucomicrobia were suggested to use the algal exudate
284 glycolate in humic lakes, based on the retrieval of genes encoding subunit D (*glcD*) of
285 glycolate oxidase, which converts glycolate to glyoxylate (26). However, these recovered
286 genes might not be bona fide *glcD* due to the lack of other essential subunits as revealed in
287 our study (see **Supplementary Text**). Among the MAGs, only TE4605 possesses all three
288 essential subunits of glycolate oxidase (*glcDEF)* (**Fig. S5**). However, genetic context analysis
289 suggests that TE4605 likely uses glycolate for amino acid assimilation, instead of energy
290 generation (**Fig. S5** and **Supplementary Text**). These results are consistent with the
291 absence of the glyoxylate shunt and especially the malate synthase, which converts
292 glyoxylate to malate to be used through the TCA cycle for energy generation in the 19
293 Verrucomicrobia MAGs (**Fig. S6**). Therefore, Verrucomicrobia populations represented by
294 the 19 MAGs are not likely key players in glycolate degradation, but more likely important
295 (poly)saccharide-degraders in freshwater, as suggested by the high abundance of GH,
296 sulfatase, and carbohydrate transporter genes, metabolic pathways for degrading diverse
297 carbohydrate monomers, and other genome features adapted to the saccharolytic life style.
298

299 **Nitrogen (N) metabolism and adaptation to different N availabilities**

300 Most Verrucomicrobia MAGs in our study do not appear to reduce nitrate or other
301 nitrogenous compounds, and they seem to uptake and use ammonia (**Fig. 6**), and
302 occasionally amino acids (**Fig. S4**), as an N-source. Further, some Trout Bog populations

303 may have additional avenues to generate ammonia, including genetic machineries for
304 assimilatory nitrate reduction in TH2746, nitrogenase genes for nitrogen fixation and
305 urease genes in some of the Trout Bog MAGs (**Fig. 6**), probably as adaptions to N-limited
306 conditions in Trout Bog.

307 Although Mendota is a eutrophic lake, N can become temporarily limiting during the
308 high-biomass period when N is consumed by large amounts of phytoplankton and
309 bacterioplankton (47). For some bacteria, when N is temporarily limited while C is in
310 excess, cells convert and store the extra C as biopolymers. For example, the
311 verrucomicrobial methanotroph *M. fumariolicum* SolV accumulated a large amount of
312 glycogen (up to 36% of the total dry weight of cells) when the culture was N-limited (48).
313 Similar to this verrucomicrobial methanotroph, genes in glycogen biosynthesis are present
314 in most MAGs from Mendota and Trout Bog (**Fig. 6**). Indeed, a glycogen synthesis pathway
315 is also present in most genomes of cultivated Verrucomicrobia in the public database (data
316 not shown), suggesting that glycogen accumulation might be a common feature for this
317 phylum to cope with the changing pools of C and N in the environment and facilitate their
318 survival when either is temporally limited.

319

320 **Phosphorus (P) metabolism and other metabolic features**

321 Verrucomicrobia populations represented by these MAGs may be able to survive under low
322 P conditions, as suggested by the presence of genes responding to P limitation, such as the
323 two-component regulator (*phoRB*), alkaline phosphatase (*phoA*), phosphonoacetate
324 hydrolase (*phnA*), and high-affinity phosphate-specific transporter system (*pstABC*) (**Fig.**
325 **6**). Detailed discussion in P acquisition and metabolism and other metabolic aspects, such

326 as acetate metabolism, sulfur metabolism, oxygen tolerance, and the presence of the
327 alternative complex III and cytochrome *c* oxidase genes in the oxidative phosphorylation
328 pathway, are discussed in the Supplementary Text (**Fig. S6**).
329

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

331 Respiration using alternative electron acceptors is important for overall lake metabolism in
332 the DOC-rich humic Trout Bog, as the oxygen levels decrease quickly with depth in the
333 water column. We therefore searched for genes involved in anaerobic respiration, and
334 found that genes in the dissimilatory reduction of nitrate, nitrite, sulfate, sulfite, DMSO, and
335 TMAO are largely absent in all MAGs (**Supplementary Text, Fig. S6)**. Compared to those
336 anaerobic processes, genes for dissimilatory metal reduction are less well understood. In
337 more extensively studied cultured iron [Fe(III)] reducers, outer surface *c*-type cytochromes
338 (cyt*c*), such as OmcE and OmcS in *Geobacter sulfurreducens* are involved in Fe(III)
339 reduction at the cell outer surface (49). Further, a periplasmic multiheme cytochrome *c*
340 (MHC, e.g. MtrA in *Shewanella oneidensis* and OmaB/OmaC in *G. sulfurreducens*) can be
341 embedded into a porin (e.g. MtrB in *S. oneidensis* and OmbB/OmbC in *G. sulfurreducens*),
342 forming a porin-MHC complex as an extracellular electron transfer (EET) conduit to reduce
343 extracellular Fe(III) (50, 51). Such outer surface cyt*c* and porin-MHC systems involved in
344 Fe(III) reduction were also suggested to be important in reducing the quinone groups in
345 humic substances (HS) at the cell surface (52-54). The reduced HS can be re-oxidized by
346 Fe(III) or oxygen, thus HS can serve as electron shuttles to facilitate Fe(III) reduction (55,
347 56) or as regenerable electron acceptors at the anoxic-oxic interface or over redox cycles
348 (57).

349 Outer surface cyt*c* or porin-MHC systems homologous to the ones in *G.*
350 *sulfurreducens* and *S. oneidensis* are not present in Verrucomicrobia MAGs. Instead, we
351 identified a novel porin-coding gene clustering with MHC genes in six MAGs (**Fig. 7**). These
352 porins were predicted to have at least 20 transmembrane motifs, and their adjacent cyt*c*
353 were predicted to be periplasmic proteins with eight conserved heme-binding sites. In
354 several cases, a gene encoding an extracellular MHC is also located in the same gene cluster.
355 As their gene organization is analogous to the porin-MHC gene clusters in *G. sulfurreducens*
356 and *S. oneidensis*, we hypothesize that these genes in Verrucomicrobia may encode a novel
357 porin-MHC complex involved in EET.

358 As these porin-MHC gene clusters are novel, we further confirmed that they are
359 indeed from Verrucomicrobia. Their containing contigs were indeed classified to
360 Verrucomicrobia based on the consensus of the best BLASTP hits for genes on these
361 contigs. Notably, the porin-MHC gene cluster was only observed in MAGs recovered from
362 the HS-rich Trout Bog, especially from the anoxic hypolimnion environment. Searching the
363 NCBI and IMG databases for the porin-MHC gene clusters homologous to those in Trout
364 Bog, we identified homologs in genomes within the Verrucomicrobia phylum, including
365 *Opitutus terrae* PB90-1 isolated from rice paddy soil, *Opitutus* sp. GAS368 isolated from
366 forest soil, "*Candidatus* Udaeobacter copiosus" recovered from prairie soil, Opititae-40 and
367 Opititae-129 recovered from freshwater sediment, and Verrucomicrobia bacterium
368 IMCC26134 recovered from freshwater; some of their residing environments are also rich
369 in HS. Therefore, based on the occurrence pattern of porin-MHC among Verrucomicrobia
370 genomes, we hypothesize that such porin-MHCs might participate in EET to HS in anoxic
371 HS-rich environments, and HS may further shuttle electrons to poorly soluble metal oxides

372 or be regenerated at the anoxic-oxic interface, thereby diverting more C flux to respiration
373 instead of fermentation and methanogenesis, which could impact the overall energy
374 metabolism and green-house gas emission in the bog environment.
375
376 **Occurrence of Planctomycete-specific cytochrome *c* and domains**
377 One of the interesting features of Verrucomicrobia and its sister phyla in the PVC
378 superphylum is the presence of a number of novel protein domains in some of their
379 member genomes (58, 59). These domains were initially identified in marine
380 planctomycete *Rhodopirellula baltica* (58) and therefore, were referred to as
381 "Planctomycete-specific", although some of them were later identified in other PVC
382 members (59). In our Verrucomicrobia MAGs, most genes containing Planctomycete-
383 specific cytochrome *c* domains (PSCyt1 to PSCyt3) also contain other Planctomycete-
384 specific domains (PSD1 through PSD5) with various combinations and arrangements (**Fig.**
385 **8** and **S7a**). Further, PSCyt2-containing and PSCyt3-containing genes are usually next to
386 two different families of unknown genes, respectively (**Fig. S7b**). Such conserved domain
387 architectures and gene organizations, as well as their high occurrence frequencies in some
388 of the Verrucomicrobia MAGs are intriguing, yet nothing is known about their functions.
389 However, some of the PSCyt-containing genes also contain protein domains identifiable as
390 carbohydrate-binding modules (CBMs), suggesting a role in carbohydrate metabolism (see
391 detailed discussion in **Supplementary Text**).
392 The coding density of PSCyt-containing genes indicates that they tend to be more
393 abundant in the epilimnion (either ME or TE) genomes (**Fig. 2c**) and exhibit an inverse
394 correlation with the GH coding density (r = -0.62). Interestingly, sulfatase-coding genes are

395 often in the neighborhood of PSCyt-containing genes in ME and TE genomes, whereas
396 sulfatase-coding genes often neighbor with GH genes in TH genomes. The genomic context
397 suggests PSCyt-containing gene functions somewhat mirror those of GHs (although their
398 reaction mechanisms likely differ fundamentally). However, these PSCyt-containing genes
399 were predicted to be periplasmic or cytoplasmic proteins rather than extracellular or outer
400 membrane proteins. Hence, if they are indeed involved in carbohydrate degradation, they
401 likely act on mono- or oligomers that can be transported into the cell. Further, the
402 distribution patterns of GH versus PSCyt-containing genes between the epilimnion and
403 hypolimnion may reflect the difference in oxygen availability and their carbohydrate
404 substrate complexity between the two layers, suggesting some niche differentiation within
405 Verrucomicrobia in freshwater systems. Therefore, we suggest that a combination of
406 carbohydrate composition, electron acceptor availability and C accessibility drive gene
407 distributions in these populations.
408
409 **Summary**
410 Verrucomicrobia MAGs recovered from the two contrasting lakes greatly expanded the
411 known genomic diversity of freshwater Verrucomicrobia, revealed the ecophysiology and
412 some interesting adaptive features of this ubiquitous yet less understood freshwater
413 lineage. The overrepresentation of GH, sulfatase, and carbohydrate transporter genes, the
414 genetic potential to use various sugars, and the microcompartments for fucose and
415 rhamnose degradation suggest that they are potentially (poly)saccharide degraders in
416 freshwater. Most of the MAGs encode machineries to cope with the changing availability of
417 N and P and can survive nutrient limitation. Despite these generalities, these

418 Verrucomicrobia differ significantly between lakes in the abundance and functional profiles
419 of their GH genes, which may reflect different C sources of the two lakes. Interestingly, a
420 number of MAGs in Trout Bog possess gene clusters potentially encoding a novel porin-
421 multiheme cytochrome *c* complex, and might be involved in extracellular electron transfer
422 in the anoxic humic-rich environment. Intriguingly, large numbers of Planctomycete-
423 specific cytochrome *c*-containing genes are present in MAGs from the epilimnion,
424 exhibiting nearly opposite distribution patterns with GH genes. Future studies are needed
425 to elucidate the functions of these novel and fascinating genomic features.

426 In this study, we focused on using genome information to infer ecophysiology of
427 Verrucomicrobia. The rich time-series metagenome dataset and the many diverse
428 microbial genomes recovered in these two lakes also provide an opportunity for the future
429 study of Verrucomicrobia population dynamics in the context of the total community and
430 their interactions with environmental variables and other microbial groups.

431 As some of the MAGs analyzed here represent first genome representatives of
432 several Verrucomicrobia subdivisions from freshwater, an interesting question is whether
433 populations represented by the MAGs are native aquatic residents and active in aquatic
434 environment, or merely present after having been washed into the lake from surrounding
435 soil. Previous studies on freshwater Verrucomicrobia were largely based on 16S rRNA
436 genes, yet 16S rRNA genes were not recovered in most MAGs, making it difficult to directly
437 link our MAGs to previously identified freshwater Verrucomicrobia. Notably, our MAGs
438 were only distantly related to the ubiquitous and abundant soil Verrucomicrobia,
439 "*Candidatus* Udaeobacter copiosus" (10) (**Fig. 1**). In addition, Verrucomicrobia were
440 abundant in Trout Bog and other bogs from a five-year bog lake bacterial community

441 composition and dynamics study (60), with average relative abundance of 7.1% and 8.6%,
442 and maximal relative abundance of 25.4% and 39.5% in Trout Bog epilimnion and
443 hypolimnion respectively. Since the MAGs were presumably from the most abundant
444 Verrucomicrobia populations, they were not likely soil immigrants due to their high
445 abundance in the aquatic environment. To confirm their aquatic origin, future experiments
446 should be designed to test their activities and physiology in the aquatic environment based
447 on the genomic insights gained in this study.
448
449
450 **MATERIALS AND METHODS**
451 **Study sites.** Samples for metagenome sequencing were collected from two temperate lakes
452 in Wisconsin, USA, Lake Mendota and Trout Bog Lake, during ice-off periods of each year
453 (May to November). Mendota is an urban eutrophic lake with most of its C being
454 autochthonous (in-lake produced), whereas Trout Bog is a small, acidic and nutrient-poor
455 dystrophic lake with mostly terrestrially-derived (allochthonous) C. General lake
456 characteristics are summarized in **Table 1**.
457
458 **Sampling.** For Mendota, we collected depth-integrated water samples from the surface 12
459 m (mostly consisting of the epilimnion layer) at 94 time points from 2008 to 2012, and
460 samples were referred to as "ME" (38). For Trout Bog, we collected the integrated
461 hypolimnion layer at 45 time points from 2007 to 2009 and the integrated epilimnion layer
462 at 45 time points from 2007 to 2009, and samples were referred to as "TH" and "TE",
463 respectively (37). All samples were filtered through 0.22 μm polyethersulfone filters and

464 stored at -80°C until extraction. DNA was extracted from the filters using the FastDNA kit
465 (MP Biomedicals) according to manufacturer's instruction with some minor modifications
466 as described previously (34).
467
468 **Metagenome sequencing, assembly, and draft genome recovery.** Details of
469 metagenome sequencing, assembly, and binning were described in Bendall *et al.* (37) and
470 Hamilton et al (61). Briefly, shotgun Illumina HiSeq 2500 metagenome libraries were
471 constructed for each of the DNA samples. Three combined assemblies were generated by
472 co-assembling reads from all metagenomes within the ME, TE, and TH groups, respectively.
473 Binning was conducted on the three combined assemblies to recover "metagenome-
474 assembled genomes" (MAGs) based on the combination of contig tetranucleotide frequency
475 and differential coverage patterns across time points using MetaBAT (62). Subsequent
476 manual curation of MAGs was conducted to remove contigs that did not correlate well with
477 the median temporal abundance pattern of all contigs within a MAG, as described in
478 Bendall *et al.* (37).
479
480 **Genome annotation and completeness estimation.** MAGs were submitted to the DOE
481 Joint Genome Institute's Integrated Microbial Genome (IMG) database for gene prediction
482 and function annotation (63). The IMG Taxon Object IDs for Verrucomicrobia MAGs are
483 listed in **Table 2**. The completeness and contamination of each MAG was estimated using
484 checkM with both the lineage-specific and Verrucomicrobia-specific workflows (35). The
485 Verrucomicrobia-specific workflow provided more accurate estimates (i.e. higher genome
486 completeness and lower contamination) than the lineage-specific workflow when tested on

487 11 complete genomes of Verrucomicrobia isolates available at IMG during our method
488 validation. We therefore only reported the estimates from Verrucomicrobia-specific
489 workflow (**Table 2**). MAGs with an estimated completeness lower than 50% were not
490 included in this study.
491
492 **Taxonomic and phylogenetic analysis.** A total of 19 MAGs were classified to the
493 Verrucomicrobia phylum based on taxonomic assignment by PhyloSift using 37 conserved
494 phylogenetic marker genes (64), as described in Bendall *et al.* (37). A phylogenetic tree was
495 reconstructed from the 19 Verrucomicrobia MAGs and 24 reference genomes using an
496 alignment concatenated from individual protein alignments of five conserved essential
497 single-copy genes (represented by TIGR01391, TIGR01011, TIGR00663, TIGR00460, and
498 TIGR00362) that were recovered in all Verrucomicrobia MAGs. Individual alignments were
499 first generated with MUSCLE (65), concatenated, and trimmed to exclude columns that
500 contain gaps for more than 30% of all sequences. A maximum likelihood phylogenetic tree
501 was constructed using PhyML 3.0 (66), with the LG substitution model and the gamma
502 distribution parameter estimated by PhyML. Bootstrap values were calculated based on
503 100 replicates. *Kiritimatiella glycovorans* L21-Fru-AB was used as an outgroup in the
504 phylogenetic tree. This bacterium was initially designated as the first (and so far the only)
505 cultured representative of Verrucomicrobia subdivision 5. However, this subdivision was
506 later proposed as a novel sister phylum associated with Verrucomicrobia (67), making it an
507 ideal outgroup for this analysis.
508

509 **Estimate of metabolic potential.** IMG provides functional annotation based on KO (KEGG
510 orthology) term, COG (cluster of orthologous group), pfam, and TIGRfam. To estimate
511 metabolic potential, we primarily used KO terms due to their direct link to KEGG pathways.
512 COG, pfam, and TIGRfam were also used when KO terms were not available for a function.
513 Pathways are primarily reconstructed according to KEGG modules, and MetaCyc pathway is
514 used if a KEGG module is not available for a pathway. As these MAGs are incomplete
515 genomes, a fraction of genes in a pathway may be missing due to genome incompleteness.
516 Therefore, we estimated the completeness of a pathway as the fraction of recovered
517 enzymes in that pathway (e.g. a pathway is 100% complete if all enzymes in that pathway
518 are encoded by genes recovered in a MAG). As some genes are shared by multiple
519 pathways, signature genes specific for a pathway were used to indicate the presence of a
520 pathway. If signature genes for a pathway were missing in all MAGs, that pathway was
521 likely absent in all genomes. Based on this, we established criteria for estimating pathway
522 completeness in each MAG. If a signature gene in a pathway was present, we report the
523 percentage of genes in the pathway that we found. If a signature gene was absent in a MAG,
524 but present in at least one third of all MAGs (i.e. >=7), we still report the pathway
525 completeness for that MAG in order to account for genome incompleteness. Otherwise, we
526 considered the pathway to be absent (i.e. completeness is 0%).
527
528 **Glycoside hydrolase identification.** Glycoside hydrolase (GH) genes were identified using
529 the dbCAN annotation tool (http://csbl.bmb.uga.edu/dbCAN/annotate.php) (68) using
530 HMMER search against hidden Markov models (HMMs) built for all GHs, with an E-value
531 cutoff of 1e-7, except GH109, for which we found that the HMM used by dbCAN is

532 pfam01408, which is a small domain at the N-terminus of GH109 proteins, but is not
533 specific for GH109. Therefore, to identify verrucomicrobial GH109, BLASTP was performed
534 using the two GH109 sequences (GenBank accession ACD03864 and ACD04752) from
535 verrucomicrobial *Akkermansia muciniphila* ATCC BAA-835 listed in the CAZy database
536 (http://www.cazy.org), with E-value cutoff of 1e-6 and query sequence coverage cutoff of
537 50%.
538
539 **Other bioinformatic analyses.** Protein cellular location was predicted using CELLO v.2.5
540 (http://cello.life.nctu.edu.tw) (69) and PSORTb v.3.0 (http://www.psort.org/psortb) (70).
541 The beta-barrel structure of outer membrane proteins was predicted by PRED-TMBB
542 (http://bioinformatics.biol.uoa.gr//PRED-TMBB) (71).
543
544 **CONFLICT OF INTEREST**
545 The authors declare no conflict of interest.
546
547 **ACKNOWLEDGEMENTS**
548 We thank the North Temperate Lakes Microbial Observatory 2007-2012 field crews, UW-
549 Trout Lake Station, the UW Center for Limnology, and the Global Lakes Ecological
550 Observatory Network for field and logistical support. We give special thanks to past
551 McMahon lab graduate students Ashley Shade, Ryan Newton, Emily Read, and Lucas
552 Beversdorf. We acknowledge efforts by many McMahon Lab undergrads and technicians
553 related to sample collection and DNA extraction, particularly Georgia Wolfe. 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.

The project was supported by 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). The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

FUNDING INFORMATION

U.S. NSF Microbial Observatories program (MCB-0702395)
Katherine D McMahon

U.S. NSF Long Term Ecological Research program (NTL-LTER DEB-1440297)
Katherine D McMahon

U.S. NSF INSPIRE award (DEB-1344254)
Katherine D McMahon

U.S. National Institute of Food and Agriculture, U.S. Department of Agriculture (Hatch Project 1002996)
Katherine D McMahon

581

582 **REFERENCE**

583 1. **Zwart G, van Hannen EJ, Kamst-van Agterveld MP, Van der Gucht K, Lindström**
584 **ES, Van Wichelen J, Lauridsen T, Crump BC, Han S-K, Declerck S**. 2003. Rapid
585 Screening for Freshwater Bacterial Groups by Using Reverse Line Blot
586 Hybridization. Applied and Environmental Microbiology **69**:5875-5883.
587 2. **Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S**. 2011. A Guide to the
588 Natural History of Freshwater Lake Bacteria. Microbiology and Molecular Biology
589 Reviews : MMBR **75**:14-49.
590 3. **Eiler A, Bertilsson S**. 2004. Composition of freshwater bacterial communities
591 associated with cyanobacterial blooms in four Swedish lakes. Environ Microbiol
592 **6**:1228-43.
593 4. **Parveen B, Mary I, Vellet A, Ravet V, Debroas D**. 2013. Temporal dynamics and
594 phylogenetic diversity of free-living and particle-associated Verrucomicrobia
595 communities in relation to environmental variables in a mesotrophic lake. FEMS
596 Microbiol Ecol **83**:189-201.
597 5. **Arnds J, Knittel K, Buck U, Winkel M, Amann R**. 2010. Development of a 16S
598 rRNA-targeted probe set for Verrucomicrobia and its application for fluorescence in
599 situ hybridization in a humic lake. Syst Appl Microbiol **33**:139-48.
600 6. **Hedlund BP, Gosink JJ, Staley JT**. 1997. Verrucomicrobia div. nov., a new division
601 of the bacteria containing three new species of Prosthecobacter. Antonie van
602 Leeuwenhoek **72**:29-38.
603 7. **Yoon J, Yasumoto-Hirose M, Katsuta A, Sekiguchi H, Matsuda S, Kasai H, Yokota**
604 **A**. 2007. Coraliomargarita akajimensis gen. nov., sp. nov., a novel member of the
605 phylum 'Verrucomicrobia' isolated from seawater in Japan. Int J Syst Evol Microbiol
606 **57**:959-63.
607 8. **Yoon J, Matsuo Y, Katsuta A, Jang JH, Matsuda S, Adachi K, Kasai H, Yokota A**.
608 2008. Haloferula rosea gen. nov., sp. nov., Haloferula harenae sp. nov., Haloferula
609 phyci sp. nov., Haloferula helveola sp. nov. and Haloferula sargassicola sp. nov., five
610 marine representatives of the family Verrucomicrobiaceae within the phylum
611 'Verrucomicrobia'. Int J Syst Evol Microbiol **58**:2491-500.
612 9. **Sangwan P, Kovac S, Davis KE, Sait M, Janssen PH**. 2005. Detection and cultivation
613 of soil verrucomicrobia. Appl Environ Microbiol **71**:8402-10.
614 10. **Brewer TE, Handley KM, Carini P, Gilbert JA, Fierer N**. 2016. Genome reduction
615 in an abundant and ubiquitous soil bacterium 'Candidatus Udaeobacter copiosus'.
616 **2**:16198.
617 11. **Qiu YL, Kuang XZ, Shi XS, Yuan XZ, Guo RB**. 2014. Terrimicrobium sacchariphilum
618 gen. nov., sp. nov., an anaerobic bacterium of the class 'Spartobacteria' in the phylum
619 Verrucomicrobia, isolated from a rice paddy field. Int J Syst Evol Microbiol **64**:1718-
620 23.
621 12. **da Rocha UN, van Elsas JD, van Overbeek LS**. 2010. Real-time PCR detection of
622 Holophagae (Acidobacteria) and Verrucomicrobia subdivision 1 groups in bulk and
623 leek (Allium porrum) rhizosphere soils. Journal of Microbiological Methods **83**:141-
624 148.

625 13. **Wertz JT, Kim E, Breznak JA, Schmidt TM, Rodrigues JLM**. 2012. Genomic and
626 Physiological Characterization of the Verrucomicrobia Isolate Diplosphaera
627 colitermitum gen. nov., sp. nov., Reveals Microaerophily and Nitrogen Fixation
628 Genes. Applied and Environmental Microbiology **78**:1544-1555.
629 14. **Derrien M, Vaughan EE, Plugge CM, de Vos WM**. 2004. Akkermansia muciniphila
630 gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. Int J Syst Evol
631 Microbiol **54**:1469-76.
632 15. **Scheuermayer M, Gulder TA, Bringmann G, Hentschel U**. 2006. Rubritalea
633 marina gen. nov., sp. nov., a marine representative of the phylum 'Verrucomicrobia',
634 isolated from a sponge (Porifera). Int J Syst Evol Microbiol **56**:2119-24.
635 16. **Chin KJ, Liesack W, Janssen PH**. 2001. Opitutus terrae gen. nov., sp. nov., to
636 accommodate novel strains of the division 'Verrucomicrobia' isolated from rice
637 paddy soil. Int J Syst Evol Microbiol **51**:1965-8.
638 17. **Otsuka S, Suenaga T, Vu HT, Ueda H, Yokota A, Senoo K**. 2013. Brevifollis
639 gellanilyticus gen. nov., sp. nov., a gellan-gum-degrading bacterium of the phylum
640 Verrucomicrobia. Int J Syst Evol Microbiol **63**:3075-8.
641 18. **Sangwan P, Chen X, Hugenholtz P, Janssen PH**. 2004. Chthoniobacter flavus gen.
642 nov., sp. nov., the first pure-culture representative of subdivision two,
643 Spartobacteria classis nov., of the phylum Verrucomicrobia. Appl Environ Microbiol
644 **70**:5875-81.
645 19. **Hedlund BP, Gosink JJ, Staley JT**. 1996. Phylogeny of Prosthecobacter, the fusiform
646 caulobacters: members of a recently discovered division of the bacteria. Int J Syst
647 Bacteriol **46**:960-6.
648 20. **Otsuka S, Ueda H, Suenaga T, Uchino Y, Hamada M, Yokota A, Senoo K**. 2013.
649 Roseimicrobium gellanilyticum gen. nov., sp. nov., a new member of the class
650 Verrucomicrobiae. Int J Syst Evol Microbiol **63**:1982-6.
651 21. **Pol A, Heijmans K, Harhangi HR, Tedesco D, Jetten MS, Op den Camp HJ**. 2007.
652 Methanotrophy below pH 1 by a new Verrucomicrobia species. Nature **450**:874-8.
653 22. **Freitas S, Hatosy S, Fuhrman JA, Huse SM, Welch DB, Sogin ML, Martiny AC**.
654 2012. Global distribution and diversity of marine Verrucomicrobia. ISME J **6**:1499-
655 505.
656 23. **Cardman Z, Arnosti C, Durbin A, Ziervogel K, Cox C, Steen AD, Teske A**. 2014.
657 Verrucomicrobia are candidates for polysaccharide-degrading bacterioplankton in
658 an arctic fjord of Svalbard. Appl Environ Microbiol **80**:3749-56.
659 24. **Martinez-Garcia M, Brazel DM, Swan BK, Arnosti C, Chain PS, Reitenga KG, Xie
660 G, Poulton NJ, Lluesma Gomez M, Masland DE, Thompson B, Bellows WK,
661 Ziervogel K, Lo CC, Ahmed S, Gleasner CD, Detter CJ, Stepanauskas R**. 2012.
662 Capturing single cell genomes of active polysaccharide degraders: an unexpected
663 contribution of Verrucomicrobia. PLoS One **7**:e35314.
664 25. **Herlemann DP, Lundin D, Labrenz M, Jurgens K, Zheng Z, Aspeborg H,
665 Andersson AF**. 2013. Metagenomic de novo assembly of an aquatic representative
666 of the verrucomicrobial class Spartobacteria. MBio **4**:e00569-12.
667 26. **Paver SF, Kent AD**. 2010. Temporal patterns in glycolate-utilizing bacterial
668 community composition correlate with phytoplankton population dynamics in
669 humic lakes. Microb Ecol **60**:406-18.

670 27. **Haukka K, Kolmonen E, Hyder R, Hietala J, Vakkilainen K, Kairesalo T, Haario**
671 **H, Sivonen K**. 2006. Effect of nutrient loading on bacterioplankton community
672 composition in lake mesocosms. Microb Ecol **51**:137-46.
673 28. **Lindström ES, Vrede K, Leskinen E**. 2004. Response of a member of the
674 Verrucomicrobia, among the dominating bacteria in a hypolimnion, to increased
675 phosphorus availability. Journal of Plankton Research **26**:241-246.
676 29. **Kolmonen E, Sivonen K, Rapala J, Haukka K**. 2004. Diversity of cyanobacteria and
677 heterotrophic bacteria in cyanobacterial blooms in Lake Joutikas, Finland. Aquatic
678 Microbial Ecology **36**:201-211.
679 30. **Lindström ES, Kamst-Van Agterveld MP, Zwart G**. 2005. Distribution of Typical
680 Freshwater Bacterial Groups Is Associated with pH, Temperature, and Lake Water
681 Retention Time. Applied and Environmental Microbiology **71**:8201-8206.
682 31. **Schlesner H**. 1987. Verrucomicrobium spinosum gen. nov., sp. nov.: a fimbriated
683 prosthecate bacterium. Systematic and applied microbiology **10**:54-56.
684 32. **Zwart G, Huismans R, van Agterveld MP, Van de Peer Y, De Rijk P, Eenhoorn H,**
685 **Muyzer G, van Hannen EJ, Gons HJ, Laanbroek HJ**. 1998. Divergent members of
686 the bacterial division Verrucomicrobiales in a temperate freshwater lake. FEMS
687 Microbiology Ecology **25**:159-169.
688 33. **Read JS, Rose KC**. 2013. Physical responses of small temperate lakes to variation in
689 dissolved organic carbon concentrations. Limnology and Oceanography **58**:921-931.
690 34. **Shade A, Jones SE, McMahon KD**. 2008. The influence of habitat heterogeneity on
691 freshwater bacterial community composition and dynamics. Environmental
692 Microbiology **10**:1057-1067.
693 35. **Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW**. 2015. CheckM:
694 assessing the quality of microbial genomes recovered from isolates, single cells, and
695 metagenomes. Genome Res **25**:1043-55.
696 36. **Schlesner H, Jenkins C, Staley J**. 2006. The Phylum Verrucomicrobia: A
697 Phylogenetically Heterogeneous Bacterial Group, p 881-896. *In* Dworkin M, Falkow
698 S, Rosenberg E, Schleifer K-H, Stackebrandt E (ed), The Prokaryotes doi:10.1007/0-
699 387-30747-8_37. Springer New York.
700 37. **Bendall ML, Stevens SLR, Chan L-K, Malfatti S, Schwientek P, Tremblay J,**
701 **Schackwitz W, Martin J, Pati A, Bushnell B, Froula J, Kang D, Tringe SG,**
702 **Bertilsson S, Moran MA, Shade A, Newton RJ, McMahon KD, Malmstrom RR**.
703 2016. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial
704 populations. ISME J doi:10.1038/ismej.2015.241.
705 38. **Garcia SL, Stevens SLR, Crary B, Martinez-Garcia M, Stepanauskas R, Woyke T,**
706 **Tringe SG, Andersson S, Bertilsson S, Malmstrom RR, McMahon KD**. 2016.
707 Contrasting patterns of genome-level diversity across distinct co-occurring bacterial
708 populations. bioRxiv doi:10.1101/080168.
709 39. **Shieh WY, Jean WD**. 1998. Alterococcus agarolyticus, gen.nov., sp.nov., a halophilic
710 thermophilic bacterium capable of agar degradation. Can J Microbiol **44**:637-45.
711 40. **Kent AD, Jones SE, Yannarell AC, Graham JM, Lauster GH, Kratz TK, Triplett EW**.
712 2004. Annual Patterns in Bacterioplankton Community Variability in a Humic Lake.
713 Microbial Ecology **48**:550-560.
714 41. **Kent AD, Yannarell AC, Rusak JA, Triplett EW, McMahon KD**. 2007. Synchrony in
715 aquatic microbial community dynamics. ISME J **1**:38-47.

716 42. **Erbilgin O, McDonald KL, Kerfeld CA**. 2014. Characterization of a planctomycetal
717 organelle: a novel bacterial microcompartment for the aerobic degradation of plant
718 saccharides. Appl Environ Microbiol **80**:2193-205.
719 43. **Blanvillain S, Meyer D, Boulanger A, Lautier M, Guynet C, Denance N, Vasse J,**
720 **Lauber E, Arlat M**. 2007. Plant carbohydrate scavenging through tonB-dependent
721 receptors: a feature shared by phytopathogenic and aquatic bacteria. PLoS One
722 **2**:e224.
723 44. **Jiao G, Yu G, Zhang J, Ewart HS**. 2011. Chemical structures and bioactivities of
724 sulfated polysaccharides from marine algae. Mar Drugs **9**:196-223.
725 45. **Filali Mouhim R, Cornet J-F, Fontane T, Fournet B, Dubertret G**. 1993.
726 Production, isolation and preliminary characterization of the exopolysaccharide of
727 the cyanobacterium Spirulina platensis. Biotechnology Letters **15**:567-572.
728 46. **Dantas-Santos N, Gomes DL, Costa LS, Cordeiro SL, Costa MS, Trindade ES,**
729 **Franco CR, Scortecci KC, Leite EL, Rocha HA**. 2012. Freshwater plants synthesize
730 sulfated polysaccharides: heterogalactans from Water Hyacinth (Eicchornia
731 crassipes). Int J Mol Sci **13**:961-76.
732 47. **Beversdorf LJ, Miller TR, McMahon KD**. 2013. The role of nitrogen fixation in
733 cyanobacterial bloom toxicity in a temperate, eutrophic lake. PLoS One **8**:e56103.
734 48. **Khadem AF, van Teeseling MC, van Niftrik L, Jetten MS, Op den Camp HJ, Pol A**.
735 2012. Genomic and Physiological Analysis of Carbon Storage in the
736 Verrucomicrobial Methanotroph "Ca. Methylacidiphilum Fumariolicum" SolV. Front
737 Microbiol **3**:345.
738 49. **Mehta T, Coppi MV, Childers SE, Lovley DR**. 2005. Outer membrane c-type
739 cytochromes required for Fe(III) and Mn(IV) oxide reduction in Geobacter
740 sulfurreducens. Appl Environ Microbiol **71**:8634-41.
741 50. **Liu Y, Wang Z, Liu J, Levar C, Edwards MJ, Babauta JT, Kennedy DW, Shi Z,**
742 **Beyenal H, Bond DR, Clarke TA, Butt JN, Richardson DJ, Rosso KM, Zachara JM,**
743 **Fredrickson JK, Shi L**. 2014. A trans-outer membrane porin-cytochrome protein
744 complex for extracellular electron transfer by Geobacter sulfurreducens PCA.
745 Environmental Microbiology Reports **6**:776-785.
746 51. **Shi L, Fredrickson JK, Zachara JM**. 2014. Genomic analyses of bacterial porin-
747 cytochrome gene clusters. Frontiers in Microbiology **5**:657.
748 52. **Bucking C, Piepenbrock A, Kappler A, Gescher J**. 2012. Outer-membrane
749 cytochrome-independent reduction of extracellular electron acceptors in
750 Shewanella oneidensis. Microbiology **158**:2144-57.
751 53. **Shyu JBH, Lies DP, Newman DK**. 2002. Protective Role of tolC in Efflux of the
752 Electron Shuttle Anthraquinone-2,6-Disulfonate. Journal of Bacteriology **184**:1806-
753 1810.
754 54. **Voordeckers JW, Kim BC, Izallalen M, Lovley DR**. 2010. Role of Geobacter
755 sulfurreducens outer surface c-type cytochromes in reduction of soil humic acid and
756 anthraquinone-2,6-disulfonate. Appl Environ Microbiol **76**:2371-5.
757 55. **Lovley DR, Blunt-Harris EL**. 1999. Role of humic-bound iron as an electron
758 transfer agent in dissimilatory Fe(III) reduction. Appl Environ Microbiol **65**:4252-4.
759 56. **Lovley DR, Coates JD, Blunt-Harris EL, Phillips EJP, Woodward JC**. 1996. Humic
760 substances as electron acceptors for microbial respiration. Nature **382**:445-448.

761 57. **Klupfel L, Piepenbrock A, Kappler A, Sander M**. 2014. Humic substances as fully
762 regenerable electron acceptors in recurrently anoxic environments. Nature Geosci
763 **7**:195-200.
764 58. **Studholme DJ, Fuerst JA, Bateman A**. 2004. Novel protein domains and motifs in
765 the marine planctomycete Rhodopirellula baltica. FEMS Microbiol Lett **236**:333-40.
766 59. **Kamneva OK, Knight SJ, Liberles DA, Ward NL**. 2012. Analysis of genome content
767 evolution in pvc bacterial super-phylum: assessment of candidate genes associated
768 with cellular organization and lifestyle. Genome Biol Evol **4**:1375-90.
769 60. **Linz AM, Crary BC, Shade A, Owens S, Gilbert JA, Knight R, McMahon KD**. 2017.
770 Bacterial Community Composition and Dynamics Spanning Five Years in Freshwater
771 Bog Lakes. mSphere **2**.
772 61. **Hamilton JJ, Garcia SL, Brown BS, Oyserman BO, Moya-Flores F, Bertilsson S,**
773 **Malmstrom RR, Forest KT, McMahon KD**. 2007. High-Throughput Metabolic
774 Network Analysis and Metatranscriptomics of a Cosmopolitan and Streamlined
775 Freshwater Lineage. bioRxiv doi:10.1101/106856.
776 62. **Kang DD, Froula J, Egan R, Wang Z**. 2015. MetaBAT, an efficient tool for accurately
777 reconstructing single genomes from complex microbial communities. PeerJ **3**:e1165.
778 63. **Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A,**
779 **Huang J, Woyke T, Huntemann M, Anderson I, Billis K, Varghese N, Mavromatis**
780 **K, Pati A, Ivanova NN, Kyrpides NC**. 2013. IMG 4 version of the integrated
781 microbial genomes comparative analysis system. Nucleic Acids Research
782 doi:10.1093/nar/gkt963.
783 64. **Darling AE, Jospin G, Lowe E, Matsen IV FA, Bik HM, Eisen JA**. 2014. PhyloSift:
784 phylogenetic analysis of genomes and metagenomes. PeerJ **2**:e243.
785 65. **Edgar RC**. 2004. MUSCLE: a multiple sequence alignment method with reduced time
786 and space complexity. BMC Bioinformatics **5**:113-113.
787 66. **Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O**. 2010.
788 New algorithms and methods to estimate maximum-likelihood phylogenies:
789 assessing the performance of PhyML 3.0. Syst Biol **59**:307-21.
790 67. **Spring S, Bunk B, Sproer C, Schumann P, Rohde M, Tindall BJ, Klenk HP**. 2016.
791 Characterization of the first cultured representative of Verrucomicrobia subdivision
792 5 indicates the proposal of a novel phylum. ISME J **10**:2801-2816.
793 68. **Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y**. 2012. dbCAN: a web resource for
794 automated carbohydrate-active enzyme annotation. Nucleic Acids Res **40**:W445-51.
795 69. **Yu CS, Chen YC, Lu CH, Hwang JK**. 2006. Prediction of protein subcellular
796 localization. Proteins **64**:643-51.
797 70. **Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M,**
798 **Foster LJ, Brinkman FSL**. 2010. PSORTb 3.0: improved protein subcellular
799 localization prediction with refined localization subcategories and predictive
800 capabilities for all prokaryotes. Bioinformatics **26**:1608-1615.
801 71. **Bagos PG, Liakopoulos TD, Spyropoulos IC, Hamodrakas SJ**. 2004. PRED-TMBB:
802 a web server for predicting the topology of beta-barrel outer membrane proteins.
803 Nucleic Acids Res **32**:W400-4.
804
805

806

807 **FIGURE LEGENDS**

808 **Fig. 1.** Phylogenetic tree constructed with a concatenated alignment of protein sequences
809 from five conserved essential single-copy genes (represented by TIGR01391, TIGR01011,
810 TIGR00663, TIGR00460, and TIGR00362) that were recovered in all Verrucomicrobia
811 MAGs. ME, TE and TH MAGs are labeled with red, green and blue, respectively. Genome ID
812 in IMG or NCBI is indicated in the bracket. The outgroup is *Kiritimatiella glycovorans* L21-
813 Fru-AB, which was initially assigned to subdivision 5, but this subdivision was recently
814 proposed as a novel sister phylum to Verrucomicrobia (67).

815

816 **Fig. 2. A.** Coding densities of glycoside hydrolase genes, **B.** sulfatase genes, and **C.**
817 Planctomycete-specific cytochrome *c* (PSCyt)-containing genes. Data from ME, TE and TH
818 MAGs are labeled with red, green and blue, respectively. The three plots share the same *x*-
819 axis label as indicated by the genome clustering on the top, which is based on a subtree
820 extracted from the phylogenetic tree in Fig. 1 to indicate the phylogenetic relatedness of
821 the 19 MAGs. The vertical dashed lines divide these MAGs to different subdivisions.

822

823

824 **Fig. 3.** Gene counts for the top 10 most abundant GH families, total gene counts for all GH
825 families, and the number of GH families represented by these genes. MAGs are ordered as
826 in the clustering in Fig. 2.

827

828 **Fig. 4.** Gene clusters encoding bacterial microcompartments (BMCs) involved in L-fucose
829 and L-rhamnose degradation. The vertical line indicates the end of a contig, and IMG gene
830 locus tag for the first gene in each presented gene cluster is indicated in the parentheses.
831 The BMC is schematically represented by a hexagon with the two building blocks labeled in
832 red and green, respectively. The two building blocks and reactions inside the BMC are
833 colored according to their encoding genes' color labels on the left side.
834
835 **Fig. 5.** Gene clusters encoding putative tonB-dependent carbohydrate utilization (CUT) loci.
836 IMG gene locus tag for the first gene in each presented gene cluster is indicated in the
837 parentheses. The horizontal solid lines below genes indicate predicted extracellular or
838 outer membrane proteins.
839
840 **Fig. 6.** Completeness estimates of key metabolic pathways. MAGs are ordered as in the
841 clustering in Fig. 2. Completeness value of "1" indicates a pathway is complete; "0" indicates
842 no genes were found in that pathway; and "(0)" indicates that although some genes in a
843 pathway are present, the pathway is likely absent because signature genes for that pathway
844 were not found in that draft genome AND signature genes are missing in more than two
845 thirds of all draft genomes.
846
847 **Fig. 7.** Gene clusters encoding putative porin-multiheme cytochrome *c* complex (PCC). IMG
848 gene locus tag for the first gene in each presented gene cluster is indicated in the
849 parentheses. The vertical line indicates the end of a contig, and horizontal lines below
850 genes indicate predicted cellular locations of their encoded proteins. These putative PCC

851 genes are in 18.1, 9.0, 6.1, 18.4, 70.0, 10.6 and 10.8 kbp long contigs, respectively. A
852 hypothesized model of extracellular electron transfer is shown on the right with yellow
853 arrows indicating electron flows. "IM" and "OM" refer to inner and outer membranes,
854 respectively, "ET in IM" refers to electron transfer in the inner membrane, and "$EA_{(ox)}$" and
855 "$EA_{(red)}$" refer to oxidized and reduced forms of the electron acceptor, respectively.
856
857 **Fig. 8.** Domain architecture and occurrence of PSCyt-containing genes. Based on the
858 combination of specific PSCyt and PSD domains, these domain structures can be classified
859 into three groups (I, II, and III). "CBM" refers to carbohydrate-binding modules, which
860 include pfam13385 (Laminin_G_3), pfam08531 (Bac_rhamnosid_N), pfam08305 (NPCBM),
861 pfam03422 (CBM_6), and pfam07691 (PA14). "PPI" refers to protein-protein interaction
862 domains, which include pfam02368 (Big_2), pfam00400 (WD40), and pfam00754
863 (F5_F8_type_C).
864
865
866 **LIST OF SUPPLEMENTARY MATERIAL**
867 **Supplementary Text**
868 **Supplementary Figures S1 through S8**
869
870 **Supplementary Figure Legends**
871 **Fig. S1.** A tiled display of an emergent self-organizing map (ESOM) based on the
872 tetranucleotide frequency (TNF) of the 19 Verrucomicrobia MAGs. TNF was calculated with
873 a window size of 5 kbp, with each dot on the ESOM representing a 5-kbp fragment (or a

874 contig if its length is shorter than 5 kbp). Dots (i.e. fragments) are colored according to
875 MAGs. A numeric ID is assigned to each MAG, and IDs from Mendota are labeled in black
876 and IDs from Trout Bog labeled in white. A red outline was drawn to indicate the clustering
877 of MAGs from Mendota on the ESOM.
878
879 **Fig. S2.** Counts of GH genes among the 78 different GH families present in MAGs**.**
880
881 **Fig. S3.** Heat map based on GH abundance profile patterns showing the clustering of MAGs
882 by different lakes.
883
884 **Fig. S4.** Counts of carbohydrate and amino acid transporter genes.
885
886 **Fig. S5.** Comparison of glycolate oxidase gene operons in *E. coli*, *C. flavus* and TE4605.
887
888 **Fig. S6**. Summary of important metabolic genes and pathways.
889
890 **Fig. S7.** Occurrence and gene organization of Planctomycetes-specific domains, DUF1501,
891 and DUF1552. (**a**) Counts of PSCyt, PSD, DUF1501, and DUF1552 domains in the MAGs. (**b**)
892 Clustering of PUF1501- and PSCyt2-containing genes, and clustering of PUF1552- and
893 PSCyt3-containing genes in the genome.
894

895 **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)	5.0	20.0
Total N (mg/L)	1.5	1.3
Total P (µg/L)	131	71
DOC/N	3.3	15.6
DOC/P	38.0	281.9
Trophic state	Eutrophic	Dystrophic

896 [a]Data from NTL-LTER https://lter.limnology.wisc.edu, averaged from the study years. DOC = Dissolved
897 organic carbon. N = Nitrogen. P = Phosphorus.
898
899
900

TABLE 2. Summary of Verrucomicrobia MAGs[a]

Genome	IMG Taxon OID	Subdivision	Recovered MAG Size (Mbp)[b]	Genome Completeness Estimate (%)[c]	Genome Contamination Estimate (%)[c]	GC Content (%)	Coding Base (%)	Gene Count	Normalized Coverage Depth[d]		
									Median	Mean	Coefficient of Variation (%)
ME3880	2582580573	1	1.6	70	2	58	90.9	1585	0.2	2.9	217
TH2746	2582580664	1	6.5	81	3	62	86.7	5430	3.3	4.7	82
ME12612	2582580523	1	2.2	79	3	59	89.0	2335	0.0	0.9	261
ME12173	2582580521	1	2.1	63	3	52	91.3	2070	0.0	0.8	583
TE4605	2582580638	1	4.7	91	0	59	91.1	4380	1.0	4.8	198
ME6381	2582580593	1	2.4	62	0	57	92.5	2221	0.0	0.4	285
ME8366	2582580607	2	3.6	87	5	63	87.4	3450	0.0	1.2	326
TH2747	2582580665	3	5.2	93	8	58	89.6	4846	1.8	2.8	99
TH3004	2582580668	3	4.5	93	6	57	91.4	3798	1.9	5.8	139
TH0989	2556921153	3	7.2	91	8	62	90.3	5583	6.1	6.3	61
TH2519	2593339181	4	1.8	69	2	42	94.3	1654	6.2	6.7	60
TE1800	2593339189	4	2.2	84	2	42	94.3	1998	10.8	11.3	77
TH4590	2582580688	4	3.3	87	1	65	90.7	3132	2.4	3.6	99
ME2014	2582580546	4	1.9	77	5	66	93.7	1700	1.0	3.3	174
ME12657	2582580524	4	1.9	81	7	68	94.0	1838	0.0	0.8	344
TE1301	2582580616	4	2.0	95	0	54	94.7	1943	4.1	14.2	187
TH4093	2582580682	Unclassified	4.7	77	6	48	86.5	3982	4.3	3.9	61
ME30509	2582580559	Unclassified	1.2	51	2	63	92.3	1160	0.0	0.5	479
TH4820	2582580691	Unclassified	3.0	56	3	63	86.7	2794	1.0	2.1	110

[a]MAGs from Lake Mendota are shaded.
[b]Recovered MAG size is the sum of the length of all contigs within a MAG.
[c]Genome completeness and contamination was estimated with checkM using Verrucomicrobia-specific marker gene sets.

[d]Normalized coverage depths of MAGs were calculated from the 94, 45, or 45 individual ME, TE, or TH metagenomes respectively, and were used to comparatively infer relative population abundance at the different sampling points. In addition to the median and mean coverage depths, the coefficient of variation is also shown to indicate variation among the sampling points.

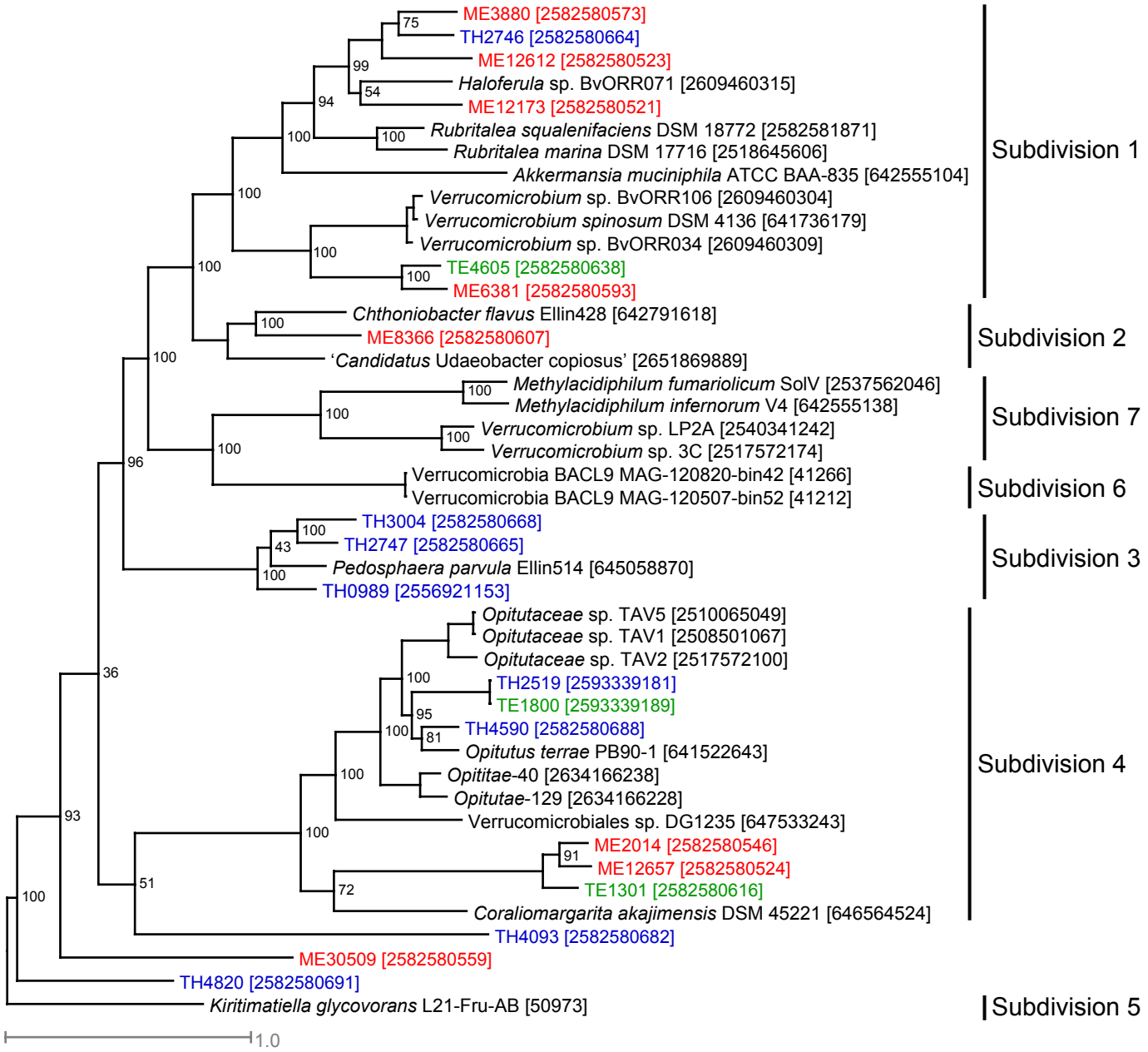

Fig. 1. Phylogenetic tree constructed with a concatenated alignment of protein sequences from five conserved essential single-copy genes (represented by TIGR01391, TIGR01011, TIGR00663, TIGR00460, and TIGR00362) that were recovered in all Verrucomicrobia MAGs. ME, TE and TH MAGs are labeled with red, 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 (67).

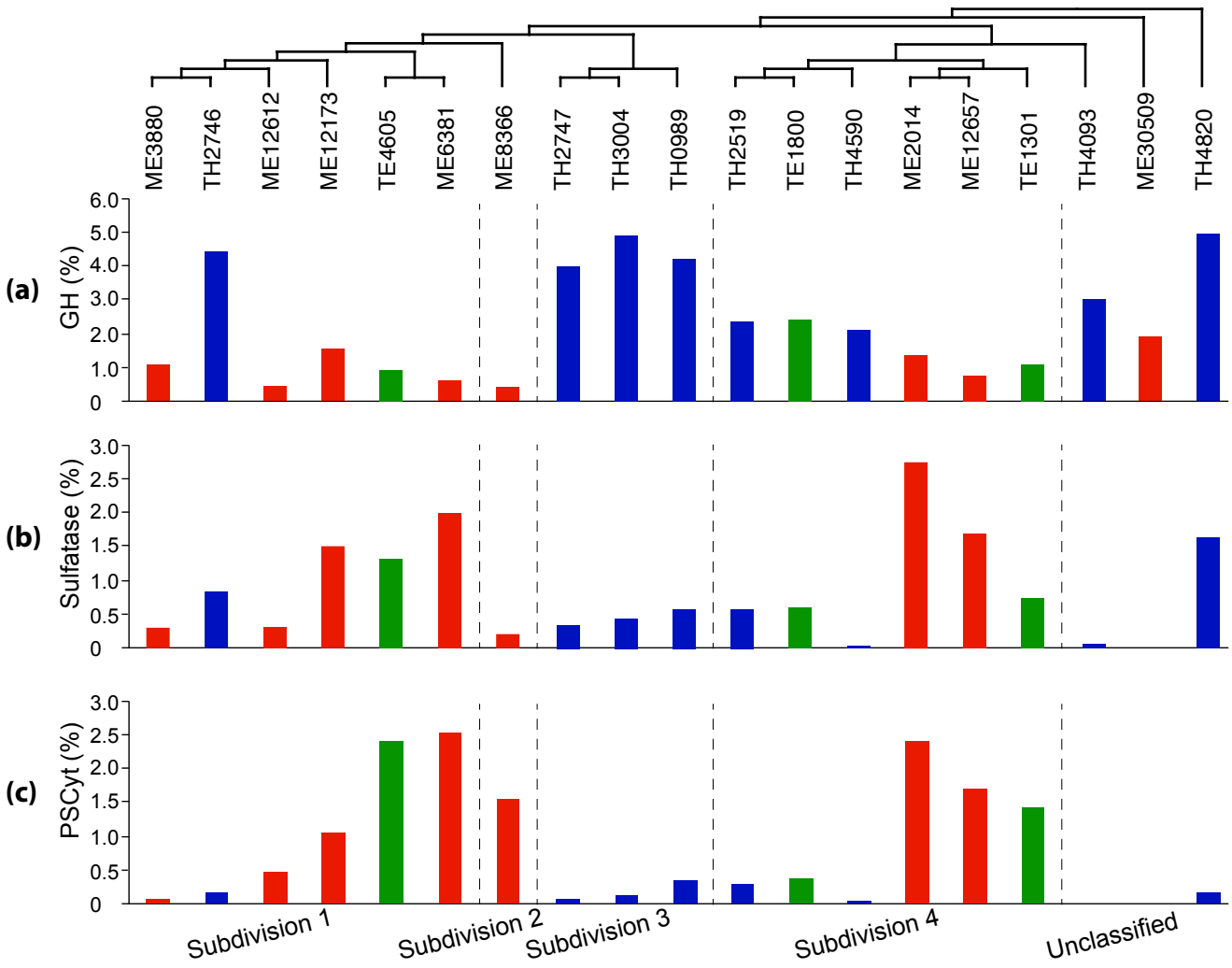

Fig. 2. Coding densities of glycoside hydrolase genes (a), sulfatase genes (b) and Planctomycete-specific cytochrome c (PSCyt)-containing genes (c). Data from ME, TE and TH MAGs are labeled with red, green and blue, respectively. The three plots share the same x-axis label indicated by the genome clustering on the top, which is is based on a subtree extracted from the phylogenetic tree in Fig. 1 to indicate the phylogenetic relatedness of the 19 MAGs. The vertical dashed lines divide these MAGs to different subdivisions.

GH	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**

Fig. 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 Fig. 2.

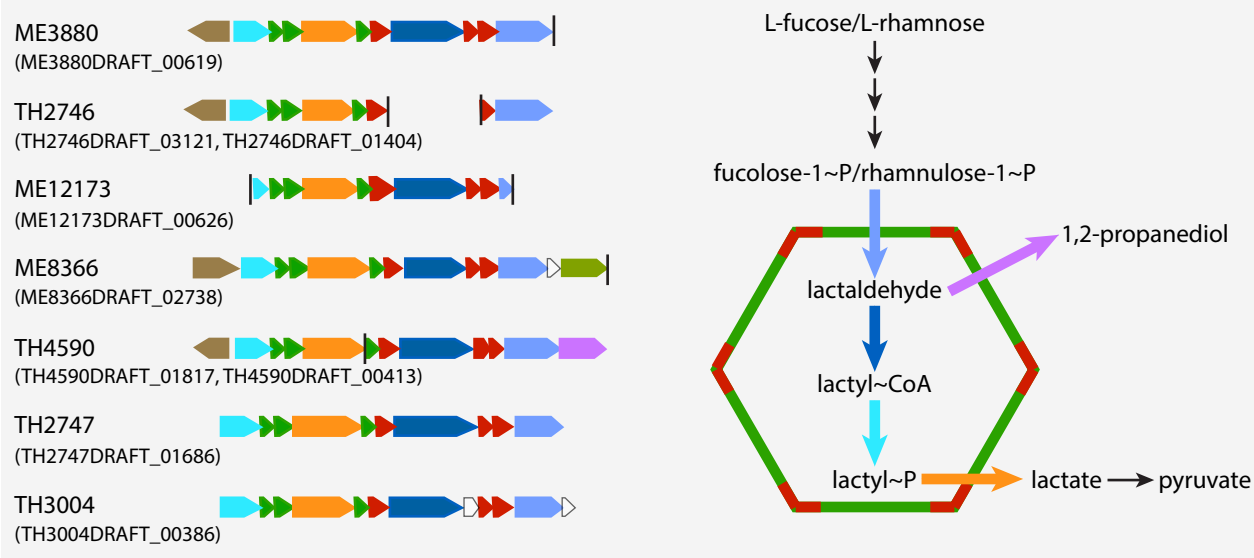

Fig. 4. Gene clusters encoding bacterial microcompartments (BMCs) involved in L-fucose and L-rhamnose degradation. The vertical line indicates the end of a contig, and IMG gene locus tag for the first gene in each presented gene cluster is indicated in the parenthesis. The BMC is schematically represented by a hexagon with the two building blocks labeled in red and green, respectively. The two building blocks and reactions inside the BMC are colored according to their encoding genes' color labels on the left side.

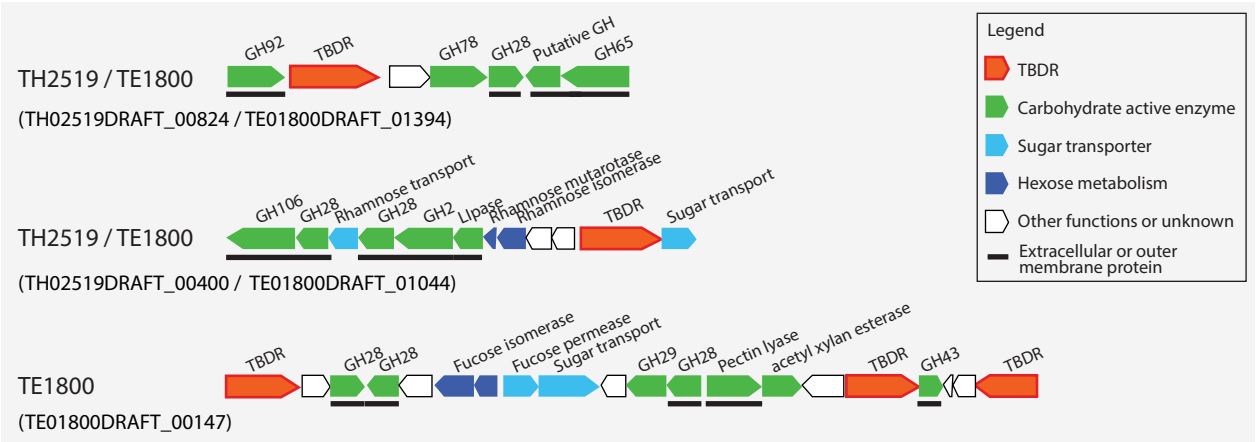

Fig. 5. Gene clusters encoding putative tonB-dependent carbohydrate utilization (CUT) loci. IMG gene locus tag for the first gene in each presented gene cluster is indicated in the parenthesis. The horizontal solid lines below genes indicate predicted extracellular or outer membrane proteins.

Genes and Pathways	ME3880	TH2746	ME12612	ME12173	TE4605	ME6381	ME8366	TH2747	TH3004	TH0989	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

Fig. 6. Completeness estimates of key metabolic pathways. MAGs are ordered as in the clustering in Fig. 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.

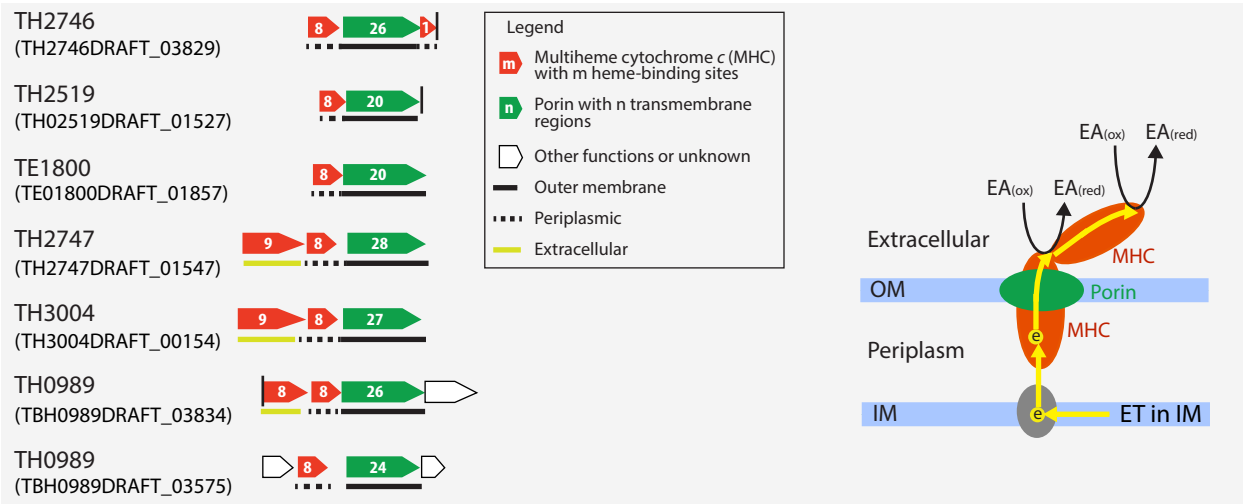

Fig. 7. Gene clusters encoding putative porin-multiheme cytochrome c complex (PCC). IMG gene locus tag for the first gene in each presented gene cluster is indicated in the parenthesis. The vertical line indicates the end of a contig, and horizontal lines below genes indicate predicted cellular locations of their encoded proteins. These putative PCC genes are in 18.1, 9.0, 6.1, 18.4, 70.0, 10.6 and 10.8 kbp long contigs, respectively. A hypothesized model of extracellular electron transfer is shown on the right with yellow arrows indicating electron flows. "IM" and "OM" refer to inner and outer membranes, respectively, "ET in IM" refers to electron transfer in the inner membrane, and "$EA_{(ox)}$" and "$EA_{(red)}$" refer to oxidized and reduced forms of the electron acceptor, respectively.

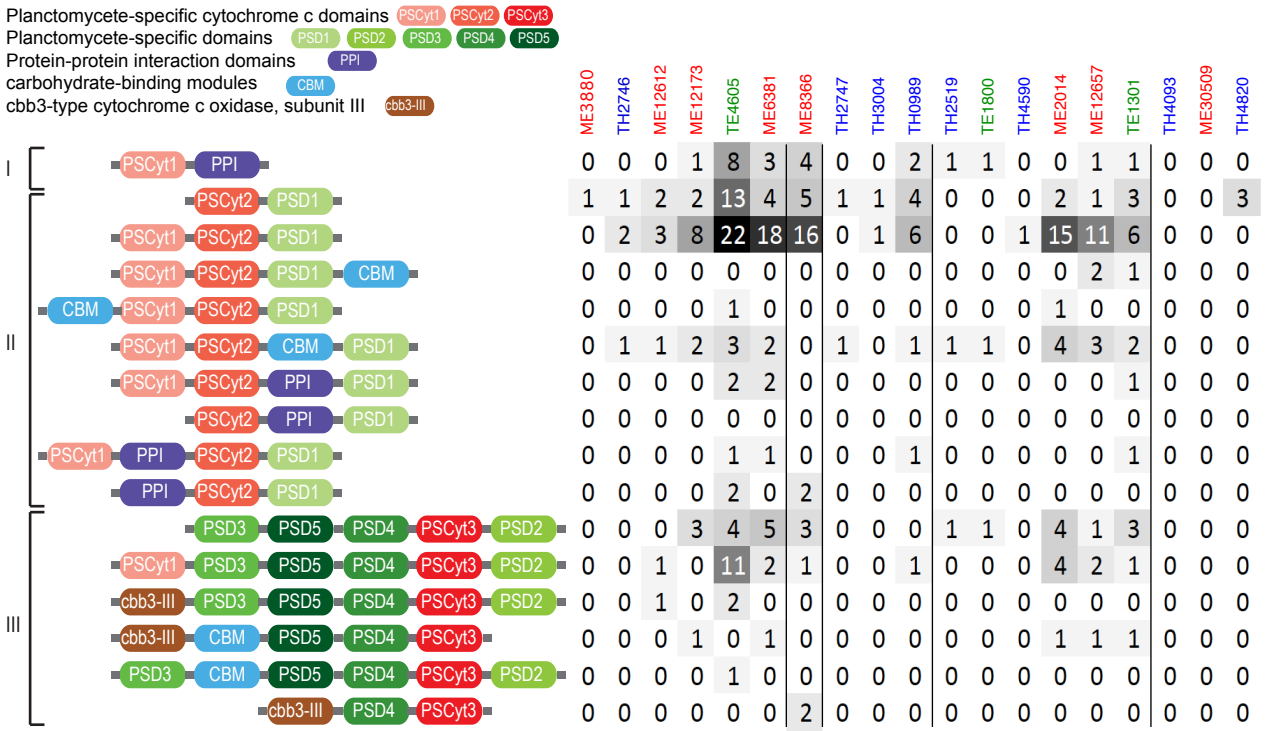

Fig. 8. Domain architecture and occurrence of PSCyt-containing genes. Based on the combination of specific PSCyt and PSD domains, these domain structures can be classified into three groups (I, II, and III). "CBM" refers to carbohydrate-binding modules, which include pfam13385 (Laminin_G_3), pfam08531 (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).