# Metabolic roles of uncultivated bacterioplankton lineages in the northern Gulf of Mexico "Dead Zone"

J. Cameron Thrash[1*], Kiley W. Seitz[2], Brett J. Baker[2*], Ben Temperton[3], Lauren E. Gillies[4], Nancy N. Rabalais[5,6], Bernard Henrissat[7,8,9], and Olivia U. Mason[4]

1. Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, USA
2. Department of Marine Science, Marine Science Institute, University of Texas at Austin, Port Aransas, TX, USA
3. School of Biosciences, University of Exeter, Exeter, UK
4. Department of Earth, Ocean, and Atmospheric Science, Florida State University, Tallahassee, FL, USA
5. Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, LA, USA
6. Louisiana Universities Marine Consortium, Chauvin, LA USA
7. Architecture et Fonction des Macromolécules Biologiques, CNRS, Aix-Marseille Université, 13288 Marseille, France
8. INRA, USC 1408 AFMB, F-13288 Marseille, France
9. Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia

*Correspondence:
JCT thrashc@lsu.edu
BJB acidophile@gmail.com

Running title: Decoding microbes of the Dead Zone

Abstract word count: 250
Text word count: XXXX

**Abstract**

Marine regions that have seasonal to long-term low dissolved oxygen (DO) concentrations, sometimes called 'dead zones,' are increasing in number and severity around the globe with deleterious effects on ecology and economics. One of the largest of these coastal dead zones occurs on the continental shelf of the northern Gulf of Mexico (nGOM), which results from eutrophication-enhanced bacterioplankton respiration and strong seasonal stratification. Previous research in this dead zone revealed the presence of multiple cosmopolitan bacterioplankton lineages that have eluded cultivation, and thus their metabolic roles in this ecosystem remain unknown. We used a coupled shotgun metagenomic and metatranscriptomic approach to determine the metabolic potential of Marine Group II Euryarchaeota, SAR406, and SAR202. We recovered multiple high-quality, nearly complete genomes from all three groups as well as near-complete genomes belonging to Candidate Phyla usually associated with anoxic environments-Parcubacteria (OD1) and Peregrinibacteria. Two additional groups with putative assignments to ACD39 and PAUC34f supplement the metabolic contributions by uncultivated taxa. Our results indicate active metabolism in all groups, including prevalent aerobic respiration, with concurrent expression of genes for nitrate reduction in SAR406 and SAR202, and dissimilatory nitrite reduction to ammonia and sulfur reduction by SAR406. We also report a variety of active heterotrophic carbon processing mechanisms, including degradation of complex carbohydrate compounds by SAR406, SAR202, ACD39, and PAUC34f. Together, these data help constrain the metabolic contributions from uncultivated groups in the nGOM during periods of low DO and suggest roles for these organisms in the breakdown of recalcitrant organic matter.

**Importance**

Dead zones receive their name primarily from the reduction of eukaryotic macrobiota (demersal fish, shrimp, etc.) that are also key coastal fisheries. Excess nutrients contributed from anthropogenic activity such as fertilizer runoff result in algal blooms and therefore ample new carbon for aerobic microbial metabolism. Combined with strong stratification, microbial respiration reduces oxygen in shelf bottom waters to levels unfit for many animals (termed hypoxia). The nGOM shelf remains one of the largest eutrophication-driven hypoxic zones in the world, yet despite its potential as a model study system, the microbial metabolisms underlying and resulting from this phenomenon—many of which occur in bacterioplankton from poorly understood lineages—have received only preliminary study. Our work details the metabolic potential and gene expression activity for uncultivated lineages across several low DO sites in the nGOM, improving our understanding of the active biogeochemical cycling mediated by these "microbial dark matter" taxa during hypoxia.

## Introduction

Hypoxia (dissolved oxygen [DO] below 2 $mg \cdot L^{-1}$/~62.5 $\mu mol \cdot kg^{-1}$) becomes dangerous or lethal to a wide variety of marine life, including organisms of economic importance (1). Hypoxia results from oxygen consumption by aerobic microbial metabolism combined with strong stratification that prevents reoxygeneation of bottom waters. These taxa are fueled primarily by autochthonous organic matter generated from phytoplankton responding to nitrogen input (1). Hypoxic zones have become more widespread globally through the proliferation of nitrogen-based fertilizers and the resulting increases in transport to coastal oceans via runoff (2). In the nGOM, nitrogen runoff from the Mississippi and Atchafalaya Rivers leads to one of the world's largest seasonal regions of coastal hypoxia, with bottom water hypoxia that can extend over 20,000 $km^2$ (1). Action plans to mitigate nGOM hypoxia have stressed that removing continued uncertainties and increasing our "understanding of nutrient cycling and transformations" remain vital for plan implementation (3). These needs motivated our current study of the engines of hypoxic zone nutrient transformation: microorganisms.

Much of our current knowledge regarding microbial contributions to regions of low DO comes from numerous studies investigating naturally occurring, deep-water oxygen minimum zones (OMZs), such as those in the Eastern Tropical North and South Pacific, the Saanich Inlet, and the Arabian, Baltic, and Black Seas (4-11). In many of these systems, continual nutrient supply generates permanent or semi-permanent decreases in oxygen, sometimes to the point of complete anoxia (4). During these conditions, anaerobic metabolisms, such as nitrate and sulfate reduction and anaerobic ammonia oxidation, become prevalent (5, 9, 11-13). In contrast, nGOM hypoxia is distinguished by a seasonal pattern of formation, persistence, and dissolution (1); benthic contributions to bottom water oxygen consumption (14, 15); and a shallow shelf that places much of the water column within the euphotic zone (16). While parts of the nGOM hypoxic zone can become anoxic (1, 17), many areas maintain low oxygen concentrations even during peak hypoxia while the upper water column remains oxygenated (18-20).

The first studies of bacterioplankton assemblages during nGOM hypoxia showed nitrifying Thaumarchaea dominated (21) and could be highly active (22), suggesting a major role for these taxa in nGOM nitrogen cycling. However, many more poorly understood organisms from cosmopolitan, but still uncultivated lineages, such as Marine Group II Euryarchaeota (MGII), SAR406, and SAR202, also occurred in abundance (21, 22). While the likely functions of some of these groups have become clearer recently, all of them contain multiple sublineages that may have distinct metabolic roles. For example, the SAR202 lineage of Chloroflexi contains at least five subclades with distinct ecological profiles (23, 24), and the best understood examples have been examined in the context of complex carbon degradation in the deep ocean (24). Likewise, SAR406 represents a distinct phylum with numerous sublineages, and the bulk of metabolic inference comes from taxa in deep water OMZs (25-28).

None of these groups have been studied in detail in the shallow coastal waters, particularly in the context of seasonal hypoxia. Thus, we pursued a combined metagenomic/metatranscriptomic approach to i) elucidate the specific contributions of these

111 uncultivated lineages to biogeochemical cycling in the nGOM during hypoxia, ii) evaluate the
112 relative similarity of these organisms to their counterparts elsewhere, and iii) determine if other
113 uncultivated lineages had eluded previous microbial characterization in the region due to primer
114 bias (29),16S rRNA gene introns (30), or low abundance. Metagenomic binning recovered 20
115 genomes across seven uncultivated lineages including MGII, SAR406, and SAR202, and also
116 from Candidate Phyla previously uncharacterized in the nGOM: Parcubacteria (28),
117 Peregrinibacteria (31), and possibly PAUC34f (32) and ACD39 (33). Our results provide the first
118 information on the likely potential function and activity of these taxa during hypoxia in the
119 shallow nGOM and suggest novel roles for some of these groups that possibly reflect sublineage-
120 specific adaptations.
121
122 **Results**
123
124 ***Study area***
125 Our previous work used 16S rRNA gene amplicon data and qPCR to examine correlations
126 between whole microbial communities, nutrients, and DO across the geographic range of the
127 2013 seasonal hypoxia (21). Here we selected six of those samples from offshore of the region
128 between Atchafalaya Bay and Terrebonne Bay (D', D, and E transects). These sites ranged
129 considerably in DO concentration (~2.2 – 132 $\mu mol \cdot kg^{-1}$), and we chose them to facilitate a
130 detailed investigation of the metabolic repertoire of individual taxa across the span of suboxic (1-
131 20 $\mu mol \cdot kg^{-1}$ DO) to oxic (> 90 $\mu mol \cdot kg^{-1}$ DO) (5) water. Microbial samples from these sites
132 were collected at the oxygen minimum near the bottom. Site depth ranged from 8 – 30 m, with
133 the hypoxic (< 2 $mg \cdot L^{-1}$/62.5 $\mu mol \cdot kg^{-1}$) layer (at sites D2, D3, E2A, and E4) extending up to
134 ~5 m off the bottom (Table S1).
135
136 ***Metagenomic assembly yielded high quality genomes from multiple uncultivated lineages***
137 Our initial assembly and binning efforts recovered 76 genomes. Using a concatenated ribosomal
138 protein tree that included members of the Candidate Phylum Radiation (CPR)(34) (Fig. S1),
139 CheckM (35) (Fig. S2), 16S rRNA genes and other single copy markers where available, and
140 analyses of individual gene taxonomy (Fig. S3), we assigned 20 genomes to uncultivated
141 "microbial dark matter" groups. These were six Marine Group II Euryarchaeota (MGII), five
142 Marinimicrobia (SAR406), three in the SAR202 clade of Chloroflexi, and within Candidate
143 Phyla (CP), one Parcubacteria (OD1), two Peregrinibacteria, and putatively, one ACD39 and two
144 PAUC34f (Table 1). All genomes, with the exception of the Parcubacteria Bin 40, had estimated
145 contamination of less than 6%, and in the majority of cases, less than 2%. Four of the six MGII
146 genomes had estimated completeness (via CheckM) of greater than 61%, four of the five
147 SAR406 greater than 73%, and all three SAR202 genomes were estimated to be greater than
148 83% complete. All CP lineages had at least one genome estimated to be greater than 71%
149 complete (Table 1).

150 The six MGII genomes spanned a range of GC content from 41-62%, which may indicate
151 the presence of sublineages, although their clustering within the ribosomal protein tree did not
152 correspond with GC content (Fig. S1). Average amino acid identity (AAI) supported a close
153 relationship between Bins 14, 17-1, and 18, congruent with the ribosomal phylogeny (Table S1).
154 Expected complete genome sizes, based on estimated completion percentages, were between 1.7
155 and 3.2 Mbp (Table 1). Within SAR406, GC content groupings did correspond to phylogenetic
156 inference. A low GC group (Group A) made up of Bins 51-1, and 51 with 39% and 41% GC,
157 respectively, and a higher GC group (Group B) with Bins 45-2, 45, and 45-1 having 46%, 47%,
158 and 49% GC, respectively. This pattern matched the ribosomal protein tree phylogeny where
159 Bins 45-1 and 45-2 were sister and separated by large branch lengths from Bin 51-1 (Fig. S1).
160 Bins 45 and 51 could not be included in the tree because of missing ribosomal proteins, but they
161 were grouped with SAR406 through 16S rRNA genes, markers on other contigs like the RNA
162 polymerase in Bin 45, similar patterns of protein family expansions (see below), based on their
163 bin of origin, and via AAI calculations (Table S1). During quality control, Bin 51-1 was
164 separated out of Bin 51, and Bins 45-1 and 45-2 out of Bin 45. The pairwise identities of the 16S
165 rRNA genes between bins also support the 45/45-1 and 51/51-1 groupings, as does AAI, where
166 Group A were more similar to each other than Group B, and vice versa (Table S1). The
167 aggregate data therefore supported Bins 45 and 51 as SAR406. Based on 16S rRNA gene
168 phylogeny (Fig. S4), Group A are sister to the Arctic96B-7 clade, and Group B form a
169 monophyletic group with the early diverging SHBH1141 clade identified by Wright and
170 colleagues (26). Expected genome sizes for SAR406 Group A were 2.6-2.7 Mbp and for Group
171 B, 2.8-3.5 Mbp (Table 1). The three SAR202 genomes had very similar GC content (52-53%),
172 were tightly grouped on the ribosomal protein tree on their own branch as early diverging
173 members of the Chloroflexi (Fig. S1), and closely related- possibly at the level of Genus- based
174 on AAI (36) (Table S1). We classified these genomes as subclade I based on 16S rRNA gene
175 phylogeny (Fig. S5), which distinguishes them from the subclade III and V genomes recently
176 reported (24). Estimated complete genome sizes were between 2.2-3.0 Mbp (Table 1). These
177 genomes also had the lowest coding density of all the groups: 88-91% (Table S1).
178 CP lineages had fewer representative genomes. A single genome from Parcubacteria had
179 a GC content of 52% and the two Perigrinibacteria genomes had 39 and 45% GC content. Both
180 of these groups had well-supported positions in the ribosomal protein tree (Fig. S1) that were
181 corroborated by BLASTP-derived annotation of encoded proteins against the NCBI nr database
182 (Fig. S3). We estimated complete genome sizes for Parcubacteria and Perigrinibacteria at near
183 1.5 Mbp (Table 1). Bins 50 and 48 grouped with WS3 bacteria in the ribosomal protein tree,
184 although their relative positions inside the clade were poorly supported at the internal nodes.
185 AAI generally supports the designation of these two bins in the same phylum (37) (Table S1), as
186 does their relationship in the ribosomal protein tree. However, the nearly complete Bin 50 16S
187 rRNA gene sequence had best blast hits to the PAUC34f CP in the GreenGenes database. 16S
188 rRNA gene phylogenetic inference grouped Bin 50 with other PAUC34f sequences sister to the
189 entire WS3 clade (Fig. S6), in conflict with the ribosomal protein phylogeny. The lack of

190 PAUC34f genomes makes it possible that the ribosomal protein tree grouping of Bins 50 and 48
191 occurred simply due to inadequate taxon selection, as these are nearby clades in the 16S rRNA
192 gene tree. Although a rare group, PAUC34f was more abundant than WS3 in the amplicon
193 dataset and we detected 18 PAUC34f operational taxonomic units (OTUs) compared to three
194 assigned to WS3 (21), so reconstruction of PAUC34f genomes was more likely than WS3. Bin
195 13 grouped with genomes designated as ACD39, a group previously identified in an aquifer (33),
196 however Bin 13 was on a long branch and the relationship was poorly supported (Fig. S1). For
197 the Bin 13 and PAUC34f genomes, blast-based identification via best hit annotation did not
198 return clear results- hit annotations included a wide variety of taxa. These results are consistent
199 with what we would expect from novel genomes that have few or no members in the database,
200 however this analysis did not directly corroborate either the ribosomal protein tree phylogeny for
201 all three bins, or the 16S rRNA gene phylogeny for Bin 50. We therefore consider the
202 designation of Bins 50/48 as PAUC34f, and Bin 13 as ACD39, putative. Bins 50 and 48 had 58
203 and 55% GC content, and larger estimated complete genomes than the others at 6.3 and 4.9 Mbp,
204 respectively. Bin 13 also had a large expected genome size at 4.8 Mbp, and GC content of 47%.
205
206 ***Unique roles for the ubiquitous MGII, SAR406, and SAR202 lineages in nGOM hypoxia***
207 Marine group II (MGII) is a broadly distributed archaeal clade, with members found in different
208 marine (38, 39), and sedimentary (40), environments. Previous work during 2012 and 2013
209 hypoxia indicated a proliferation of archaeal taxa in both the Thaumarchaea and MGII phyla (21,
210 22). MGII comprised over 10% of the total community in some samples from 2013, and one
211 MGII OTU also had a strong negative correlation with DO during 2013 hypoxia (21). Within the
212 metagenomics dataset herein, MGII were more abundant in lower oxygen samples than in fully
213 oxic ones, and the most abundant of the lineages reported here (Fig. S7). The majority encoded
214 for aerobic, chemoheterotrophic metabolism, with no predicted genes for nitrogen or sulfur
215 respiration except for a putative nitrite reductase (*nirK*) in a single genome- Bin 15 (Fig.1, Table
216 S1). Abundance of MGII genomic abundance correlated well with transcriptional abundance in
217 most samples (Fig. 3), and we specifically found MGII cytochrome c oxidase expression
218 throughout, though the levels and patterns differed depending on the gene and the source genome
219 (Fig. 4, Table S1). Expression of the *nirK* gene occurred in the D2 and E2A samples- both
220 suboxic. All but the most incomplete genome encoded for ammonia assimilation, making this a
221 likely nitrogen source. Aggregate metabolic construction also indicated a complete TCA cycle,
222 glycolysis via the pentose phosphate pathway, and gluconeogenesis (Fig. 1). Carbohydrate active
223 enzyme (CAZy) genes can provide critical information on the relationships between microbes
224 and possible carbon sources (41). We found few and these were largely restricted to
225 glycosyltransferases (GT) in families 2 and 4, with activities related to cellular synthesis. In
226 general, CAZy expression occurred for at least one gene in every genome and we detected
227 expression of GT cellular synthesis genes in the E2A sample (Fig. 5), likely indicating actively
228 growing cells.

229 Another cosmopolitan group found in our samples was SAR406 or Marine Group A.
230 These organisms were discovered over 20 years ago (27, 42), and the clade has recently been
231 proposed as the phylum "Marinimicrobia" (28). SAR406 occur in numerous marine (5, 25, 27,
232 28, 43), sedimentary (28), and even oil reservoir (44) environments. They are prevalent in deeper
233 ocean waters (27, 43, 45) and prefer lower oxygen concentrations in OMZs (5, 25, 46). SAR406
234 represented over 5% of the population in some locations during hypoxia in 2013, and one
235 abundant OTU was negatively correlated with DO (21). Metagenomic read recruitment to the
236 SAR406 bins confirmed this trend, with greater recruitment in the suboxic samples relative to
237 dysoxic or oxic (Fig. S7). Total RNA recruitment was strongest to Bins 45 and 51-1, though
238 most bins showed a positive RNA to DNA recruitment ratio in at least one sample, indicating
239 these taxa were likely active (Fig. 3).
240 Despite their affinity for low oxygen environments, the SAR406 genomes encoded a
241 predicted capacity for aerobic respiration (Fig. 1), and we found expression of cytochrome c
242 oxidases in even the lowest oxygen samples (Fig. 4). The Group B genomes encoded both high
243 and low-affinity cytochrome c oxidases (47), whereas the high affinity ($cbb_3$-type) oxidases were
244 not recovered in the Group A genomes (Table S1), which may indicate sublineage-specific
245 optimization for different oxygen regimes. Sublineage variation also appeared in genes for the
246 nitrogen and sulfur cycles. Group B genomes all contained predicted nitrous oxide reductases
247 (*nosZ*) and *nrfAH* genes for dissimilatory nitrite reduction to ammonium (defined here as DNRA,
248 although this acronym frequently refers to nitrate, even though that is a misnomer (48)). The
249 *nrfA* genes formed a monophyletic group with *Anaeromyxobacter dehalogenans* 2CP-1, an
250 organism with demonstrated DNRA activity (48) (Fig. S8). The genes also contained conserved
251 motifs diagnostic of the *nrfA* gene (48) (Fig. S9). We observed expression of *nrfAH* and *nosZ* at
252 the sites with the lowest DO concentrations (D2, E2A, and E4) and expression appeared to have
253 a negative relationship with DO concentration (Fig. 4). The Bin 51-1 Group A genome contained
254 predicted *narHI* genes for dissimilatory nitrate reduction, which we did not find in the Group B
255 genomes. We observed expression of SAR406 *narHI* only in the lowest DO sample from station
256 E2A (Fig. 4). Two Group B SAR406 genomes had predicted *phsA* genes for thiosulfate reduction
257 to sulfide (and/or polysulfide reduction (49)), as previously described from fosmid sequences
258 (26). We detected transcripts for these genes only in samples E2A and E4, the two lowest DO
259 samples (Fig. 4). Interestingly, many of the anaerobic respiratory genes were co-expressed with
260 cytochrome c oxidases, indicating a potential for either co-reduction of these alternative terminal
261 electron acceptors or poising of these organisms for rapid switching between aerobic and
262 anaerobic metabolism (50).
263 All SAR406 genomes had numerous genes for heterotrophy. We found CAZy genes in
264 all major categories except polysaccharide lyases, and expression for most of these genes in both
265 Group A and Group B genomes in one or more samples (Fig. 5). Notable carbohydrate
266 compounds for which degradation capacity was predicted include cellulose (glycoside hydrolase
267 (GH) families GH3, GH5; carbohydrate binding module (CBM) family CBM6), starch (GH13),
268 agar and other sulfated galactans (GH2, GH16), chitin (GH18), xylan (GH30, CBM9), and

269 peptidoglycan (GH23, GH103, CBM50). The genomes contained putative transporters for a
270 variety of dissolved organic matter (DOM) components including nucleosides, amino and fatty
271 acids, and oligopeptides (Table S1). We also found numerous outer membrane transporters,
272 including cation symporters; Outer Membrane Receptors (OMR- TonB-dependent), which play
273 important roles in transport of metals, vitamins, colicins, and other compounds; Outer Membrane
274 Factors (OMF); and most genomes also had large numbers of duplicated genes (24 in Bin 45-2),
275 identified via hidden Markov model searches against the SFam database (51), annotated as "Por
276 secretion system C-terminal sorting domain-containing protein," some of which were associated
277 with GH16. These genes likely play a role in sorting C-terminal tags of proteins targeted for
278 secretion via the Por system, which is essential for gliding motility and chitinase secretion in
279 some Bacteroidetes (52). The extensive gene duplication may indicate expanded and/or
280 specialized sorting functionality, and suggests an emphasis on protein secretion in this group.
281 Expression of a membrane-bound lytic murein transglycosylase D (GH23) involved in
282 membrane remodeling, also supports the idea of active and growing cells from Group A in all
283 samples (Table S1). Thus, SAR406 likely perform chemoorganoheterotrophic metabolism,
284 possibly degrading extracellular material, with flexibility in terminal electron accepting
285 processes that implicate them in nitrogen and sulfur cycling in the lowest suboxic DO samples.
286 Members of the SAR202 clade of Chloroflexi also inhabit a wide variety of marine
287 environments (23), frequently in deeper waters (23, 43, 45, 53, 54), and remain functionally
288 understudied, although recent work on subclades III and V has suggested roles in degrading
289 recalcitrant DOM in the mesopelagic (24). We detected Chloroflexi 16S rRNA gene sequences
290 during 2013 hypoxia at up to 5% of the community (21). In contrast to the single-cell genomes
291 examined by Landry et al. (24), the three SAR202 genomes that we recovered belong to subclade
292 I (23) (Fig. S5). Although present at lower abundance than MGII and SAR406 (Fig. S7), these
293 genomes showed relatively high activity in some samples (Fig. 3). Like subclade III and V,
294 subclade I organisms likely respire oxygen, however we also found *napAB* and *nosZ* genes for
295 nitrate and nitrous oxide reduction, respectively (Fig. 1). As in SAR406, we detected concurrent
296 expression of these genes with cytochrome c oxidases in the lowest DO samples (Fig. 4) (Fig. 4,
297 Table S1).
298 The SAR202 genomes have numerous transporters with predicted roles in organic matter
299 transport, which supports previous observations of DOM uptake (53). In particular, SAR202
300 genomes also had considerably more Major Facilitator Superfamily (MFS) transporters than the
301 other genomes in this study (Table S1) and those of the subclade III genomes (24), and SFam
302 searches revealed a the majority of these shared annotation as a "Predicted arabinose efflux
303 permease" (SFam 346742). MFS genes transport numerous diverse substrates, such as sugars and
304 amino acids, through coupling with an ion gradient, and can be associated with either uptake or
305 export of compounds (55). SAR202 genomes also had between 53 and 66 predicted ABC
306 transporters. The largest gene family expansion that we observed in SAR202 was associated with
307 SFam 6706, with between 46 and 48 genes in this family encoded in each genome. Most of these
308 (121/142) were annotated as either a "galactonate dehydratase" or a "L-alanine-DL-glutamate

309 epimerase." Galactonate dehydratase catalyzes the first step of the pathway to utilize D-
310 galactonate in central carbon metabolism via the pentose phosphate pathway. The large number
311 of genes in these categories likely indicate some divergence for alternative roles as this group
312 belongs broadly to the COG4948 "L-alanine-DL-glutamate epimerase or related enzyme of
313 enolase superfamily." All genomes also had numerous dehydrogenases as reported for the
314 subclade III genomes (24). Specifically, SFams 346640 and 1639 were the third and fourth most
315 abundant, with 16-18 and 13-15 genes in each family, respectively, across the three genomes.
316 Genes in these families were annotated as "short-chain alcohol dehydrogenase family," "3-alpha
317 (or 20-beta)-hydroxysteroid dehydrogenase," "meso-butanediol dehydrogenase," and others.
318 These match the annotations of the subclade III genomes, and suggest a similar role in
319 conversion of alcohols to ketones (24). The SAR202 genomes have comparatively few CAZy
320 genes relative to the other genomes. GH15 and GH63 suggest starch degradation, and GH105
321 pectin degradation, and we detected expression of multiple genes in these categories across
322 samples (Fig. 5, Table S1).
323
324 ***Other Candidate Phyla organisms in nGOM hypoxia***
325 In contrast to the abundant and cosmopolitan MGII, SAR406, and SAR202 clades, we also
326 recovered genomes from several groups that were either previously undetected in the nGOM or
327 very rare. Although these taxa likely do not contribute the biomass of more populous clades,
328 their genomes provide important insight into their functional potential during hypoxia. The Bin
329 13 genome (possibly ACD39) also had the highest relative activity compared to all the other
330 genomes in our study (Fig. 3), underlining the point that low abundance does not automatically
331 equate to low metabolic impact. Bin 13 had predicted aerobic respiration with both high and low-
332 affinity cytochrome c oxidases (Fig. 2), however, the low affinity oxidases contributed more
333 reads in the samples where we could detect expression (Table S1). The genome contained
334 numerous predicted CAZy genes in the glycosyltransferase and glycoside hydrolase categories,
335 spread across numerous families in each (Table S1). Notable degradation capacity included
336 starch (GH13) and peptidoglycan (GH23, GH103, GH104).
337 Bin 13 had ~ 80 ABC transporter genes, and similarly to the SAR406 genomes,
338 numerous outer membrane transporters, including the OMR and OMF families. We predict
339 complete glycolysis/gluconeogenesis pathways and a TCA cycle. We recovered paralogous pilus
340 subunit genes, chemotaxis genes, and a partial flagellar assembly. Furthermore we detected
341 relatively high expression of the *flgLN* flagellin genes in samples D2, D3, E2A, and E4 (Table
342 S1) suggesting active motility in these environments. Several other Bin 13 genes were among the
343 most highly expressed in all samples, but could only be classified as hypothetical (Table S1).
344 Similarly, the three most populous SFams in Bin 13, according to number of genes (n=16, 15,
345 and 13) also linked to genes annotated as hypothetical proteins with either tetratricopeptide,
346 HEAT, TPR, or Sel1 repeats. Although currently obscure, these and the highly expressed
347 hypothetical genes represent important targets for future research into the function of this group.
348 The organism likely leads an aerobic heterotrophic lifestyle, with substantial capacity for

349 interacting with the environment through numerous transporters and possible surface attachment
350 via multiple motility genes.
351 Bins 50 and 48 were lower in abundance than SAR202 genomes, with no observable
352 trend associated with oxygen levels (Fig. S7). These genomes encoded flagellar motility, aerobic
353 respiration, glycolysis via the pentose-phosphate pathway, gluconeogenesis, assimilatory sulfate
354 reduction, and DNRA (Fig. 2). The *nrfA* subunit from both genomes grouped in the same
355 monophyletic clade as those from SAR406 (Fig. S8), and had similar conserved motifs (Fig. S9).
356 However, we note that the *nrfAH* gene sets for Bins 50 and 48 occurred on relatively short
357 contigs (5650 and 5890 bp, respectively), so the metabolic assignment cannot be corroborated as
358 definitively as that for SAR406. The Bin 50 genome was among the more active in our analysis
359 (Fig. 3), and we detected highest expression of cytochrome c oxidase components in samples
360 E2A and E4 (Fig. 4). DNRA gene expression was low but observable in the same samples. We
361 also recovered a partial gene for the ribulose-bisphosphate carboxylase (RuBisCO) large subunit,
362 however this fragment was on a very short contig (3954 bp), and we did not detect expression in
363 any of our samples, so we cannot rule out that this gene occurred on a contaminating contig.
364 These genomes had abundant CAZy genes in all categories, suggesting a highly flexible
365 metabolic repertoire for carbon acquisition. They contain possible capacity for breakdown of
366 starch (GH13, CBM48), peptidoglycan (GH23, CBM50), fructose-based oligosaccharides
367 (GH32), and hemicellulose (GH2, GH3, GH43). Notably these genomes were the only ones with
368 predicted polysaccharide lyases (PL) among those compared (with the exception of a single
369 predicted PL gene in SAR406- Table S1). PL genes cleave uronic-acid containing
370 polysaccharides (56). These organisms seem particularly adapted for pectin (PL1, PL2, PL9,
371 PL10, PL11, PL22, GH78) and alginate (PL15, PL17) degradation- both compounds are
372 common cell wall components of green and brown algae, respectively. In line with the algal cell
373 wall degradation ability, we detected a large expansion (102 genes in Bin 50) of sulfatase genes
374 in SFam 1534, annotated predominantly as either "arylsulfatase A" or "choline-sulfatase."
375 Arylsulfatases cleave sulfate esters, usually to supply microbes with a source of sulfur, and can
376 be located intracellularly or in membranes (57). Choline-sulfatases cleave choline sulfate to
377 choline and sulfate, with downstream use for the former as a carbon source or osmoprotectant
378 and the latter as a sulfur source (58). Given the predicted assimilatory sulfate reduction pathway
379 in Bins 50 and 48, this is a logical means to obtain sulfur for the group. We observed large
380 expansions in galactonate and other dehydratases (as in SAR202, above- SFam 6706 n=42 in Bin
381 50), as well as numerous ABC transporter permeases (SFam 4442), which match the transporter
382 prediction via IMG: 117 predicted genes for ABC transporters in all. These genomes also had
383 numerous OMF and OMR transporter genes (Table S1). The large number of transporters and
384 protein family expansions correspond to the relatively large expected genome sizes (between 5
385 and 6 Mbp). Thus, we expect these organisms to have an aerobic, potentially facultatively
386 anaerobic, multifaceted chemoorganoheterotrophic metabolism with roles in complex carbon
387 compound degradation (like that of algal cell walls) and the nitrogen cycle.

388 We also recovered genomes associated with CPR taxa usually associated with anoxic
389 environments: two Peregrinibacteria and one from the Uhrbacteria subclade of the Parcubacteria
390 (formerly OD1). All three genomes could be assigned taxonomically with high confidence based
391 on their positions in the ribosomal protein tree (Fig. S1) and via gene annotations (Fig. S3). We
392 note that although the Peregrinibacteria bins (16 and 39) had very low predicted contamination,
393 the Parcubacteria Bin 40 has 15% predicted contamination (75% of which we attribute to strain
394 heterogeneity in the bin) (Table 1). Recovery of Parcubacteria from a coastal marine system is
395 unusual, but not unprecedented. Parcubacteria single-cell genomes have been identified in
396 marine and brackish sources (28), and we previously identified 26 rare OTUs assigned to the
397 phylum in nGOM hypoxia (21). That number of low abundance OTUs may explain why we have
398 20 single copy marker genes present in two copies in Bin 40 (Table S1). In contrast to
399 Parcubacteria, Peregrinibacteria have thus far only been found in terrestrial subsurface aquifers
400 (31, 33, 59, 60), and remained undetected in our amplicon survey (21). Both groups occurred in
401 low relative abundance to the other taxa in this study, and showed the lowest activity (Fig. 3).
402 Consistent with previous reports of obligate fermentative metabolism by Parcubacteria and
403 Peregrinibacteria (28, 30, 31, 59), we identified no respiratory pathways for these taxa (Fig. 2)
404 and they trended towards greater abundances in the lowest DO samples (Fig. S7). In spite of
405 relatively high predicted genome completion, we found very few CAZy genes, and those were
406 mostly restricted to glycosyltransferases (Table S1), hence probably involved in capsular
407 polysaccharide synthesis. While these organisms had low relative abundance to the other groups
408 (Fig. S7), we did observe activity in some samples (E2, E2A, D1), suggesting these organisms
409 are either insensitive to oxygen or occupy microanoxic niches (Fig. 3).
410
411 **Discussion**
412 This work provides the first reconstruction of multiple nearly complete genomes from
413 uncultivated bacterioplankton during nGOM hypoxia. Although we define roles for MGII,
414 SAR406, SAR202, Bin 13 and Bins 50/48 as aerobic heterotrophs, we also observed concurrent
415 expression of genes associated with anaerobic metabolism in SAR406 (nitrate reduction, DNRA,
416 nitrous oxide reduction, and sulfur reduction), SAR202 (nitrate and nitrous oxide reduction),
417 MGII (nitrite reduction), and Bins 50/48 (DNRA) in suboxic samples with the lowest measured
418 DO concentrations. Simultaneous utilization of multiple electron acceptors with different redox
419 potentials likely indicates an abundant supply of electron donors (61), may denote niche
420 partitioning within group sublineages at a finer level of taxonomic resolution than we observed,
421 or indicate poising of taxa for rapidly changing chemical gradients (50). An organism's set of
422 carbohydrate active enzyme (CAZy) genes often gives insights into its biology, in particular into
423 nutrient sensing and acquisition. All taxa examined in this study had predicted
424 chemoorganoheterotrophic metabolism, and the CAZy genes found in these genomes suggest
425 that SAR406, SAR202, Bin 13, and Bins 50/48 participate in the degradation of complex organic
426 matter resulting from the detritus of larger organisms. This matches the general model of
427 hypoxic zone oxygen consumption resulting from sinking organic matter provided by algal

428 blooms in surface waters (1). The observed activity of obligate fermentative groups
429 Parcubacteria and Peregrinibacteria also suggests that anoxic pockets occur in the water column
430 where these organisms can thrive.
431 The prevalence of MGII among lower oxygen samples in the hypoxic zone is somewhat
432 surprising, considering that they are commonly associated with aerobic environments (39).
433 However, oxygen was still present in even the lowest DO samples, and MGII success likely had
434 more to do with the carbon content than oxygen levels. These nGOM MGII appear to be
435 metabolically similar to those described in previous work: MGII have been shown to be
436 dominant in water column environments associated with blooms in productivity, for example at
437 deep-sea hydrothermal plumes (38). Thus, the increased availability of organic matter (proteins
438 and carbohydrates), thought to be preferred substrates for MGII (62, 63), probably explains their
439 abundance.
440 Members of the SAR406 clade have frequently been observed in OMZs (5, 25), and our
441 data now define roles for them in the eutrophication-driven hypoxia of the nGOM. Our genomes
442 had larger estimated genome sizes- 2.6-2.7 Mbp (Group A) and 2.8-3.5 Mbp (Group B)-
443 compared to 1.1-2.4 Mbp from single-cell genomes (28). Overall GC content, however, was in
444 the range of the 30-48% reported for fosmids (26) and single-cell genomes (28). The lower GC
445 Group A genomes specifically had a similar GC content to the Arctic96B-7 fosmids, matching
446 their predicted phylogenetic affiliation (see below) (26). Previous metabolic reconstructions of
447 SAR406 predicted aerobic metabolism (28) and sulfur reduction (26), which our data confirm,
448 although the sulfur reduction genes were only found in Group B organisms (Table S1). Our
449 genomes now also suggest multiple nitrogen cycling roles that appear to be organized by
450 sublineages within the phylum, and sublineage specific presence of both high and low affinity
451 cytochrome c oxidases. Specifically, Group B organisms contained both types of cytochrome c
452 oxidases, *nosZ* and *nrfAH* genes, whereas Group A organisms contained the low affinity
453 cytochrome c oxidases only, and additionally *narHI* genes not found in Group B. The unique
454 roles predicted for these taxa is not surprising given the diversity of the SAR406 clade and the
455 genetic distances between Group A and B (Fig. S4). Group A organisms were sister to the
456 Arctic96B-7 clade. Previous fosmids associated with this clade contained genes for oxidative
457 stress and sulfur reduction, although we only found sulfur reduction genes in the distantly related
458 Group B genomes. The ArcticB96-7 clade may be diverse enough to encompass differing
459 metabolic strategies, however the variable presence of *phsA* genes in this group may simply be
460 due to incomplete genomes (herein) and fosmids (26). The Group B organisms group with the
461 early diverging SBH1141 clade (26), for which no previous genome data exist. In addition to
462 sublineage-specific respiratory characteristics, our results also generate specific hypotheses about
463 organic matter metabolism in SAR406: likely degradation capacity for cellulose, starch, agar,
464 xylan, and peptidoglycan; transport of nucleosides, amino and fatty acids, and oligopeptides; and
465 substantial gene duplication associated with protein secretion for possible extracellular
466 metabolism. Together these data suggest that during nGOM hypoxia SAR406 members degrade

467 complex carbohydrates fueled by aerobic respiration, and supplemented with facultative
468 anaerobic respiration of nitrate, nitrite, or sulfur compounds.
469 Genome data for SAR202 have been lacking until recently. Landry and colleagues have
470 recently described the properties for several single-cell genomes representing subclades III and V
471 of SAR202 recovered from the mesopelagic (24). Our genomes have generally higher GC
472 content and much lower expected genome sizes than those predicted by Landry et al., although
473 these calculations are likely complicated by the relative incompleteness of the genomes (8-47%).
474 Their genomes indicated a role for SAR202 in the oxidation of recalcitrant dissolved organic
475 matter, and specifically cyclic alkanes, via flavin mononucleotide monooxygenases (FMNOs)
476 and different dehydrogenases that occurred in paralogous groups (24). We observed many of the
477 same gene expansions, namely that of MFS transporters and short-chain dehydrogenases (and
478 related genes), but we did not recover any FMNOs of SFams 4832 or 4965, suggesting subclade
479 and/or niche-specific adaptations. Furthermore, we observed *napAB* and *nosZ* genes for nitrate
480 and nitrous oxide reduction (and expression of these genes), which were not reported for
481 subclade III or V. The nGOM hypoxia SAR202 genomes had CAZy genes implicating them in
482 degradation of complex compounds such as chitin and pectin. The emerging picture of these taxa
483 from both shallow hypoxic waters and the mesopelagic is one of recalcitrant carbon degraders,
484 with overlapping suites of paralogous genes, but that may be specialized for specific compounds
485 more commonly available in their respective habitats.
486 This study has also developed roles for CP taxa in a shallow marine water column during
487 hypoxia. The most active organism in our survey based on the ratio of RNA to DNA reads
488 recruited, Bin 13, putatively belongs to a group with little genomic data- ACD39. The original
489 ACD39 genome was reconstructed from an aquifer community (33). Although this was only a
490 partial genome, it shared some features with our putative ACD39 member, namely pilin and
491 chemotaxis genes, those containing TPR and tetratricopeptide repeats, and CAZy genes for
492 degradation of complex compounds such as starch (33). Our study provides evidence that these
493 taxa have relatively large genomes (~4.8 Mbp), are active aerobes in nGOM hypoxia, and have
494 chemotaxis and motility genes that could facilitate scavenging and surface attachment. However,
495 most of the highly expressed genes in this organism were annotated as hypothetical proteins, so
496 much of the function of these organisms remains to be uncovered.
497 Bins 50 and 48 provide novel genome data for bacterioplankton in nGOM hypoxia,
498 although the exact taxonomic position of these bins remains in conflict. The ribosomal protein
499 tree provides evidence that these taxa belong to the Latescibacteria (WS3) (Fig. S1), but 16S
500 rRNA genes (Fig. S6) and our amplicon data point toward membership in the more poorly
501 understood PAUC34f clade. Since no previous genome data exist for PAUC34f, we cannot rule
502 out erroneous assignment in the ribosomal protein tree due to insufficient taxon selection. Bins
503 50 and 48 represented the largest genomes of the study, with estimated complete sizes of ~5-6
504 Mbp, and numerous genes suggesting degradation of a wider suite of complex organic matter
505 than any of the other genomes examined. For example, they were the only genomes with
506 numerous polysaccharide lyase genes, and these likely facilitate breakdown of algal cell wall

507 components like pectin and alginate. The Bin 50 genome was among the most active across all
508 samples (Fig. 3), and we detected expression of cytochrome c oxidase genes, and those for
509 DNRA, in both the Bin 50 and 48 genomes. Thus we expect these organisms to play a role in the
510 breakdown of algal detritus through respiration of oxygen, with nitrite as an alternative electron
511 sink.
512 If these bins belong to PAUC34f, they represent the first genomic data for the group.
513 Although originally discovered, and commonly found, in marine sponges (32, 64-66), this
514 putative bacterial phylum (via GreenGenes/SILVA) has been detected as a rare group in other
515 marine invertebrates (67) and stream sediment (68), and we identified 18 distinct but rare
516 PAUC34f OTUs in nGOM hypoxia, compared to just three from WS3 (21). Although the
517 majority of studies suggest an endosymbiotic lifestyle for PAUC34f, our representative genome
518 data point towards a free-living existence with multiple terminal electron accepting processes,
519 motility genes for seeking more favorable conditions, and a large metabolic repertoire for
520 degradation of complex compounds. On the other hand, if these genomes represent WS3, the
521 sister clade to PAUC34f (Fig. S6), they have many similarities to the lifestyles inferred from
522 recent metagenomic investigations (69, 70). Specifically, while this group was previously
523 considered anaerobic (69), new data have supported an aerobic lifestyle for some members (59),
524 and revealed complete electron transport chains and both high and low affinity cytochrome c
525 oxidases (70). The Bin 50 and 48 genomes predict aerobic metabolism as well, although only
526 with low affinity cytochrome c oxidases. Farag et al. also found little evidence of these taxa in
527 host-associated environments, contrary to PAUC34f sequence data (70). The enrichment of PL
528 family genes in Bins 50 and 48, polysaccharide degradation capability in general, and specific
529 genes for degradation of cell wall components, all corroborate previous findings on WS3 as well
530 (70). Bin 50 had 78 annotated peptidases, nearly double that in all other genomes in the study
531 (Bin 48 had 46), which also concurs with metagenomic predictions for WS3 (70). Our genomes
532 differed from WS3 metagenomes principally in the predicted DNRA metabolism and the
533 dramatic expansion of sulfatases. Although sulfatases were observed in WS3 metagenomes (70),
534 they were not present in the numbers associated with Bin 50 (n=102). A large cadre of sulfatases
535 has been previously reported for *Lentisphaera* (n=267) and *Pirellula* (n=110) genomes (71, 72)
536 and suggests specialization for degradation of sulfate-esters to satisfy carbon and/or sulfur
537 requirements.
538 Although Parcubacteria and Peregrinibacteria occurred in low abundance and we detected
539 activity in only a few samples, their recovery in the hypoxic zone is notable because these
540 organisms have generally been associated with anoxic environments. Our predicted genome sizes
541 (~1.5 Mbp) corroborate previous reports of these organisms having small genomes (31, 59). We
542 did not observe any genes associated with nitrogen or sulfur redox transitions, although we
543 cannot rule these capabilities entirely due to incomplete genomes. Regardless, we can
544 hypothesize that Parcubacteria and Peregrinibacteria persist as members of the rare biosphere
545 until they can take advantage of microanoxic niches in the water column where they participate
546 in carbon cycling as obligately fermentative organisms.

547
548 ***Conclusions***
549 Our study helps define the metabolism for many of the most enigmatic organisms observed in
550 cultivation-independent studies during nGOM hypoxia. We also provide new information about
551 sublineage specific attributes for the important cosmopolitan marine clades SAR406 and
552 SAR202, for which previous metabolic information has been limited. Notably, although
553 Parcubacteria and Peregrinibacteria likely subsist via an obligate anaerobic lifestyle, the
554 remaining uncultivated groups in the nGOM hypoxic zone had one or more genomes that
555 encoded cytochrome c oxidases (and other electron transport chain components) for respiring
556 oxygen, making these taxa likely only facultative anaerobes. Pervasive aerobic metabolism in an
557 oxygen-depleted water column may seem counterintuitive, yet despite DO being as low as 2.2
558 $\mu$mol $kg^{-1}$ in the E2A sample, oxygen probably remained high enough to sustain aerobic
559 microbes. As little as ~0.3 $\mu$mol $kg^{-1}$ oxygen inhibited denitrification in OMZ populations by
560 50% (73), and even *Eschericia coli* K-12 could grow aerobically at oxygen concentrations as low
561 as 3 nM (74). Thus, for many organisms, active aerobic respiration likely persists even in
562 suboxic waters. Nevertheless, our data also suggests pervasive co-reduction of alternative
563 terminal electron acceptors (oxygen, nitrate, nitrite, nitrous oxide, and sulfur), sometimes within
564 the same organism (Fig. 4). Co-reduction of electron acceptors with different redox potentials
565 across a community could indicate microniches and/or aggregates in the water column where DO
566 concentrations drop below bulk values (50). Alternatively this can occur with an abundance of
567 electron donor, and overlapping redox processes have been reported in multiple environments,
568 including aquatic ones (61, 75). Concurrent expression of genes for multiple terminal electron
569 accepting processes within a single organism has been proposed as a means of improved
570 readiness for dynamic conditions, albeit at the cost of lower productivity (50). Given that many
571 uncultivated taxa likely perform multiple terminal electron accepting processes (and possibly do
572 so simultaneously), and we found a comparative cornucopia of genes for degradation of
573 chemoorganohetrotrophic energy sources, we hypothesize that niche differentiation within
574 uncultivated hypoxic zone bacterioplankton occurs predominantly via specialization for different
575 oxidizable substrates rather than for distinct roles in the canonical redox cascade (4, 5).
576 Importantly, many of the active uncultivated taxa also appeared adapted for degradation of
577 complex carbon substrates. Such compounds might comprise the bulk of available organic matter
578 during the later stages of hypoxia after initial oxygen depletion by microorganisms feeding on
579 more labile carbon sources. Selection for chemoorganotrophic microbes adapted to utilize
580 recalcitrant organic matter could also explain why organisms that do not require an exogenous
581 carbon source, such as the chemolithoautotrophic *Nitrosopumilus*, proliferate during hypoxia
582 (21, 22) compared to their levels during spring before DO decreases (76, 77). Temporal data on
583 the relative abundance and activity of these organisms, and of organic matter composition in the
584 water column, will be critical to more fully understand the relationship of bacterioplankton to the
585 creation, maintenance, and dissolution of nGOM hypoxia.
586

587
588 **Materials and Methods**
589 *Sample selection and nucleic acid processing.* Six samples representing hypoxic (n=4) and oxic
590 (n=2) DO concentrations were picked from among those previously reported (21) at stations D1,
591 D2, D3, E2, E2A, and E4 (Table S1). DO, and nutrient collection information is detailed in
592 Gillies et al., 2015. Nucleic acids were collected as follows: At these six stations 10 L of
593 seawater was collected and filtered with a peristaltic pump. A 2.7 µM Whatman GF/D pre-filter
594 was used and samples were concentrated on 0.22 µM Sterivex filters (EMD Millipore). Sterivex
595 filters were immediately sparged, filled with RNAlater, and placed at -20˚C, at which they were
596 maintained until extraction. DNA and RNA were extracted directly off of the filter by placing
597 half of the Sterivex filter in a Lysing matrix E (LME) glass/zirconia/silica beads Tube (MP
598 Biomedicals, Santa Ana, CA) using the protocol described in Gillies *et al.* (2015) which
599 combines phenol:chloroform:isoamyalcohol (25:24:1) and bead beating. Genomic DNA and
600 RNA were stored at -80°C until purified. DNA and RNA were purified using QIAGEN
601 (Valencia, CA) AllPrep DNA/RNA Kit. DNA quantity was determined using a Qubit2.0
602 Fluorometer (Life Technologies, Grand Island, NY). RNA with an RNA integrity number (RIN)
603 (16S/23S rRNA ratio determined with the Agilent TapeStation) ≥ 8 (on a scale of 1-10, with 1
604 being degraded and 10 being undegraded RNA) was selected for metatranscriptomic sequencing.
605 Using a Ribo-Zero kit (Illumina) rRNA was subtracted from total RNA. Subsequently, mRNA
606 was reverse transcribed to cDNA as described in Mason *et al.* (2012) (78).
607
608 *Sequencing, assembly, and binning.* DNA and RNA were sequenced separately, six samples per
609 lane, with Illumina HiSeq 2000 chemistry to generate 100 bp, paired-end reads (180 bp insert
610 size) at the Argonne National Laboratory Next Generation Sequencing facility. The data are
611 available at the NCBI SRA repository under the BioSample accession numbers
612 SAMN05791315-SAMN05791320 (DNA) and SAMN05791321-SAMN05791326 (RNA).
613 DNA sequencing resulted in a total of 416,924,120 reads that were quality trimmed to
614 413,094,662 reads after adaptors were removed using Scythe
615 (https://github.com/vsbuffalo/scythe), and low-quality reads (Q < 30) were trimmed with Sickle
616 (https://github.com/najoshi/sickle). Reads with three or more Ns or with average quality score of
617 less than Q20 and a length < 50 bps were removed. Genomes were reconstructed using two
618 rounds of assembly. Metagenomic reads from all six samples were pooled, assembled, and
619 binned using previously described methods (79, 80). Briefly, quality filtered reads were
620 assembled with IDBA-UD (81) on a 1TB RAM, 40-core node at the LSU High Performance
621 Computing cluster SuperMikeII, using the following settings: -mink 65 –maxk 105 –step 10 –
622 pre_correction –seed_kmer 55. Initial binning of the assembled fragments was performed using
623 tetra-nucleotide frequency signatures using 5 kbp fragments of the contigs. Emergent self-
624 organizing maps (ESOM) were manually delineated and curated based on clusters within the
625 map (Fig. S8, S9). The primary assembly utilized all reads and produced 28,080 contigs ≥ 3 kb
626 totaling 217,715,956 bp. Of these, 303 contigs were over 50 kb, 72 over 100 kb, and the largest

627 contig was just under 495 kb. Binning produced 76 genomes, of which 20 genomes were
628 assigned to lineages with uncultivated representatives using CheckM, ribosomal protein trees,
629 and 16S rRNA gene sequences (below).
630
631 *DNA and RNA mapping.* Metagenomic and metatranscriptomic sequencing reads from each
632 sample were separately mapped to binned contigs using BWA (82) to compare bin abundance
633 across samples and facilitate bin cleanup (below). Contigs within each bin were concatenated
634 into a single fasta sequence and BWA was used to map the reads from each sample to all bins.
635 All commands used for these steps are available in supplementary information.
636
637 *Bin QC.* Bins were examined for contamination and completeness with CheckM (35), and we
638 attempted to clean bins with > 10% estimated contamination using a combination of methods.
639 First, the CheckM modify command removed contigs determined to be outliers by GC content,
640 coding density, and tetranucleotide frequency. Next, in bins that still showed > 10%
641 contamination, contigs were separated according to comparative relative abundance of mean
642 DNA read coverage by sample. Final bins were evaluated with CheckM again to generate the
643 statistics in Table S1 and final bin placements in the CheckM concatenated gene tree (Fig. S2).
644
645 *Ribosomal protein tree.* The concatenated ribosomal protein tree was generated using 16
646 syntenic genes that have been shown to undergo limited lateral gene transfer (rpL2, 3, 4, 5, 6, 14,
647 15, 16, 18, 22, 24 and rpS3, 8, 10, 17, 19) (83). Ribosomal proteins for each bin were identified
648 with Phylosift (84). Amino acid alignments of the individual ribosomal proteins were generated
649 using MUSCLE (85) and trimmed using BMGE (86) (with the following settings: -m
650 BLOSUM30 –g 0.5). The curated alignments were then concatenated for phylogenetic analyses
651 and phylogeny inferred via RAxML v 8.2.8 (87) with 100 bootstrap runs (with the following
652 settings: mpirun -np 4 -npernode 1 raxmlHPC-HYBRID-AVX -f a -m PROTCATLG -T 16 -p
653 12345 -x 12345 -# 100). Note this is similar to the number utilized in a previous publication for
654 this tree with automated bootstrapping (88), and required just over 56 hours of wall clock time.
655 The alignment is available in SI.
656
657 *Average amino acid identity*. AAI was calculated with Get Homologues (89) v. 02032017, with
658 the following settings: –M –t 0 –n 16 –A.
659
660 *Taxonomic assignment*. Taxonomy for each bin was assigned primarily using the ribosomal
661 protein tree. However, for bins that did not have enough ribosomal proteins to be included in the
662 tree, or for which the placement within the tree was poorly supported, assignments were made
663 based on the concatenated marker gene tree as part of the CheckM analysis (Fig. S2), or via 16S
664 rRNA gene sequences, when available. 16S rRNA genes were identified via CheckM, and these
665 sequences were aligned against the NCBI nr database using BLASTN to corroborate CheckM
666 assignments. In the case of the SAR202 genomes, which did not have representative genomes in

667 either the ribosomal protein tree or the CheckM tree, the 16S rRNA gene sequences for two of
668 the three bins (43-1, 43-2) were available and aligned with the sequences used to define the
669 SAR202 clade (23) (Fig. S5). Alignment, culling, and inference were completed with MUSCLE
670 (85), Gblocks (90), and FastTree2 (91), respectively, with the FT_pipe script. The script is
671 provided in SI. The 16S rRNA gene tree for subclade assignment of SAR406 (Fig. S4) was
672 assembled by blasting the four 16S sequences predicted by CheckM against a local GenBank nt
673 database using blastn (v. 2.2.28+) (92), selected the top 100 non-redundant hits to each sequence,
674 and manually removing all hits to genome sequences. These were combined with previously
675 defined SAR406 subclade reference sequences (25), fosmid 16S sequences (26), single-cell
676 genome sequences (28), and run through alignment, culling, and inference with FT_pipe. Taxa
677 with identical alignments were removed with RAxML v 8.2.8 (87) using default settings, and the
678 final tree was inferred using FastTree2 (91). For putative CP genomes, taxonomy was also
679 evaluated by examining the taxonomic identification for each of the predicted protein sequences
680 after a BLASTP search against the NCBI nr database. Post-blast, the number of assignments to
681 the dominant one or two taxonomic names, along with the number of assignments to "uncultured
682 bacterium," was plotted for each genome according to the bit score quartile (Fig. S3). Quartiles
683 were determined in R using the summary function. Bin 56 has two ribosomal protein operons on
684 scaffold_2719/Ga0113622_1153 and scaffold_21777/Ga0113622_1009. In the ribosomal protein
685 tree, the former placed the organism in the *Planctomycetaceae*, while the latter (which was much
686 smaller) placed the organism in CP WS3. The majority of BLASTP annotations to the nr
687 database matched *Planctomycetaceae* taxa, as did the 16S rRNA gene sequences found in the
688 genome, so Bin 56 organism was designated a Planctomycetes and not WS3, and excluded from
689 this study. The 16S rRNA gene from Bin 50 was also used to infer taxonomic identity using and
690 established phylogeny for the WS3 clade (70) and relevant outgroups. The Bin 50 sequence was
691 blasted against the greengenes database (Dec. 2013) with megablast, and since many of the top
692 hits belonged to the PAUC34f clade, these were included with the sequences from Farag et al.
693 2017. Alignment, culling, and inference was completed with FT_pipe. Node labels were
694 constructed with the newick utilities (93) script nw_rename.
695
696 *Metabolic reconstruction*. Post-binning, genomes were submitted individually to IMG (94) for
697 annotation. Genome accession numbers are in Table S1, and all are publically available.
698 Metabolic reconstruction found in Table S1 and Figs S5-7/ S11-13, came from these annotations
699 and inspection with IMG's analysis tools, including KEGG pathway assignments and transporter
700 predictions. Transporters highlighted for DOM uptake were identified based on information at
701 the Transporter Classification Database (95). Carbohydrate-active enzymes (CAZymes) were
702 predicted using the same routines as those in operation for the updates of the carbohydrate-active
703 enzymes database (www.cazy.org) (96).
704
705 *RPKM abundance of taxa and genes.* Abundance of taxa within the sample was quantified by
706 evaluating mapped reads using Reads-Per-Kilobase-Per-Million (RPKM) normalization (97)

707 according to $A_{ij} = (N_{ij}/L_i) \times (1/T_j)$, where $A_{ij}$ is the abundance of bin $i$ in sample $j$, $N_{ij}$ is the
708 number of reads that map to bin from sample $j$, $L_i$ is the length of bin $i$ in kilobases, and $T_j$ is the
709 total number of reads in sample $j$ divided by $10^6$. These values were generated for all bins, with
710 only the data for the 20 uncultivated bins reported here. All contigs within a given bin were
711 artificially concatenated into "supercontigs" prior to mapping. $N_{ij}$ was generated using the
712 samtools (82) idxstats function after mapping with BWA. The data in Fig. S7 were created by
713 summing $(N_{ij}/L_i)$ for groups of taxa defined in Table S1 prior to multiplying by $(1/T_j)$. RNA
714 coverage was used to evaluate both bin and gene activity for all bins. Mean coverage for each
715 supercontig was calculated using bedtools (98) and bins were assigned a rank from lowest mean
716 recruitment (1) to highest mean recruitment (2). Bins with particularly high or low activity
717 (transcript abundance) relative to their abundance (genome abundance) were identified using
718 rank-residuals, calculated as follows: On a plot of DNA coverage rank vs. RNA coverage rank,
719 residuals for each bin or gene were calculated from the identity. As the rank-residuals followed a
720 Gaussian distribution, bins with a residual that was > 1 s.d. from the rank-residual mean were
721 classified as having higher-than-expected transcriptional activity; bins with a residual that was <
722 1 s.d. from the mean were classified as having lower-than-expected transcriptional activity.
723 RPKM values were also calculated for every gene in every bin analogously to that for bins, using
724 RNA mapping values extracted with the bedtools multicov function. Sample E2 was omitted
725 from gene-specific calculations as only 4588 transcriptomic reads mapped successfully from this
726 sample, compared to >100,000 from other samples. 17,827 of 140,347 genes had no evidence of
727 expression in any sample and so were removed from further analysis. 3,840 genes recruited reads
728 in all remaining samples. All calculations are available in Table S1 or the R markdown document
729 Per.gene.RPKM.Rmd in Supplemental Information. Table S1 includes only analyzed data for the
730 uncultivated bins reported in this study. Note that RPKM values indicate abundance
731 measurements across a small number of samples. While we can evaluate the relative expression
732 of genes for those samples, our dataset lacks sufficient power to evaluate estimates of
733 significance in differential expression.
734
735 *nrfA sequence assessment.* Initial annotation of our bins identified putative homologs to the
736 *nrfAH* genes associated with dissimilatory nitrite reduction to ammonia. Since *nrfA*-type nitrite
737 reductases can be misannotated due to homology with other nitrite reductases, annotation for
738 these genes was curated with phylogenetic analysis using known *nrfA* genes (48) (Fig. 3)
739 obtained via Dr. Welsh (personal communication). Alignment, culling, and inference were
740 completed with the FT_pipe script. The tree was rooted on the designated outgroup octaheme
741 nitrite reductase sequence from *Thioalkalivibrio nitratireducens* ONR. Node labels were
742 constructed with the newick utilities (93) script nw_rename. Visualization of the alignment (Fig.
743 S9) to confirm the presence of the first CXXCK/CXXCH and highly conserved KXQH/KXRH
744 catalytic site was completed with the MSAViewer (99) online using the un-culled *nrfA* alignment
745 as input.
746

747 *SFam homology searches.* To identify group specific expansions in particular gene families, we
748 performed a homology search of all predicted protein coding sequences in each bin against the
749 Sifted Families (SFam) database (51) using hmmsearch (HMMER 3.1b (100)) with default
750 settings except for the utilization of 16 cpus per search.
751
752 **Funding Information**
753 Funding for this work was provided to JCT through the Oak Ridge Associated Universities
754 Ralph E. Powe Junior Faculty Enhancement Award and the Louisiana State University
755 Department of Biological Sciences. A portion of the funding for this work was provided by a
756 Planning Grant award to OUM from Florida State University. Funding for the research vessel
757 and collection of oceanographic data was provided by the National Oceanic and Atmospheric
758 Administration, Center for Sponsored Coastal Ocean Research Award Number
759 NA09NOS4780204 to NNR.
760
761 **Acknowledgements**
762 The authors thank the crew of the R/V *Pelican*, Dr. Allana Welsh and Dr. Mostafa Elshahed for
763 providing fasta files for *nrfA* and WS3 phylogenetic comparisons, and Dr. Elshahed for helpful
764 comments regarding WS3 phylogeny. Portions of this research were conducted with high
765 performance computing resources provided by Louisiana State University
766 (http://www.hpc.lsu.edu).
767
768 **Author contributions**
769 JCT and OUM designed the study. LEG, NNR, and JCT collected samples. NNR provided
770 processed oceanographic data. LEG and OUM extracted, quantified, and determined quality of
771 nucleic acids. JCT, KWS, and BJB reconstructed the genomes. KWS, BJB, BT, BH, and JCT
772 conducted downstream analyses. JCT led manuscript writing and all co-authors evaluated and
773 contributed edits.
774
775 **Competing financial interests**
776 The authors declare no competing financial interests.
777

778 **Table 1.** Genome characteristics for the 20 bins associated with uncultivated lineages.
779

| IMG Genome ID | Bin Id | Taxonomy | Compl % | Contam % | Strain het. | Scaff. | Longest scaff. (bp) | Size (bp) | Genes | GC (fract.) | Coding density (fract.) | Estim. Compl. Size (Mbp) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **2651870035** | **43-1** | Chloroflexi (SAR202) | 90.3 | 0 | 0 | 69 | 161242 | 1972793 | 1882 | 0.52 | 0.88 | 2.2 |
| **2651870036** | **43-2** | Chloroflexi (SAR202) | 88.6 | 4.1 | 15.4 | 134 | 157345 | 2402386 | 2373 | 0.52 | 0.91 | 2.7 |
| **2651870034** | **43** | Chloroflexi (SAR202) | 83.2 | 0.1 | 100 | 179 | 90503 | 2475308 | 2392 | 0.53 | 0.89 | 3.0 |
| **2693429801** | **45** | Marinimicrobia (SAR406) –B | 89.8 | 5.5 | 71.4 | 124 | 105140 | 2811623 | 2487 | 0.47 | 0.93 | 3.1 |
| **2651870052** | **45-1** | Marinimicrobia (SAR406) –B | 85.2 | 0.1 | 100 | 144 | 100582 | 2410233 | 2235 | 0.49 | 0.93 | 2.8 |
| **2693429802** | **45-2** | Marinimicrobia (SAR406) –B | 79.3 | 1.8 | 12.5 | 287 | 61408 | 2811444 | 2554 | 0.46 | 0.94 | 3.5 |
| **2651870053** | **51-1** | Marinimicrobia (SAR406) –A | 73.6 | 1.7 | 50 | 75 | 191525 | 1901306 | 1835 | 0.39 | 0.95 | 2.6 |
| **2651870051** | **51** | Marinimicrobia (SAR406) –A | 21.3 | 0 | 0 | 115 | 16923 | 578802 | 686 | 0.41 | 0.95 | 2.7 |
| **2651870038** | **15** | Euryarchaeota (MGII) | 83.2 | 1.6 | 0 | 43 | 255599 | 1885130 | 1614 | 0.62 | 0.95 | 2.3 |
| **2651870039** | **17** | Euryarchaeota (MGII) | 81.9 | 0.1 | 50 | 88 | 137319 | 1803861 | 1564 | 0.43 | 0.96 | 2.2 |
| **2651870037** | **14** | Euryarchaeota (MGII) | 71.2 | 0.8 | 100 | 80 | 90069 | 1389909 | 1305 | 0.54 | 0.94 | 2.0 |
| **2651870040** | **18** | Euryarchaeota (MGII) | 61.1 | 0.8 | 100 | 155 | 20633 | 1033226 | 1025 | 0.50 | 0.96 | 1.7 |
| **2651870042** | **38** | Euryarchaeota (MGII) | 27.1 | 0 | 0 | 138 | 12893 | 615290 | 610 | 0.55 | 0.96 | 2.3 |
| **2651870041** | **17-1** | Euryarchaeota (MGII) | 16.6 | 2.8 | 33.3 | 123 | 16952 | 538052 | 566 | 0.41 | 0.95 | 3.2 |
| **2693429807** | **13** | Unclassified (ACD39) | 89.8 | 5.1 | 20 | 401 | 75104 | 4269849 | 3686 | 0.47 | 0.93 | 4.8 |
| **2693429799** | **40** | Parcubacteria (OD1) | 71.9 | 15.1 | 75 | 97 | 65104 | 1086283 | 1208 | 0.52 | 0.92 | 1.5 |
| **2693429797** | **16** | Peregrinibacteria | 83.2 | 0.3 | 100 | 67 | 59308 | 1384712 | 1318 | 0.39 | 0.93 | 1.7 |
| **2693429798** | **39** | Peregrinibacteria | 49.9 | 0.3 | 100 | 131 | 18806 | 747520 | 809 | 0.45 | 0.95 | 1.5 |
| **2693429804** | **50** | Unclassified (PAUC34f) | 84.8 | 1.4 | 0 | 455 | 76269 | 5346994 | 5484 | 0.58 | 0.92 | 6.3 |
| **2693429803** | **48** | Unclassified (PAUC34f) | 51.9 | 2.2 | 0 | 470 | 20176 | 2566149 | 2596 | 0.55 | 0.94 | 4.9 |

780
781

782 **Figure 1.** Metabolic reconstruction of Marine Group II Euryarchaeota, SAR406, and SAR202,
783 based on the top three or four most complete genomes. Colors indicate pathway elements based
784 on the number of genomes in which they were recovered, according to the key. Black outlines
785 and/or arrows indicate genes that were not observed. Numbers correspond to annotations
786 supplied in Table S2.
787
788 **Figure 2.** Metabolic reconstruction of the Candidate Phylum members PAUC34f, Parcubacteria
789 (OD1), Peregrinibacteria, and ACD39 (Bin 13). Colors indicate pathway elements based on the
790 number of genomes in which they were recovered, according to the key. Black outlines and/or
791 arrows indicate genes that were not observed. Numbers correspond to annotations supplied in
792 Table S2.
793
794 **Figure 3.** Relative DNA to RNA recruitment rank for each genome, by sample. Colors indicate
795 the relative difference in the ratio of rank based on total RNA and DNA mapping. Red indicates
796 a higher RNA recruitment rank compared to DNA recruitment rank, and vice-versa for blue. +
797 and - symbols indicate bins where the rank-residual from the identity in RNA vs. DNA read
798 mapping was more or less than one standard deviation beyond 0, respectively. Dendrograms
799 were calculated using an Unweighted Pair Group Method with Arithmetic Mean (UPGMA) from
800 Euclidian distances of rank residuals across all samples and bins.
801
802 **Figure 4.** Expression of predicted respiratory genes. RPKM values of RNA recruitment for each
803 gene, by sample, are depicted with colors according to the key (yellow to blue follows increasing
804 intensity). Genes are grouped by bin, taxonomic affiliation, and specific respiratory process.
805 DNRA- dissimilatory nitrite reduction to ammonia.
806
807 **Figure 5.** Expression of predicted CAZy genes. RPKM values or RNA recruitment for each
808 gene, by sample, are depicted with colors according to the key (yellow to red follows increasing
809 intensity). Genes are grouped by bin, taxonomic affiliation, and general CAZy categories. CE-
810 carbohydrate esterase; GH- glycoside hydrolase; GT- glycosyltransferase; CBM- carbohydrate
811 binding module.
812
813
814

Supplemental Information

**Table S1.** Spreadsheet (Table_S1.xlsx) containing information on taxonomy, CheckM results, IMG statistics, partial metabolic reconstruction, transporter classifications, CAZy predictions, sample chemical data, RPKM values and gene neighborhoods for WS3 cytochrome c oxidases and *nrfA* genes from SAR406 and WS3.

**Table S2.** Gene annotations associated with Figures 1 and 2. Numbers correspond to those in the figures.

**Figure S1.** Maximum likelihood tree of concatenated ribosomal protein coding genes. Values at internal nodes indicate bootstrap support (n=100). Scale bar indicates changes per position.

**Figure S2.** Phylogenetic placement of bins based on CheckM.

**Figure S3.** Annotations of protein-coding gene sequence best blastp hits in the nr database, divided into quartiles by bit score, for CP bacteria.

**Figure S4.** SAR406 16S rRNA gene phylogeny. Genes recovered from assembled bins have 45* or 51* designations. Values at nodes indicate Shimodaira-Hasegawa "like" values (91). Scale bar indicates changes per position.

**Figure S5.** SAR202 16S rRNA gene phylogeny. Genes recovered from assembled bins are indicated as 43-*. Subclades are designated according to Morris et al. 2004, and the tree is rooted according to Figure 1 in that publication. Values at nodes indicate Shimodaira-Hasegawa "like" values (91). Scale bar indicates changes per position.

**Figure S6.** 16S rRNA gene phylogeny of the WS3 clade (70) with added PAUC34f sequences from the GreenGenes database and the Bin 50 sequence. Tree is rooted on the Archaea according to Fig. S1 in Farag et al. 2017. Values at nodes indicate Shimodaira-Hasegawa "like" values (91). Scale bar indicates changes per position.

**Figure S7.** Metagenomic RPKM values for each group, comprised of aggregated values for each bin within the group. Values are plotted according to sample and colored according to the dissolved oxygen (DO) concentration from where the sample was taken.

**Figure S8.** Phylogenetic tree of predicted *nrfA* genes in SAR406 and WS3. Additional taxa are from Figure 3 in Welch et al., 2014. The tree was rooted at the midpoint. Values at nodes indicate Shimodaira-Hasegawa "like" values (91). Scale bar indicates changes per position.

855 **Figure S9.** Conserved catalytic motifs within the *nrfA* gene. A- black square surrounds the first
856 heme-binding CXXCK/CXXCH motif. B- black square surrounds the catalytic KXQH/KXRH
857 motif. The alignment follows highlighting found in Welsh *et al.*, 2014 (48). All genes numbers
858 from this study are indicated as 26536*, corresponding to rows 64 and 66-68.
859
860 **Additional Supplemental Information** such as scripts, workflows, and key files, including fasta
861 files for each tree, are provided as a link hosted at the Thrash Lab website:
862 http://thethrashlab.com/publications.
863
864

## References Cited


1. Rabalais NN, Turner RE, Dortch Q, Justic D, Bierman VJ, Wiseman Jr WJ. 2002. Nutrient-enhanced productivity in the northern Gulf of Mexico: past, present and future. Hydrobiologia 475:39-63.
2. Diaz RJ, Rosenberg R. 2008. Spreading Dead Zones and Consequences for Marine Ecosystems. Science 321:926-929.
3. Action Plan 2008 for Reducing Mitigating, And Controlling Hypoxia in the Northern Gulf of Mexico and Improving Water Quality in the Mississippi River Basin. Mississippi River Gulf of Mexico Watershed Nutrient Task Force, U.S. Environmental Protection Agency, Office of Wetlands, Oceans, and Watersheds: Washington, D.C.,
4. Ulloa O, Canfield DE, DeLong EF, Letelier RM, Stewart FJ. 2012. Microbial oceanography of anoxic oxygen minimum zones. Proceedings of the National Academy of Sciences 109:15996-16003.
5. Wright JJ, Konwar KM, Hallam SJ. 2012. Microbial ecology of expanding oxygen minimum zones. Nature Reviews Microbiology 10:381-394.
6. Grote J, Jost G, Labrenz M, Herndl GJ, Jurgens K. 2008. Epsilonproteobacteria Represent the Major Portion of Chemoautotrophic Bacteria in Sulfidic Waters of Pelagic Redoxclines of the Baltic and Black Seas. Applied and Environmental Microbiology 74:7546-7551.
7. Glaubitz S, Kiesslich K, Meeske C, Labrenz M, Jurgens K. 2013. SUP05 Dominates the Gammaproteobacterial Sulfur Oxidizer Assemblages in Pelagic Redoxclines of the Central Baltic and Black Seas. Applied and Environmental Microbiology 79:2767-2776.
8. Friedrich J, Janssen F, Aleynik D, Bange HW, Boltacheva N, Çagatay MN, Dale AW, Etiope G, Erdem Z, Geraga M, Gilli A, Gomoiu MT, Hall POJ, Hansson D, He Y, Holtappels M, Kirf MK, Kononets M, Konovalov S, Lichtschlag A, Livingstone DM, Marinaro G, Mazlumyan S, Naeher S, North RP, Papatheodorou G, Pfannkuche O, Prien R, Rehder G, Schubert CJ, Soltwedel T, Sommer S, Stahl H, Stanev EV, Teaca A, Tengberg A, Waldmann C, Wehrli B, Wenzhöfer F. 2014. Investigating hypoxia in aquatic environments: diverse approaches to addressing a complex phenomenon. Biogeosciences 11:1215-1259.
9. Lam P, Kuypers M. 2011. Microbial Nitrogen Cycling Processes in Oxygen Minimum Zones. Annual review of marine science 3:317-345.
10. Beman JM, Popp BN, Alford SE. 2012. Quantification of ammonia oxidation rates and ammonia-oxidizing archaea and bacteria at high resolution in the Gulf of California and eastern tropical North Pacific Ocean. Limnology and Oceanography 57:711-726.
11. Saad EM, Longo AF, Chambers LR, Huang R, Benitez‐Nelson C, Dyhrman ST, Diaz JM, Tang Y, Ingall ED. 2016. Understanding marine dissolved organic matter production: Compositional insights from axenic cultures of Thalassiosira pseudonana. Limnology and Oceanography AOP. doi: 10.1002/lno.10367.
12. Canfield DE, Stewart FJ, Thamdrup B, De Brabandere L, Dalsgaard T, DeLong EF, Revsbech NP, Ulloa O. 2010. A Cryptic Sulfur Cycle in Oxygen-Minimum-Zone Waters off the Chilean Coast. Science 330:1375-1378.
13. Stewart FJ, Ulloa O, DeLong EF. 2012. Microbial metatranscriptomics in a permanent marine oxygen minimum zone. Environmental Microbiology 14:23-40.

14. Roberts BJ, Doty SM. 2015. Spatial and Temporal Patterns of Benthic Respiration and Net Nutrient Fluxes in the Atchafalaya River Delta Estuary. Estuaries and Coasts 38:1918-1936.
15. McCarthy MJ, Carini SA, Liu Z, Ostrom NE, Gardner WS. 2013. Oxygen consumption in the water column and sediments of the northern Gulf of Mexico hypoxic zone. Estuarine, Coastal and Shelf Science 123:46-53.
16. Schaeffer BA, Sinclair GA, Lehrter JC, Murrell MC, Kurtz JC, Gould RW, Yates DF. 2011. An analysis of diffuse light attenuation in the northern Gulf of Mexico hypoxic zone using the SeaWiFS satellite data record. Remote Sensing of Environment 115:3748-3757.
17. Rabalais NN, Turner RE, Gupta BKS, Boesch DF, Chapman P, Murrell MC. 2007. Hypoxia in the northern Gulf of Mexico: Does the science support the Plan to Reduce, Mitigate, and Control Hypoxia? Estuaries and Coasts 30:753-772.
18. Rabalais NN, Turner RE, Wiseman Jr WJ. 2001. Hypoxia in the Gulf of Mexico. Journal of Environmental Quality 30:320-329.
19. Rabalais NN, Turner RE, Wiseman JWJ. 2002. Gulf of Mexico hypoxia, AKA "The dead zone". Annual Review of Ecology and Systematics 33:235-63.
20. Wiseman Jr WJ, Rabalais NN, Turner RE, Dinnel SP, MacNaughton A. 1997. Seasonal and interannual variability within the Louisiana coastal current: stratification and hypoxia. Journal of Marine Systems 12:237-248.
21. Gillies LE, Thrash JC, deRada S, Rabalais NN, Mason OU. 2015. Archaeal enrichment in the hypoxic zone in the northern Gulf of Mexico. Environmental Microbiology 17:3847.
22. Bristow LA, Sarode N, Cartee J, Caro-Quintero A, Thamdrup B, Stewart FJ. 2015. Biogeochemical and metagenomic analysis of nitrite accumulation in the Gulf of Mexico hypoxic zone. Limnology and Oceanography 60:1733-1750.
23. Cho J-C, Vergin KL, Morris RM, Giovannoni SJ. 2004. Lentisphaera araneosa gen. nov., sp. nov, a transparent exopolymer producing marine bacterium, and the description of a novel bacterial phylum, Lentisphaerae. Environmental Microbiology 6:611-621.
24. Landry Z, Swan BK, Herndl GJ, Stepanauskas R, Giovannoni SJ. 2017. SAR202 Genomes from the Dark Ocean Predict Pathways for the Oxidation of Recalcitrant Dissolved Organic Matter. mBio 8:17.
25. Allers E, Wright JJ, Konwar KM, Howes CG, Beneze E, Hallam SJ, Sullivan MB. 2012. Diversity and population structure of Marine Group A bacteria in the Northeast subarctic Pacific Ocean. The ISME Journal 7:256-268.
26. Wright JJ, Mewis K, Hanson NW, Konwar KM, Maas KR, Hallam SJ. 2014. Genomic properties of Marine Group A bacteria indicate a role in the marine sulfur cycle. The ISME Journal 8:455-468.
27. Gordon DA, Giovannoni SJ. 1996. Detection of stratified microbial populations related to Chlorobium and Fibrobacter species in the Atlantic and Pacific oceans. Applied and Environmental Microbiology 62.
28. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu W-T, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T. 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature 499:431-437.

955 29. Parada AE, Needham DM, Fuhrman JA. 2015. Every base matters: assessing small
956 subunit rRNA primers for marine microbiomes with mock communities, time series and
957 global field samples. Environmental Microbiology 18:1403-1414.
958 30. Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, Wilkins MJ, Wrighton
959 KC, Williams KH, Banfield JF. 2015. Unusual biology across a group comprising more
960 than 15% of domain Bacteria. Nature 523:208-211.
961 31. Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, Verberkmoes NC, Wilkins
962 MJ, Hettich RL, Lipton MS, Williams KH, Long PE, Banfield JF. 2012. Fermentation,
963 Hydrogen, and Sulfur Metabolism in Multiple Uncultivated Bacterial Phyla. Science
964 337:1661-1665.
965 32. Hentschel U, Hopke J, Horn M, Friedrich AB, Wagner M, Hacker J, Moore BS. 2002.
966 Molecular evidence for a uniform microbial community in sponges from different oceans.
967 Applied and Environmental Microbiology 68:4431-4440.
968 33. Wrighton KC, Castelle CJ, Wilkins MJ, Hug LA, Sharon I, Thomas BC, Handley KM,
969 Mullin SW, Nicora CD, Singh A, Lipton MS, Long PE, Williams KH, Banfield JF. 2014.
970 Metabolic interdependencies between phylogenetically novel fermenters and respiratory
971 organisms in an unconfined aquifer. The ISME Journal 8:1452-1463.
972 34. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN,
973 Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM,
974 Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. Nature
975 Microbiology 1.
976 35. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM:
977 assessing the quality of microbial genomes recovered from isolates, single cells, and
978 metagenomes. Genome Research 25:1043-1055.
979 36. Konstantinidis KT, Tiedje JM. 2007. Prokaryotic taxonomy and phylogeny in the
980 genomic era: advancements and challenges ahead. Current Opinion in Microbiology
981 10:504-509.
982 37. Konstantinidis KT, Tiedje JM. 2005. Towards a genome-based taxonomy for
983 prokaryotes. Journal of bacteriology 187:6258-6264.
984 38. Li M, Baker BJ, Anantharaman K, Jain S, Breier JA, Dick GJ. 2015. Genomic and
985 transcriptomic evidence for scavenging of diverse organic compounds by widespread
986 deep-sea archaea. Nature Communications 6:8933. DOI: 10.1038/ncomms9933.
987 39. Zhang CL, Xie W, Martin-Cuadrado A-B, Rodriguez-Valera F. 2015. Marine Group II
988 Archaea, potentially important players in the global ocean carbon cycle. Frontiers in
989 Microbiology 6:1108.
990 40. Orsi WD, Smith JM, Liu S, Liu Z, Sakamoto CM, Wilken S, Poirier C, Richards TA,
991 Keeling PJ, Worden AZ, Santoro AE. 2016. Diverse, uncultivated bacteria and archaea
992 underlying the cycling of dissolved protein in the ocean. The ISME Journal 10:2158-
993 2173.
994 41. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The
995 Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics.
996 Nucleic Acids Research 37:1–6. doi:10.1093/nar/gkn663.
997 42. Fuhrman JA, McCallum K, Davis AA. 1993. Phylogenetic diversity of subsurface marine
998 microbial communities from the Atlantic and Pacific Oceans. Applied and Environmental
999 Microbiology 59:1294-1302.

1000 43. Schattenhofer M, Fuchs BM, Amann R, Zubkov MV, Tarran GA, Pernthaler J. 2009.
1001 Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean.
1002 Environmental Microbiology 11:2078-2093.
1003 44. Hu P, Tom L, Singh A, Thomas BC, Baker BJ, Piceno YM, Andersen GL, Banfield JF.
1004 2016. Genome-Resolved Metagenomic Analysis Reveals Roles for Candidate Phyla and
1005 Other Microbial Community Members in Biogeochemical Transformations in Oil
1006 Reservoirs. mBio 7:15.
1007 45. DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N-UU, Martinez A,
1008 Sullivan MB, Edwards R, Brito BR, Chisholm SW, Karl DM. 2006. Community
1009 genomics among stratified microbial assemblages in the ocean's interior. Science (New
1010 York, NY) 311:496-503.
1011 46. Fuchs BM, Woebken D, Zubkov MV, Burkill P, Amann R. 2005. Molecular
1012 identification of picoplankton populations in contrasting waters of the Arabian Sea.
1013 Aquatic Microbial Ecology 39:145-157.
1014 47. Pitcher RS, Watmough NJ. 2004. The bacterial cytochrome cbb3 oxidases. Biochimica et
1015 biophysica acta 1655:388-399.
1016 48. Welsh A, Chee-Sanford JC, Connor LM, Löffler FE, Sanford RA. 2014. Refined NrfA
1017 Phylogeny Improves PCR-Based nrfA Gene Detection. Applied and Environmental
1018 Microbiology 80:2110-2119.
1019 49. Burns JL, DiChristina TJ. 2009. Anaerobic respiration of elemental sulfur and thiosulfate
1020 by Shewanella oneidensis MR-1 requires psrA, a homolog of the phsA gene of
1021 Salmonella enterica serovar typhimurium LT2. Applied and environmental microbiology
1022 75:5209-5217.
1023 50. Chen J, Hanke A, Tegetmeyer HE, Kattelmann I, Sharma R, Hamann E, Hargesheimer T,
1024 Kraft B, Lenk S, Geelhoed JS, Hettich RL, Strous M. 2017. Impacts of chemical
1025 gradients on microbial community structure. The ISME Journal 11:920-931.
1026 51. Sharpton TJ, Jospin G, Wu D, Langille MGI, Pollard KS, Eisen JA. 2012. Sifting through
1027 genomes with iterative-sequence clustering produces a large, phylogenetically diverse
1028 protein-family resource. BMC Bioinformatics 13:264.
1029 52. McBride MJ, Zhu Y. 2013. Gliding Motility and Por Secretion System Genes Are
1030 Widespread among Members of the Phylum Bacteroidetes. Journal of Bacteriology
1031 195:270-278.
1032 53. Varela MM, Aken HM, Herndl GJ. 2008. Abundance and activity of Chloroflexi‐type
1033 SAR202 bacterioplankton in the meso‐ and bathypelagic waters of the (sub)tropical
1034 Atlantic. Environmental Microbiology 10:1903-1911.
1035 54. Treusch AH, Vergin KL, Finlay LA, Donatz MG, Burton RM, Carlson CA, Giovannoni
1036 SJ. 2009. Seasonality and vertical structure of microbial communities in an ocean gyre.
1037 The ISME Journal 3:1148-1163.
1038 55. Law CJ, Maloney PC, Wang D-N. 2008. Ins and Outs of Major Facilitator Superfamily
1039 Antiporters. Annual Review of Microbiology 62:289-305.
1040 56. Lombard V, Bernard T, Rancurel C, Brumer H, Coutinho PM, Henrissat B. 2010. A
1041 hierarchical classification of polysaccharide lyases for glycogenomics. Biochemical
1042 Journal 432:437-444.
1043 57. Cregut M, Piutti S, Slezack-Deschaumes S, Benizri E. 2013. Compartmentalization and
1044 regulation of arylsulfatase activities in Streptomyces sp., Microbacterium sp. and

1045 Rhodococcus sp. soil isolates in response to inorganic sulfate limitation. Microbiological
1046 Research 168:12-21.
1047 58. Cregut M, Durand M-J, Thouand G. 2014. The Diversity and Functions of Choline
1048 Sulphatases in Microorganisms. Microbial Ecology 67:350-357.
1049 59. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, Thomas BC,
1050 Singh A, Wilkins MJ, Karaoz U, Brodie EL, Williams KH, Hubbard SS, Banfield JF.
1051 2016. Thousands of microbial genomes shed light on interconnected biogeochemical
1052 processes in an aquifer system. Nature Communications 7:13219.
1053 60. Luef B, Frischkorn KR, Wrighton KC, Holman H-YN, Birarda G, Thomas BC, Singh A,
1054 Williams KH, Siegerist CE, Tringe SG, Downing KH, Comolli LR, Banfield JF. 2015.
1055 Diverse uncultivated ultra-small bacterial cells in groundwater. Nature Communications
1056 6:6372.
1057 61. Alewell C, Paul S, Lischeid G, Storck FR. 2008. Co-regulation of redox processes in
1058 freshwater wetlands as a function of organic matter availability? Science of The Total
1059 Environment 404:335-342.
1060 62. Ouverney CC, Fuhrman JA. 2000. Marine planktonic archaea take up amino acids.
1061 Applied and environmental microbiology 66:4829-4833.
1062 63. Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. 2012.
1063 Untangling genomes from metagenomes: revealing an uncultured class of marine
1064 Euryarchaeota. Science 335:587-590.
1065 64. Weigel BL, Erwin PM. 2017. Effects of reciprocal transplantation on the microbiome and
1066 putative nitrogen cycling functions of the intertidal sponge, Hymeniacidon heliophila.
1067 Scientific Reports 7:43247.
1068 65. Jeong I-H, Kim K-H, Park J-S. 2013. Analysis of bacterial diversity in sponges collected
1069 off Chujado, an Island in Korea, using barcoded 454 pyrosequencing: Analysis of a
1070 distinctive sponge group containing Chloroflexi. Journal of Microbiology 51:570-577.
1071 66. Cuvelier ML, Blake E, Mulheron R, McCarthy PJ, Blackwelder P, Thurber RL, Lopez
1072 JV. 2014. Two distinct microbial communities revealed in the sponge Cinachyrella.
1073 Frontiers in Microbiology 5:581.
1074 67. Erwin PM, Pineda M, Webster N, Turon X, López-Legentil S. 2013. Down under the
1075 tunic: bacterial biodiversity hotspots and widespread ammonia-oxidizing archaea in coral
1076 reef ascidians. The ISME Journal 8:575-588.
1077 68. Costa PS, Reis MP, Ávila MP, Leite LR, de Araújo FMG, Salim ACM, Oliveira G,
1078 Barbosa F, Chartone-Souza E, Nascimento AMA. 2015. Metagenome of a Microbial
1079 Community Inhabiting a Metal-Rich Tropical Stream Sediment. PLOS ONE
1080 10:e0119465.
1081 69. Youssef NH, Farag IF, Rinke C, Hallam SJ, Woyke T, Elshahed MS. 2015. In Silico
1082 Analysis of the Metabolic Potential and Niche Specialization of Candidate Phylum
1083 "Latescibacteria" (WS3). PLOS ONE 10:10.1371/journal.pone.0127499.
1084 70. Farag IF, Youssef NH, Elshahed MS. 2017. Global Distribution Patterns and Pangenomic
1085 Diversity of the Candidate Phylum “Latescibacteria” (WS3). Applied and Environmental
1086 Microbiology 83: e00521-17.
1087 71. Thrash JC, Cho JC, Vergin KL, Morris RM, Giovannoni SJ. 2010. Genome Sequence of
1088 Lentisphaera araneosa HTCC2155T, the Type Species of the Order Lentisphaerales in the
1089 Phylum Lentisphaerae. Journal of Bacteriology 192:2938-2939.

1090 72. Krüger M, Meyerdierks A, Glöckner FO, Amann R, Widdel F, Kube M, Reinhardt R,
1091 Kahnt J, Böcher R, Thauer RK, Shima S. 2003. A conspicuous nickel protein in
1092 microbial mats that oxidize methane anaerobically. Nature 426:878-881.
1093 73. Dalsgaard T, Stewart FJ, Thamdrup B, De Brabandere L, Revsbech NP, Ulloa O,
1094 Canfield DE, DeLong EF. 2014. Oxygen at Nanomolar Levels Reversibly Suppresses
1095 Process Rates and Gene Expression in Anammox and Denitrification in the Oxygen
1096 Minimum Zone off Northern Chile. mBio 5:e01966-14-e01966-14.
1097 74. Stolper DA, Revsbech NP, Canfield DE. 2010. Aerobic growth at nanomolar oxygen
1098 concentrations. Proceedings of the National Academy of Sciences 107:18755-18760.
1099 75. Eggleston EM, Lee DY, Owens MS, Cornwell JC, Crump BC, Hewson I. 2015. Key
1100 respiratory genes elucidate bacterial community respiration in a seasonally anoxic
1101 estuary. Environmental Microbiology 17:2306-2318.
1102 76. King GM, Smith CB, Tolar B, Hollibaugh JT. 2013. Analysis of composition and
1103 structure of coastal to mesopelagic bacterioplankton communities in the northern gulf of
1104 Mexico. Frontiers in Microbiology 3:doi: 10.3389/fmicb.2012.00438.
1105 77. Tolar BB, King GM, Hollibaugh JT. 2013. An analysis of thaumarchaeota populations
1106 from the northern gulf of Mexico. Frontiers in Microbiology 4:72.
1107 78. Mason OU, Hazen TC, Borglin S, Chain PSG, Dubinsky EA, Fortney JL, Han J, Holman
1108 H-YN, Hultman J, Lamendella R, Mackelprang R, Malfatti S, Tom LM, Tringe SG,
1109 Woyke T, Zhou J, Rubin EM, Jansson JK. 2012. Metagenome, metatranscriptome and
1110 single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. The
1111 ISME Journal doi:papers2://publication/doi/10.1038/ismej.2012.59:1-13.
1112 79. Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, Banfield JF.
1113 2009. Community-wide analysis of microbial genome sequence signatures. Genome
1114 Biology 10:R85.
1115 80. Sharon I, Morowitz MJ, Thomas BC, Costello EK, Relman DA, Banfield JF. 2013. Time
1116 series community genomics analysis reveals rapid shifts in bacterial species, strains, and
1117 phage during infant gut colonization. Genome Reserach 23:111-120.
1118 81. Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: a de novo assembler for
1119 single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics
1120 28:1420-1428.
1121 82. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler
1122 transform. Bioinformatics 25:1754-1760.
1123 83. Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P, Rubin EM. 2007. Genome-wide
1124 experimental determination of barriers to horizontal gene transfer. Science 318:1449-52.
1125 84. Darling AE, Jospin G, Lowe E, Matsen Iv FA, Bik HM, Eisen JA. 2014. PhyloSift:
1126 phylogenetic analysis of genomes and metagenomes. PeerJ 2:e243.
1127 85. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
1128 throughput. Nucleic Acids Research 32:1792-1797.
1129 86. Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a
1130 new software for selection of phylogenetic informative regions from multiple sequence
1131 alignments. BMC Evol Biol 10:210.
1132 87. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis
1133 of large phylogenies. Bioinformatics 30:1312-1313.
1134 88. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN,
1135 Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM,

1136 Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. Nature
1137 Microbiology 1:16048.
1138 89. Contreras-Moreira B, Vinuesa P. 2013. GET_HOMOLOGUES, a versatile software
1139 package for scalable and robust microbial pangenome analysis. Applied and
1140 Environmental Microbiology 79:7696-7701.
1141 90. Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use
1142 in phylogenetic analysis. Molecular Biology and Evolution 17:540-552.
1143 91. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood
1144 trees for large alignments. PLOS ONE 5:e9490.
1145 92. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL.
1146 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.
1147 93. Junier T, Zdobnov EM. 2010. The Newick utilities: high-throughput phylogenetic tree
1148 processing in the Unix shell. Bioinformatics 26:1669-1670.
1149 94. Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang
1150 J, Woyke T, Huntemann M, Anderson I, Billis K, Varghese N, Mavromatis K, Pati A,
1151 Ivanova NN, Kyrpides NC. 2014. IMG 4 version of the integrated microbial genomes
1152 comparative analysis system. Nucleic Acids Research 42:D560-7.
1153 95. Saier MH, Reddy VS, Tsu BV, Ahmed M, Li C, Moreno-Hagelsieb G. 2016. The
1154 Transporter Classification Database (TCDB): recent advances. Nucleic Acids Research
1155 44:D372-D379.
1156 96. Lombard V, Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-
1157 active enzymes database (CAZy) in 2013. Nucleic Acids Research 42:D490-D495.
1158 97. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and
1159 quantifying mammalian transcriptomes by RNA-Seq. Nature Methods 5:621-628.
1160 98. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing
1161 genomic features. Bioinformatics 26:841-842.
1162 99. Yachdav G, Wilzbach S, Rauscher B, Sheridan R, Sillitoe I, Procter J, Lewis SE, Rost B,
1163 Goldberg T. 2016. MSAViewer: interactive JavaScript visualization of multiple sequence
1164 alignments. Bioinformatics 32:3501-3503.
1165 100. Eddy SR. 2011. Accelerated Profile HMM Searches. PLOS Computational Biology
1166 7:e1002195.
1167

**Figure 1**

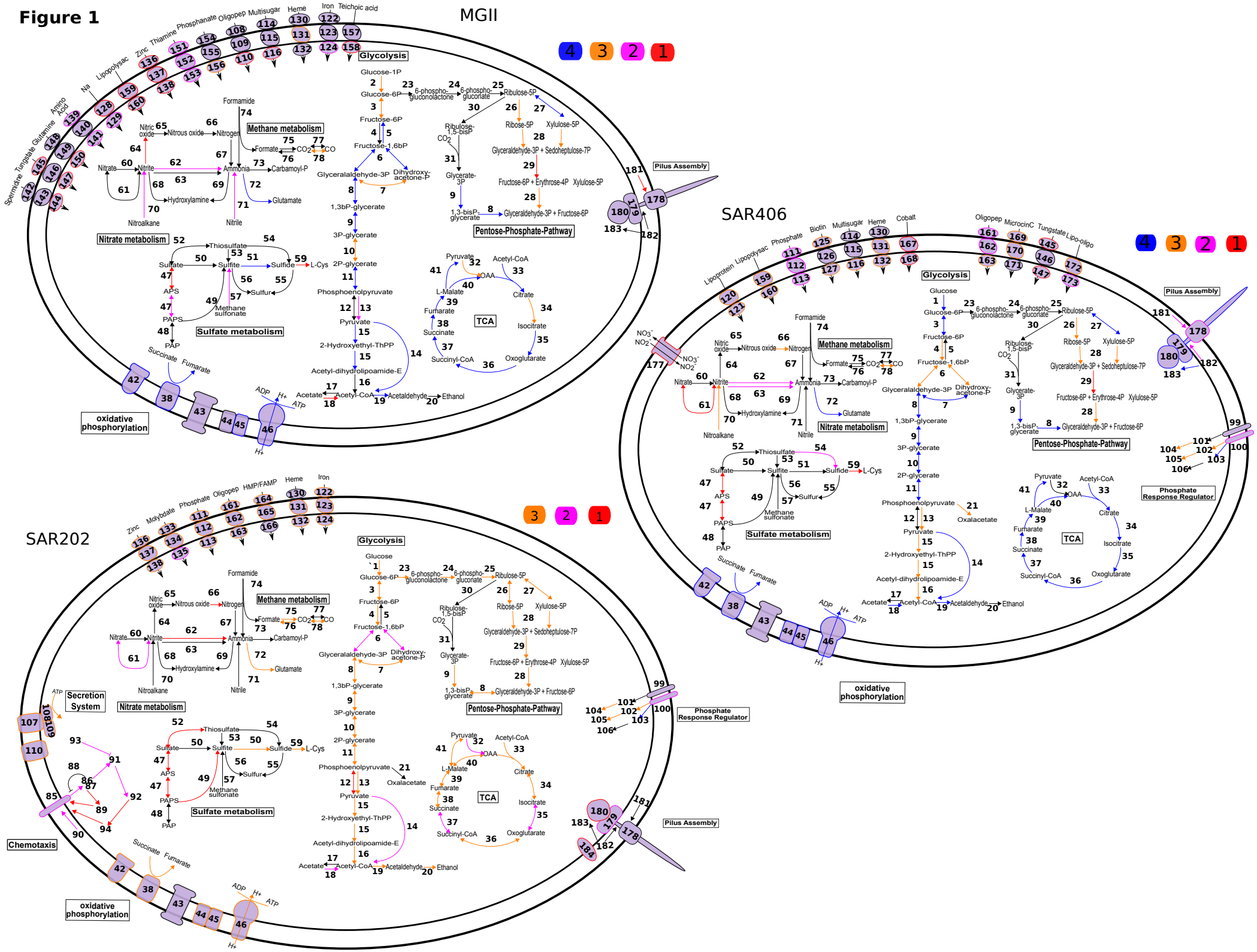
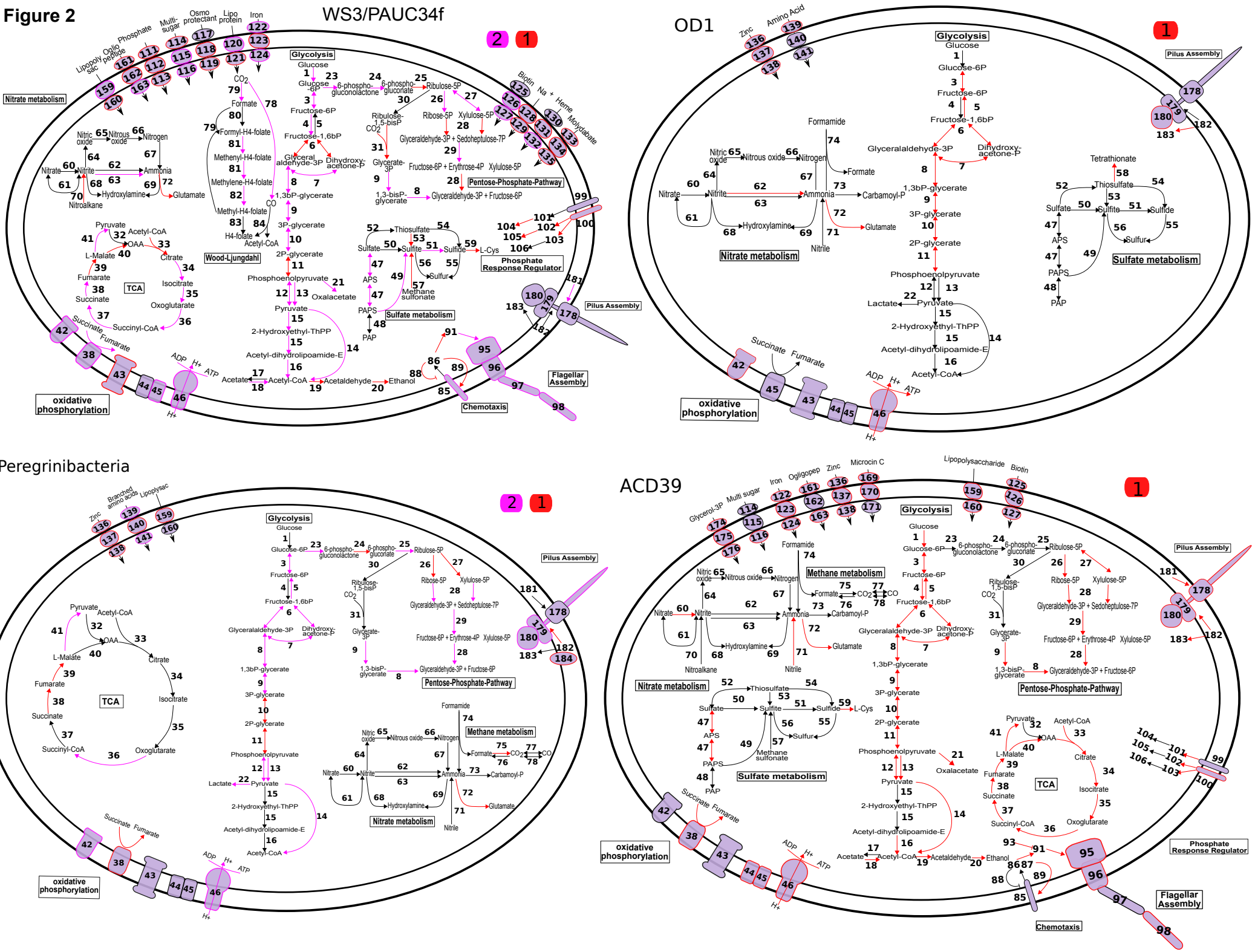
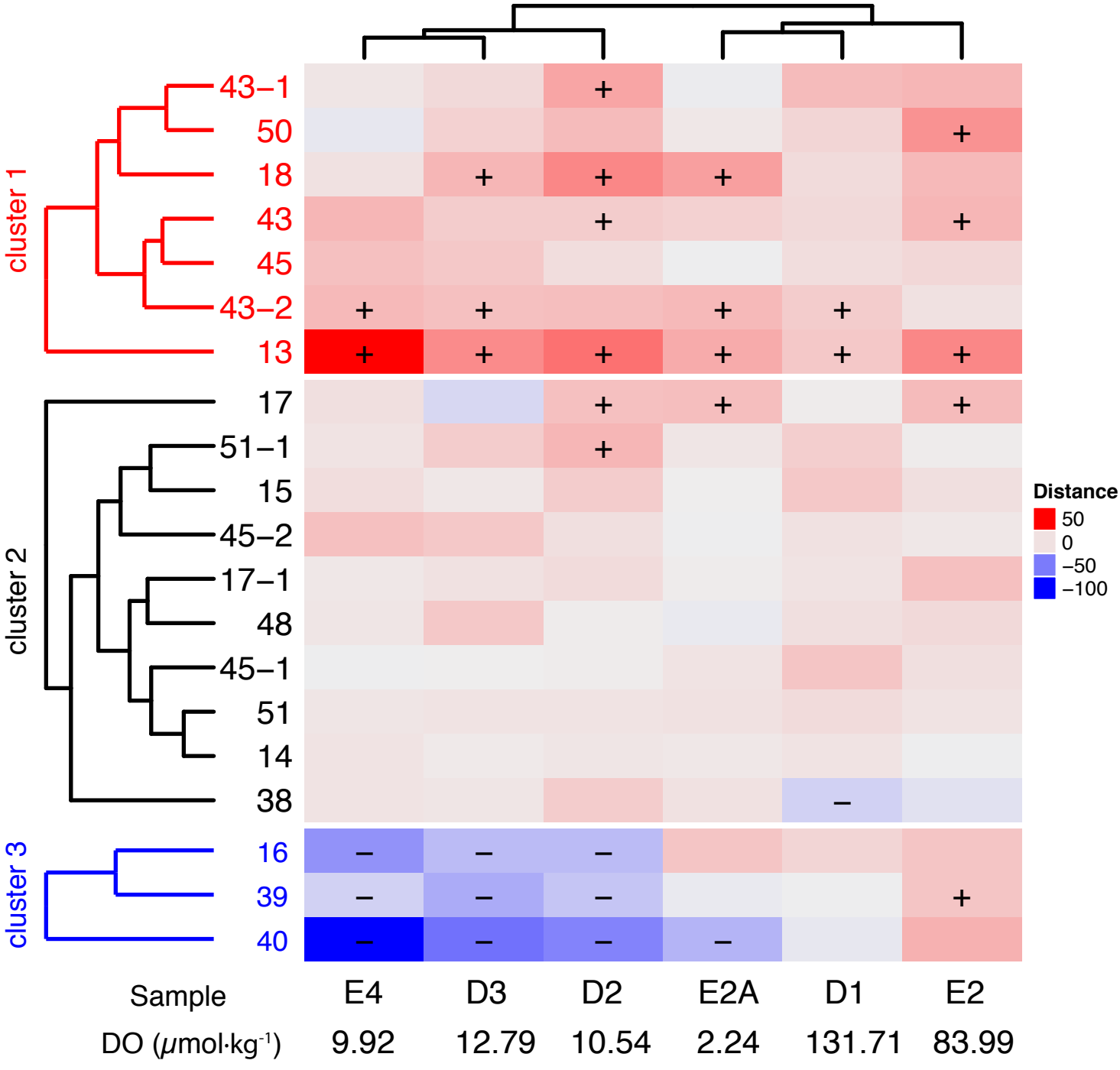
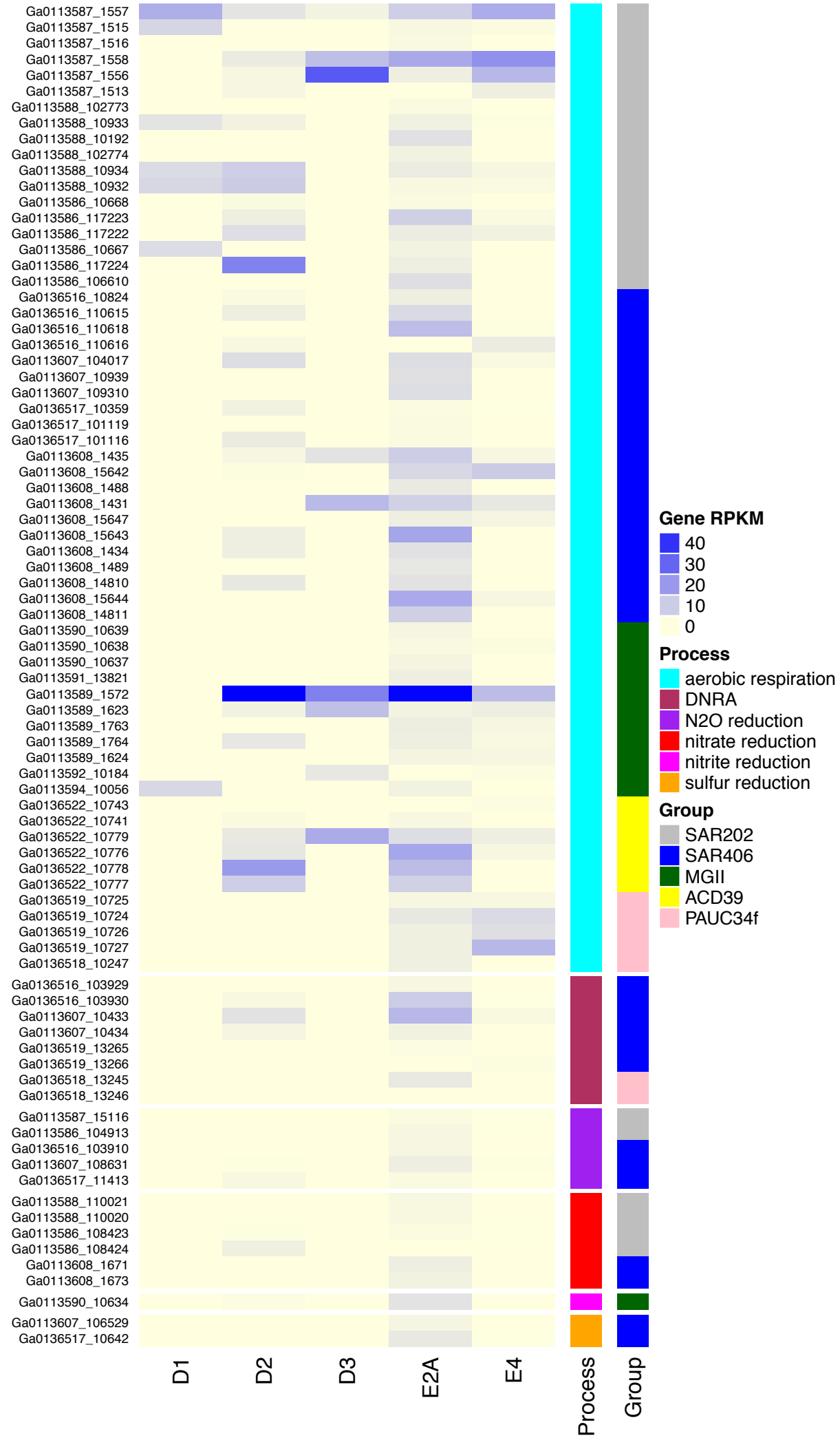
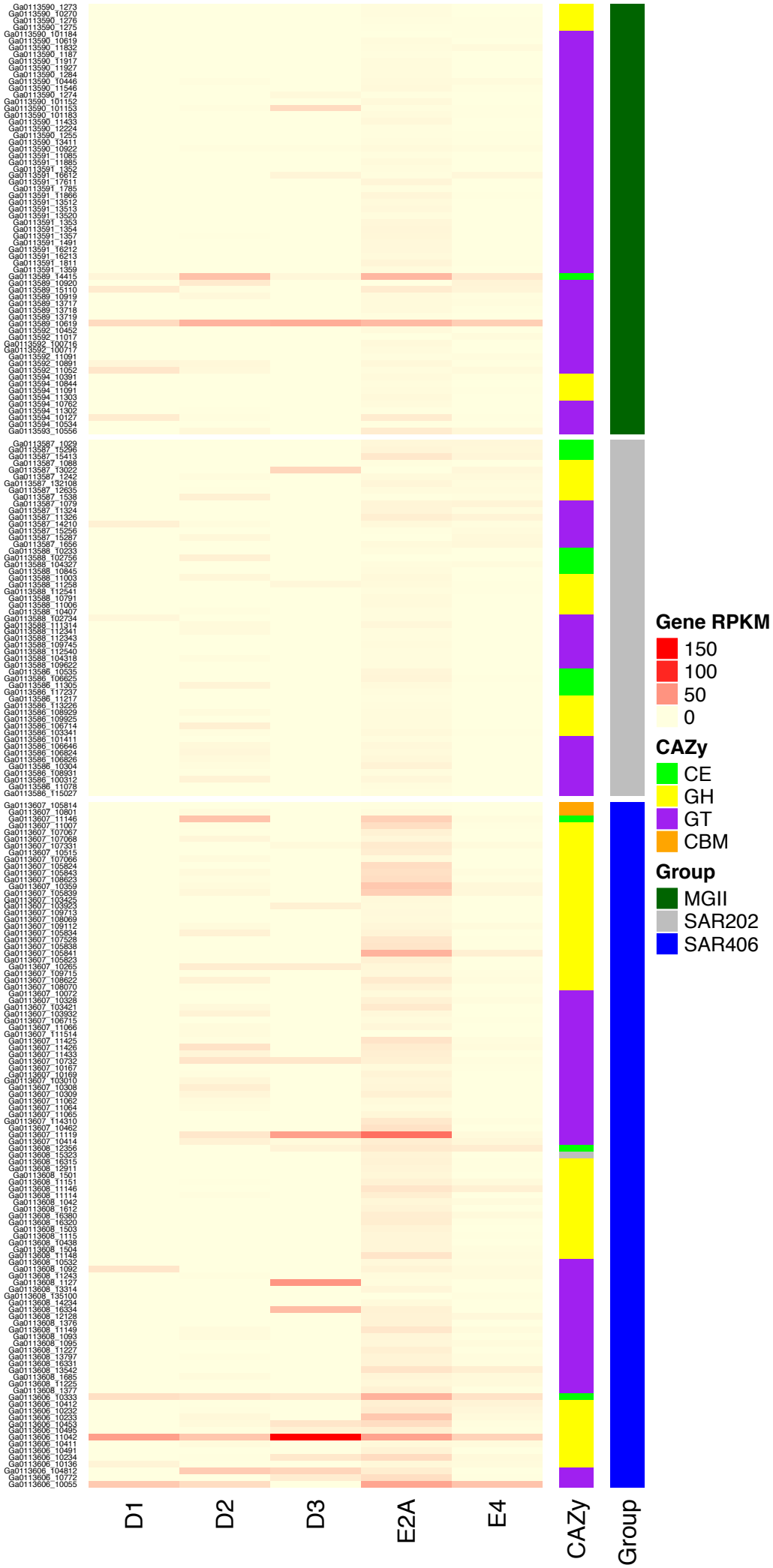

MGII
SAR406
SAR202
Glycolysis
Methane metabolism
Nitrate metabolism
Sulfate metabolism
Pentose-Phosphate-Pathway
TCA
Pilus Assembly
Phosphate Response Regulator
Secretion System
Chemotaxis
oxidative phosphorylation
Glucose-1P
Glucose-6P
6-phospho-gluconolactone
6-phospho-gluconate
Ribulose-5P
Ribose-5P
Xylulose-5P
Glyceraldehyde-3P + Sedoheptulose-7P
Fructose-6P + Erythrose-4P
Xylulose-5P
Glyceraldehyde-3P + Fructose-6P
Fructose-6P
Fructose-1,6bP
Glyceralaldehyde-3P
Dihydroxy-acetone-P
1,3bP-glycerate
3P-glycerate
2P-glycerate
Phosphoenolpyruvate
Pyruvate
2-Hydroxyethyl-ThPP
Acetyl-dihydrolipoamide-E
Acetate
Acetyl-CoA
Acetaldehyde
Ethanol
Formamide
Nitrogen
Nitrous oxide
Nitric oxide
Nitrate
Nitrite
Ammonia
Carbamoyl-P
Glutamate
Hydroxylamine
Nitroalkane
Nitrile
Thiosulfate
Sulfate
Sulfite
Sulfide
L-Cys
Sulfur
APS
PAPS
PAP
Methane sulfonate
Succinate
Fumarate
Succinyl-CoA
Oxoglutarate
Isocitrate
Citrate
OAA
L-Malate
Acetyl-CoA
Oxalacetate
Glucose
Cobalt
Lipo-oligo
Tungstate
MicrocinC
Biotin
Lipoprotein
Lipopolysac
Phosphate
Multisugar
Heme
Iron
Teichoic acid
Amino Acid
Glutamine
Tungstate
Spermidine
Na
Zinc
Thiamine
Phosphanate
Oligopep
Molybdate
HMP/FAMP
ADP
H+
ATP

**Figure 2**


WS3/PAUC34f
OD1
Peregrinibacteria
ACD39
Glycolysis
Pentose-Phosphate-Pathway
Wood-Ljungdahl
TCA
Nitrate metabolism
Sulfate metabolism
Phosphate Response Regulator
Pilus Assembly
Flagellar Assembly
Chemotaxis
oxidative phosphorylation
Methane metabolism
Glucose
Glucose-6P
Fructose-6P
Fructose-1,6bP
Glyceralaldehyde-3P
Dihydroxy-acetone-P
1,3bP-glycerate
3P-glycerate
2P-glycerate
Phosphoenolpyruvate
Pyruvate
2-Hydroxyethyl-ThPP
Acetyl-dihydrolipoamide-E
Acetyl-CoA
Acetate
Acetaldehyde
Ethanol
Formate
Formyl-H4-folate
Methenyl-H4-folate
Methylene-H4-folate
Methyl-H4-folate
H4-folate
Acetyl-CoA
Nitrate
Nitrite
Ammonia
Nitrogen
Nitrous oxide
Nitric oxide
Hydroxylamine
Glutamate
Nitroalkane
Nitrile
Formamide
Carbamoyl-P
Lactate
Succinate
Fumarate
Succinyl-CoA
Oxoglutarate
Isocitrate
Citrate
L-Malate
OAA
Pyruvate
Acetyl-CoA
Ribulose-5P
Ribose-5P
Xylulose-5P
Glyceraldehyde-3P + Sedoheptulose-7P
Fructose-6P + Erythrose-4P
Xylulose-5P
Glyceraldehyde-3P + Fructose-6P
Ribulose-1,5-bisP
Glycerate-3P
1,3-bisP-glycerate
Thiosulfate
Sulfate
Sulfite
Sulfide
L-Cys
Sulfur
Methane sulfonate
APS
PAPS
PAP
Tetrathionate
Lipopolysac
Oglio peptide
Phosphate
Multi-sugar
Osmo protectant
Lipo protein
Iron
Biotin
Na + Heme
Molydabate
Zinc
Amino Acid
Branched amino acids
Lipopolysaccharide
Microcin C
Glycerol-3P
Multi sugar
Iron
Ogligopep
Zinc
Microcin C
Lipopolysaccharide
Biotin
ADP
H+
ATP


cluster 1
cluster 2
cluster 3
43–1
50
18
43
45
43–2
13
17
51–1
15
45–2
17–1
48
45–1
51
14
38
16
39
40
Distance
50
0
−50
−100
Sample
E4
D3
D2
E2A
D1
E2
DO (μmol·kg⁻¹)
9.92
12.79
10.54
2.24
131.71
83.99


Ga0113587_1557
Ga0113587_1515
Ga0113587_1516
Ga0113587_1558
Ga0113587_1556
Ga0113587_1513
Ga0113588_102773
Ga0113588_10933
Ga0113588_10192
Ga0113588_102774
Ga0113588_10934
Ga0113588_10932
Ga0113586_10668
Ga0113586_117223
Ga0113586_117222
Ga0113586_10667
Ga0113586_117224
Ga0113586_106610
Ga0136516_10824
Ga0136516_110615
Ga0136516_110618
Ga0136516_110616
Ga0113607_104017
Ga0113607_10939
Ga0113607_109310
Ga0136517_10359
Ga0136517_101119
Ga0136517_101116
Ga0113608_1435
Ga0113608_15642
Ga0113608_1488
Ga0113608_1431
Ga0113608_15647
Ga0113608_15643
Ga0113608_1434
Ga0113608_1489
Ga0113608_14810
Ga0113608_15644
Ga0113608_14811
Ga0113590_10639
Ga0113590_10638
Ga0113590_10637
Ga0113591_13821
Ga0113589_1572
Ga0113589_1623
Ga0113589_1763
Ga0113589_1764
Ga0113589_1624
Ga0113592_10184
Ga0113594_10056
Ga0136522_10743
Ga0136522_10741
Ga0136522_10779
Ga0136522_10776
Ga0136522_10778
Ga0136522_10777
Ga0136519_10725
Ga0136519_10724
Ga0136519_10726
Ga0136519_10727
Ga0136518_10247
Ga0136516_103929
Ga0136516_103930
Ga0113607_10433
Ga0113607_10434
Ga0136519_13265
Ga0136519_13266
Ga0136518_13245
Ga0136518_13246
Ga0113587_15116
Ga0113586_104913
Ga0136516_103910
Ga0113607_108631
Ga0136517_11413
Ga0113588_110021
Ga0113588_110020
Ga0113586_108423
Ga0113586_108424
Ga0113608_1671
Ga0113608_1673
Ga0113590_10634
Ga0113607_106529
Ga0136517_10642
D1
D2
D3
E2A
E4
Process
Group
Gene RPKM
40
30
20
10
0
Process
aerobic respiration
DNRA
N2O reduction
nitrate reduction
nitrite reduction
sulfur reduction
Group
SAR202
SAR406
MGII
ACD39
PAUC34f


Gene RPKM
150
100
50
0
CAZy
CE
GH
GT
CBM
Group
MGII
SAR202
SAR406
D1
D2
D3
E2A
E4
CAZy
Group

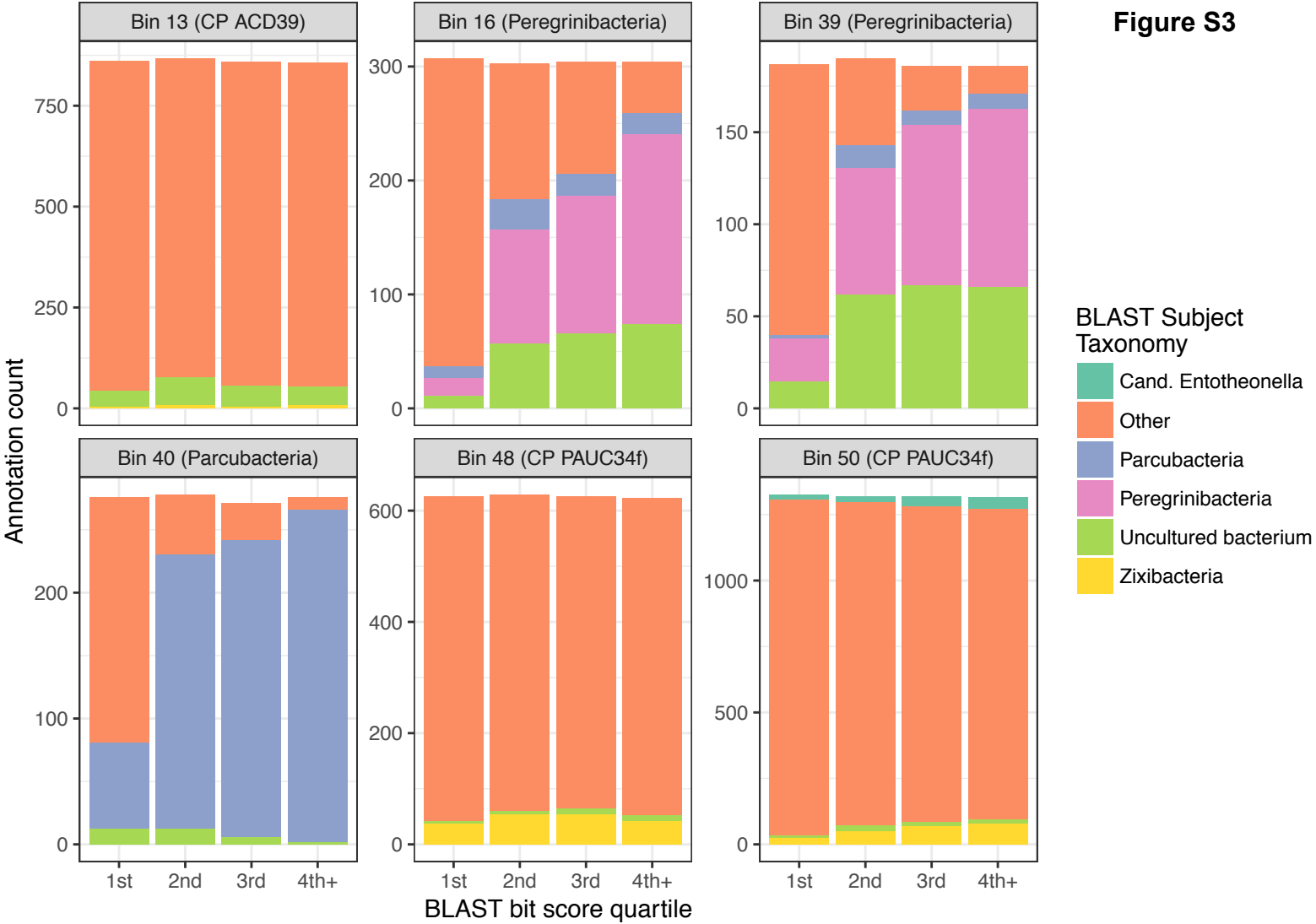


Figure S3
Bin 13 (CP ACD39)
Bin 16 (Peregrinibacteria)
Bin 39 (Peregrinibacteria)
Bin 40 (Parcubacteria)
Bin 48 (CP PAUC34f)
Bin 50 (CP PAUC34f)
Annotation count
BLAST bit score quartile
1st
2nd
3rd
4th+
BLAST Subject Taxonomy
Cand. Entotheonella
Other
Parcubacteria
Peregrinibacteria
Uncultured bacterium
Zixibacteria

**Figure S4**

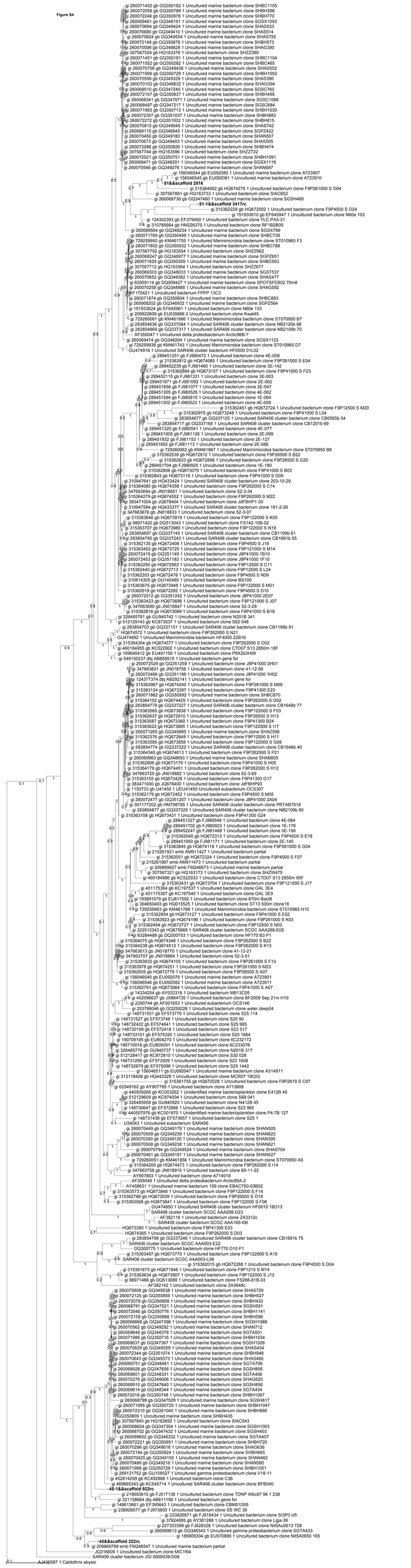

gi 260071452 gp GQ350182 1 Uncultured marine bacterium clone SHBC1105
gi 260072059 gp GQ350789 1 Uncultured marine bacterium clone SHBH396
gi 260072248 gp GQ350978 1 Uncultured marine bacterium clone SHBH770
gi 260069461 gp GQ348191 1 Uncultured marine bacterium clone SGSX1093
gi 260070694 gp GQ349424 1 Uncultured marine bacterium clone SHAS533
gi 260070680 gp GQ349410 1 Uncultured marine bacterium clone SHAS514
gi 260070824 gp GQ349554 1 Uncultured marine bacterium clone SHAS755
gi 260072148 gp GQ350878 1 Uncultured marine bacterium clone SHBH973
gi 260070098 gp GQ348828 1 Uncultured marine bacterium clone SHAG390
gi 307567524 gp HQ163376 1 Uncultured bacterium clone SHZZ390
gi 260071451 gp GQ350181 1 Uncultured marine bacterium clone SHBC1104
gi 260071552 gp GQ350282 1 Uncultured marine bacterium clone SHBC465
gi 260070706 gp GQ349436 1 Uncultured marine bacterium clone SHAS552
gi 260071999 gp GQ350729 1 Uncultured marine bacterium clone SHBH1052
gi 260070599 gp GQ349329 1 Uncultured marine bacterium clone SHAS390
gi 260070102 gp GQ348832 1 Uncultured marine bacterium clone SHAG394
gi 260068510 gp GQ347240 1 Uncultured marine bacterium clone SGSC760
gi 260072107 gp GQ350837 1 Uncultured marine bacterium clone SHBH495
gi 260068341 gp GQ347071 1 Uncultured marine bacterium clone SGSC1066
gi 260068487 gp GQ347217 1 Uncultured marine bacterium clone SGSC684
gi 260071983 gp GQ350713 1 Uncultured marine bacterium clone SHBH1030
gi 260072307 gp GQ351037 1 Uncultured marine bacterium clone SHBH883
gi 260072272 gp GQ351002 1 Uncultured marine bacterium clone SHBH815
gi 260070815 gp GQ349545 1 Uncultured marine bacterium clone SHAS742
gi 260068115 gp GQ346845 1 Uncultured marine bacterium clone SGFZ422
gi 260070459 gp GQ349180 1 Uncultured marine bacterium clone SHAN507
gi 260070673 gp GQ349403 1 Uncultured marine bacterium clone SHAS505
gi 260072096 gp GQ350826 1 Uncultured marine bacterium clone SBH474
gi 307567744 gp HQ163596 1 Uncultured bacterium clone SHZZ722
gi 260072021 gp GQ350751 1 Uncultured marine bacterium clone SHBH1091
gi 260069471 gp GQ348201 1 Uncultured marine bacterium clone SGSX1119
gi 260070546 gp GQ349276 1 Uncultured marine bacterium clone SHAN687
gi 156046544 gb EU092080 1 Uncultured marine bacterium clone A723907
gi 156046545 gb EU092081 1 Uncultured marine bacterium clone A723910
51-1&scaffold 2816
gi 315364002 gb HQ674275 1 Uncultured bacterium clone F9P261000 S O04
gi 307567881 gb HQ163733 1 Uncultured bacterium clone SIAC652
gi 260068730 gp GQ347460 1 Uncultured marine bacterium clone SGSH460
51-1&scaffold 3417rc
gi 315362229 gb HQ672502 1 Uncultured bacterium clone F9P4500 S O24
gi 151933610 gb EF645947 1 Uncultured bacterium clone N60e 103
gi 124302393 gb EF379450 1 Uncultured bacterium clone TLC-PA3-31
gi 310785664 gb HQ226370 1 Uncultured bacterium clone RF182B09
gi 260069504 gb GQ348234 1 Uncultured marine bacterium clone SGSX799
gi 260071769 gp GQ350499 1 Uncultured marine bacterium clone SHBC739
gi 729259950 gb KM461755 1 Uncultured Marinimicrobia bacterium clone ST010965 F3
gi 260071802 gp GQ350532 1 Uncultured marine bacterium clone SHBC788
gi 307567702 gb HQ163554 1 Uncultured bacterium clone SHZZ662
gi 260068247 gp GQ346977 1 Uncultured marine bacterium clone SGPZ651
gi 260071625 gp GQ350355 1 Uncultured marine bacterium clone SHBC563
gi 307567712 gb HQ163564 1 Uncultured bacterium clone SHZZ677
gi 260069033 gp GQ348033 1 Uncultured marine bacterium clone SGST537
gi 260070652 gp GQ349382 1 Uncultured marine bacterium clone SHAS477
gi 62959114 gb DQ009427 1 Uncultured marine bacterium clone SPOTSFEB02 70m8
gi 260070258 gp GQ348988 1 Uncultured marine bacterium clone SHAG592
KF170421 1 Uncultured bacterium FPPP 13C3
gi 260071874 gp GQ350604 1 Uncultured marine bacterium clone SHBC883
gi 260068202 gp GQ346932 1 Uncultured marine bacterium clone SGPZ564
gi 151933624 gb EF645961 1 Uncultured bacterium clone N60e 133
gi 209929695 gb EU035866 2 Uncultured bacterium clone Kwa45
gi 729260061 gb KM461866 1 Uncultured Marinimicrobia bacterium clone ST070950 B7
gi 283854636 gb GQ337084 1 Uncultured SAR406 cluster bacterium clone NB210b 88
gi 283854869 gb GQ337317 1 Uncultured SAR406 cluster bacterium clone NB210b 70
AF355047 1 Uncultured delta proteobacterium Arctic96B-7
gi 260069474 gp GQ348204 1 Uncultured marine bacterium clone SGSX1123
gi 729259938 gb KM461743 1 Uncultured Marinimicrobia bacterium clone ST010965 D7
GU474916 1 Uncultured SAR406 cluster bacterium HF0500 01L02
gi 289451251 gb FJ980472 1 Uncultured bacterium clone 4E-008
gi 315363812 gb HQ674085 1 Uncultured bacterium clone F9P261000 S O04
gi 289452239 gb FJ981460 1 Uncultured bacterium clone 3E-142
gi 315362884 gb HQ673157 1 Uncultured bacterium clone F9P41000 S F23
gi 289452110 gb FJ981331 1 Uncultured bacterium clone 3E-003
gi 289451871 gb FJ981092 1 Uncultured bacterium clone 2E-062
gi 289451856 gb FJ981077 1 Uncultured bacterium clone 2E-047
gi 289451305 gb FJ980526 1 Uncultured bacterium clone 4E-062
gi 289451594 gb FJ980815 1 Uncultured bacterium clone 1E-064
gi 289451302 gb FJ980523 1 Uncultured bacterium clone 4E-059
gi 315362451 gb HQ672724 1 Uncultured bacterium clone F9P12500 S M20
gi 315362975 gb HQ673248 1 Uncultured bacterium clone F9P41000 S L04
gi 283854677 gb GQ337125 1 Uncultured SAR406 cluster bacterium clone C80565b 54
gi 283854717 gb GQ337165 1 Uncultured SAR406 cluster bacterium clone CB1201b 69
gi 289451320 gb FJ980541 1 Uncultured bacterium clone 4E-077
gi 289451905 gb FJ981126 1 Uncultured bacterium clone 2E-099
gi 289451932 gb FJ981153 1 Uncultured bacterium clone 2E-127
gi 289451892 gb FJ981113 1 Uncultured bacterium clone 2E-086
gi 729260062 gb KM461867 1 Uncultured Marinimicrobia bacterium clone ST070950 B8
gi 315362539 gb HQ672812 1 Uncultured bacterium clone F9P26500 S B22
gi 315362623 gb HQ672898 1 Uncultured bacterium clone F9P26500 S G20
gi 289451704 gb FJ980925 1 Uncultured bacterium clone 1E-180
gi 315362806 gb HQ673079 1 Uncultured bacterium clone F9P41000 S B03
gi 315362843 gb HQ673116 1 Uncultured bacterium clone F9P41000 S D09
gi 310947641 gb HQ433424 1 Uncultured SAR406 cluster bacterium clone 203-10-29
gi 315364085 gb HQ674358 1 Uncultured bacterium clone F9P262000 S C14
gi 347663694 gb JN018851 1 Uncultured bacterium clone 52-3-34
gi 315364279 gb HQ674552 1 Uncultured bacterium clone F9P262000 S M22
gi 353417004 gb JQ678404 1 Uncultured bacterium clone JdFBHP1-30
gi 310947594 gb HQ433377 1 Uncultured SAR406 cluster bacterium clone 191-2-26
gi 347663676 gb JN018833 1 Uncultured bacterium clone 52-3-07
gi 315363646 gb HQ673919 1 Uncultured bacterium clone F9P122000 S K05
gi 98971420 gb DQ513043 1 Uncultured bacterium clone FS142-10B-02
gi 315363707 gb HQ673980 1 Uncultured bacterium clone F9P122000 S N19
gi 283854697 gb GQ337145 1 Uncultured SAR406 cluster bacterium clone CB1199b 61
gi 283854795 gb GQ337243 1 Uncultured SAR406 cluster bacterium clone CB1891b 55
gi 315362135 gb HQ672408 1 Uncultured bacterium clone F9P4500 S J18
gi 315363452 gb HQ673725 1 Uncultured bacterium clone F9P121000 S M14
gi 260072419 gp GQ351149 1 Uncultured marine bacterium clone JBP41000 1B10
gi 260072453 gp GQ351183 1 Uncultured marine bacterium clone JBP41000 1F10
gi 315362290 gb HQ672563 1 Uncultured bacterium clone F9P12500 S C11
gi 315362440 gb HQ672713 1 Uncultured bacterium clone F9P12500 S L24
gi 315362203 gb HQ672476 1 Uncultured bacterium clone F9P4500 S N09
gi 310814305 gb GU145485 1 Uncultured bacterium clone BS100
gi 315363675 gb HQ673948 1 Uncultured bacterium clone F9P122000 S M01
gi 315362019 gb HQ672292 1 Uncultured bacterium clone F9P4500 S D10
gi 260072512 gp GQ351242 1 Uncultured marine bacterium clone JBP41000 2E07
gi 315363423 gb HQ673696 1 Uncultured bacterium clone F9P121000 S J07
347663690 gb JN018847 1 Uncultured bacterium clone 52-3-28
gi 315362816 gb HQ673089 1 Uncultured bacterium clone F9P41000 S B19
gi 326485781 gb GU940742 1 Uncultured bacterium clone N201B 341
gi 512125014 gb KC873537 1 Uncultured bacterium clone S62 048
gi 283854703 gb GQ337151 1 Uncultured SAR406 cluster bacterium clone CB1199b 91
HQ674572 1 Uncultured bacterium clone F9P262000 S N21
GU474892 1 Uncultured Marinimicrobia bacterium HF0500 22B16
gi 315364304 gb HQ674577 1 Uncultured bacterium clone F9P262000 S O02
gi 460184055 gb KC522902 1 Uncultured bacterium clone CTD07 S13 2850m 18F
gi 169404012 gb EU491156 1 Uncultured bacterium clone P9X2b3H06
gi 549150237 dbj AB858515 1 Uncultured bacterium gene for
gi 260072529 gp GQ351259 1 Uncultured marine bacterium clone JBP41000 2H01
347663601 gb JN018758 1 Uncultured bacterium clone 41-12-06
gi 260072466 gp GQ351196 1 Uncultured bacterium clone JBP41000 1H02
gi 124373774 dbj AB292141 1 Uncultured bacterium gene for
gi 315363967 gb HQ674240 1 Uncultured bacterium clone F9P261000 S M09
gi 315363124 gb HQ673397 1 Uncultured bacterium clone F9P41000 E23
gi 260071862 gp GQ350592 1 Uncultured marine bacterium clone SHBC870
gi 315364152 gb HQ674425 1 Uncultured bacterium clone F9P262000 S G02
gi 283847779 gb GQ337227 1 Uncultured SAR406 cluster bacterium clone CB1648b 77
gi 315363565 gb HQ673838 1 Uncultured bacterium clone F9P122000 S F03
gi 315362637 gb HQ672910 1 Uncultured bacterium clone F9P26500 S H13
gi 315363087 gb HQ673360 1 Uncultured bacterium clone F9P41300 B24
gi 315363622 gb HQ673895 1 Uncultured bacterium clone F9P122000 S I17
gi 260071265 gp GQ349695 1 Uncultured marine bacterium clone SHAZ596
gi 315362376 gb HQ672649 1 Uncultured bacterium clone F9P12500 S H11
gi 315363585 gb HQ673858 1 Uncultured bacterium clone F9P122000 S G08
gi 283854774 gb GQ337222 1 Uncultured SAR406 cluster bacterium clone CB1648b 40
gi 315364340 gb HQ674613 1 Uncultured bacterium clone F9P262000 S P21
gi 260069663 gb GQ348693 1 Uncultured marine bacterium clone SHAB805
gi 315362906 gb HQ673179 1 Uncultured bacterium clone F9P41000 S H05
gi 315364178 gb HQ674451 1 Uncultured bacterium clone F9P262000 S H12
347663725 gb JN018882 1 Uncultured bacterium clone 52-3-69
gi 315363155 gb HQ673428 1 Uncultured bacterium clone F9P41300 G17
gi 383471000 gb JQ678400 1 Uncultured bacterium clone JdFBHP93
gi 1150733 gb U41450 1 UEU41450 Uncultured eubacterium OCS307
gi 315362179 gb HQ672452 1 Uncultured bacterium clone F9P4500 S M05
gi 260072477 gp GQ351207 1 Uncultured marine bacterium clone JBP41000 2A04
gi 301177202 gb HM798785 1 Uncultured SAR406 cluster bacterium clone PRTAB7918
gi 283854877 gb GQ337325 1 Uncultured SAR406 cluster bacterium clone NB210b 90
gi 315363158 gb HQ673431 1 Uncultured bacterium clone F9P41300 G24
gi 289451327 gb FJ980548 1 Uncultured bacterium clone 4E-084
gi 289451702 gb FJ980923 1 Uncultured bacterium clone 1E-178
gi 289452247 gb FJ981468 1 Uncultured bacterium clone 3E-150
gi 315362040 gb HQ672313 1 Uncultured bacterium clone F9P4500 S E18
gi 289451950 gb FJ981171 1 Uncultured bacterium clone 2E-145
gi 315363845 gb HQ674118 1 Uncultured bacterium clone F9P261000 S G04
gi 215261921 emb AM911427 1 Uncultured bacterium partial
gi 315362051 gb HQ672324 1 Uncultured bacterium clone F9P4500 S F07
gi 215261967 emb AM911473 1 Uncultured bacterium partial
gi 209869927 emb FM246673 1 Uncultured marine bacterium partial
gi 307567321 gb HQ163173 1 Uncultured bacterium clone SHZW475
gi 460184986 gb KC522933 1 Uncultured bacterium clone CTD07 S13 2850m 95F
gi 315363431 gb HQ673704 1 Uncultured bacterium clone F9P121000 S J17
gi 451175364 gb KC197537 1 Uncultured bacterium clone GAL 3E4
gi 451175367 gb KC197540 1 Uncultured bacterium clone GAL 3E9
gi 193891079 gb EU817592 1 Uncultured bacterium clone 670m Bact8
gi 304650493 gb HQ015525 1 Uncultured bacterium clone ST13 500m clone18
gi 729259963 gb KM461768 1 Uncultured Marinimicrobia bacterium clone ST010965 H10
gi 315362854 gb HQ673127 1 Uncultured bacterium clone F9P41000 S E02
gi 315363923 gb HQ674196 1 Uncultured bacterium clone F9P261000 S K03
gi 315362454 gb HQ672727 1 Uncultured bacterium clone F9P12500 S N03
gi 322512343 gb HQ675668 1 Uncultured bacterium clone SCGC AAA288-E05
gi 63284488 gb DQ300753 1 Uncultured bacterium clone HF770 B3 P1
gi 315364075 gb HQ674348 1 Uncultured bacterium clone F9P262000 S B22
gi 315364239 gb HQ674512 1 Uncultured bacterium clone F9P262000 S K13
347663613 gb JN018770 1 Uncultured bacterium clone 41-12-21
347663707 gb JN018864 1 Uncultured bacterium clone 52-3-51
gi 315363832 gb HQ674105 1 Uncultured bacterium clone F9P41000 S F13
gi 315363978 gb HQ674251 1 Uncultured bacterium clone F9P261000 S M23
gi 315362505 gb HQ672778 1 Uncultured bacterium clone F9P26500 S A07
gi 156046540 gb EU092076 1 Uncultured marine bacterium clone A723901
gi 156046546 gb EU092082 1 Uncultured marine bacterium clone A723911
gi 315362791 gb HQ673040 1 Uncultured bacterium clone F9P41000 S A07
gi 14334254 gb AY033318 1 Uncultured bacterium MB13C05
gi 452096637 gb JX864735 1 Uncultured bacterium clone BF2009 Sep 21m H10
gi 2290744 gb AF001653 1 Uncultured eubacterium OCS146
gi 253796046 gb QQ259226 1 Uncultured bacterium clone water deep04
148731551 gb EF573770 1 Uncultured bacterium clone SZS 114
148731527 gb EF573746 1 Uncultured bacterium clone SZS 90
148732422 gb EF574641 1 Uncultured bacterium clone SZS 985
148730199 gb EF572418 1 Uncultured bacterium clone SZS 517
148733101 gb EF575320 1 Uncultured bacterium clone SZS 1664
190709195 gb EU804270 1 Uncultured bacterium clone EC232172
190710016 gb EU805091 1 Uncultured bacterium clone EC233076
326485776 gb GU940737 1 Uncultured bacterium clone N201B 317
51129417 gb KC872810 1 Uncultured bacterium clone S30 028
148731290 gb EF573509 1 Uncultured bacterium clone S23 1608
148732879 gb EF575098 1 Uncultured bacterium clone SZS 1442
156046511 gb EU092047 1 Uncultured marine bacterium clone A314911
313118428 gb HQ443329 1 Uncultured bacterium clone MCR07 19C03
315361755 gb HQ672028 1 Uncultured bacterium clone F9P2610 S C07
62549162 gb AY907795 1 Uncultured bacterium clone A713808
440509268 gb KC003262 1 Unidentified marine bacterioplankton clone E412B 45
512129609 gb KC874004 1 Uncultured bacterium clone S68 041
326485959 gb GU940920 1 Uncultured bacterium clone N412B 45
148730647 gb EF572866 1 Uncultured bacterium clone S23 965
440507976 gb KC001970 1 Unidentified marine bacterioplankton clone P4-7B 127
148731438 gb EF573657 1 Uncultured bacterium clone SZS 1
U34043 1 Uncultured eubacterium SAR406
260070449 gp GQ349179 1 Uncultured marine bacterium clone SHAN505
260070509 gp GQ349239 1 Uncultured marine bacterium clone SHAN623
260070390 gp GQ349120 1 Uncultured marine bacterium clone SHAN395
260070508 gp GQ349238 1 Uncultured marine bacterium clone SHAN821
260070794 gp GQ349524 1 Uncultured marine bacterium clone SHAS704
260070461 gp GQ349191 1 Uncultured marine bacterium clone SHAN527
729260051 gb KM461856 1 Uncultured Marinimicrobia bacterium clone ST070950 A5
315364200 gb HQ674473 1 Uncultured bacterium clone F9P262000 S I14
347663768 gb JN018915 1 Uncultured bacterium clone 65-11-22
AY907803 1 Uncultured bacterium clone A714018
AF355046 1 Uncultured delta proteobacterium Arctic95A-2
AY458631 1 Uncultured bacterium clone 159 clone EBAC750-03B02
315363573 gb HQ673846 1 Uncultured bacterium clone F9P122000 S F14
315362766 gb HQ673039 1 Uncultured bacterium clone F9P26500 S O19
315363568 gb HQ673841 1 Uncultured bacterium clone F9P122000 S F06
GU474850 1 Uncultured SAR406 cluster bacterium HF0010 18O13
SAR406 cluster bacterium SCGC AAA298-D23
AF382116 1 Uncultured bacterium clone ZA3312c
SAR406 cluster bacterium SCGC AAA160-I06
HQ673390 1 Uncultured bacterium clone F9P41300 E03
HQ674365 1 Uncultured bacterium clone F9P262000 S D03
283854798 gb GQ337246 1 Uncultured SAR406 cluster bacterium clone CB1891b 75
SAR406 cluster bacterium SCGC AAA003-E22
DQ300775 1 Uncultured bacterium clone HF770 D10 P1
315363497 gb HQ673770 1 Uncultured bacterium clone F9P122000 S A18
SAR406 cluster bacterium SCGC AAA003-L08
315362015 gb HQ672288 1 Uncultured bacterium clone F9P4500 S D04
315361673 gb HQ671946 1 Uncultured bacterium clone F9P1210 S M14
315363634 gb HQ673907 1 Uncultured bacterium clone F9P122000 S J12
98971466 gb DQ513089 1 Uncultured bacterium clone FS266-81B-03
AF382142 1 Uncultured bacterium clone ZA3648c
260070808 gp GQ349538 1 Uncultured marine bacterium clone SHAS729
260072125 gp GQ350855 1 Uncultured marine bacterium clone SHBH527
260072078 gp GQ350808 1 Uncultured marine bacterium clone SHBH433
260068791 gp GQ347521 1 Uncultured marine bacterium clone SGSH591
260072246 gp GQ350776 1 Uncultured marine bacterium clone SHBH1141
260072158 gp GQ350888 1 Uncultured marine bacterium clone SHBH596
260068666 gb GQ347396 1 Uncultured marine bacterium clone SGSH1088
260070562 gp GQ349292 1 Uncultured marine bacterium clone SHAN712
260069648 gp GQ348378 1 Uncultured marine bacterium clone SGTA501
260071986 gp GQ350716 1 Uncultured marine bacterium clone SHBH1034
260068637 gp GQ347367 1 Uncultured marine bacterium clone SGSH1026
260070629 gp GQ349359 1 Uncultured marine bacterium clone SHAS434
260072344 gp GQ351074 1 Uncultured marine bacterium clone SHBH948
260070643 gp GQ349373 1 Uncultured marine bacterium clone SHAS460
260069751 gp GQ348481 1 Uncultured marine bacterium clone SGTA706
260068928 gp GQ347658 1 Uncultured marine bacterium clone SGSH895
260069601 gp GQ348331 1 Uncultured marine bacterium clone SGTA406
260070276 gp GQ349086 1 Uncultured marine bacterium clone SHAG620
260068910 gp GQ347640 1 Uncultured marine bacterium clone SGSH856
260069614 gp GQ348344 1 Uncultured marine bacterium clone SGTA434
260072018 gp GQ350748 1 Uncultured marine bacterium clone SBH1087
260068799 gp GQ347529 1 Uncultured marine bacterium clone SGSH617
260071995 gp GQ350725 1 Uncultured marine bacterium clone SBH1047
260072310 gp GQ351040 1 Uncultured marine bacterium clone SHBH890
GQ350809 1 Uncultured marine bacterium clone SHBH435
307567840 gb HQ163652 1 Uncultured bacterium clone SIAC543
260068624 gb GQ347354 1 Uncultured marine bacterium clone SGSH1003
260068702 gb GQ347432 1 Uncultured marine bacterium clone SGSH403
260069602 gb GQ348332 1 Uncultured marine bacterium clone SGTA407
260072221 gp GQ350951 1 Uncultured marine bacterium clone SBH720
260070286 gp GQ349016 1 Uncultured marine bacterium clone SHAG636
260072194 gp GQ350924 1 Uncultured marine bacterium clone SHBH665
260070425 gp GQ349155 1 Uncultured marine bacterium clone SHAN462
260070486 gp GQ349216 1 Uncultured marine bacterium clone SHAN580
260071998 gp GQ350728 1 Uncultured marine bacterium clone SHBH1051
269131752 gb GU108527 1 Uncultured gamma proteobacterium clone V18-11
452814299 gb KC492866 1 Uncultured bacterium clone C36
469665343 gb KC545714 1 Uncultured SAR406 cluster bacterium clone BFB040
45-1&scaffold 922rc
219693615 gb FJ517138 1 Uncultured bacterium clone TDNP Wbc97 96 1 238
321158664 dbj AB611156 1 Uncultured bacterium gene for
148613661 gb EF395643 1 Uncultured bacterium clone CBM01D05
238926077 gb FJ973600 1 Uncultured bacterium clone SS WC 30
233826871 gb FJ918434 1 Uncultured bacterium clone SI3P0 cl5
37624995 gb AY381288 1 Uncultured bacterium clone Lgja-39
227353399 gb FJ628328 1 Uncultured bacterium clone NitSAu0613 728
260069613 gb GQ348343 1 Uncultured gamma proteobacterium clone SGTA433
189909304 gb EU570890 1 Uncultured bacterium clone NitSA0650 165
45&scaffold 322rc
209869799 emb FM246547 1 Uncultured marine bacterium partial
JQ316609 1 Uncultured bacterium clone MIC1R04
SAR406 cluster bacterium JGI 000039-D08
AJ430587 1 Caldothrix abyssi

**Figure S5**

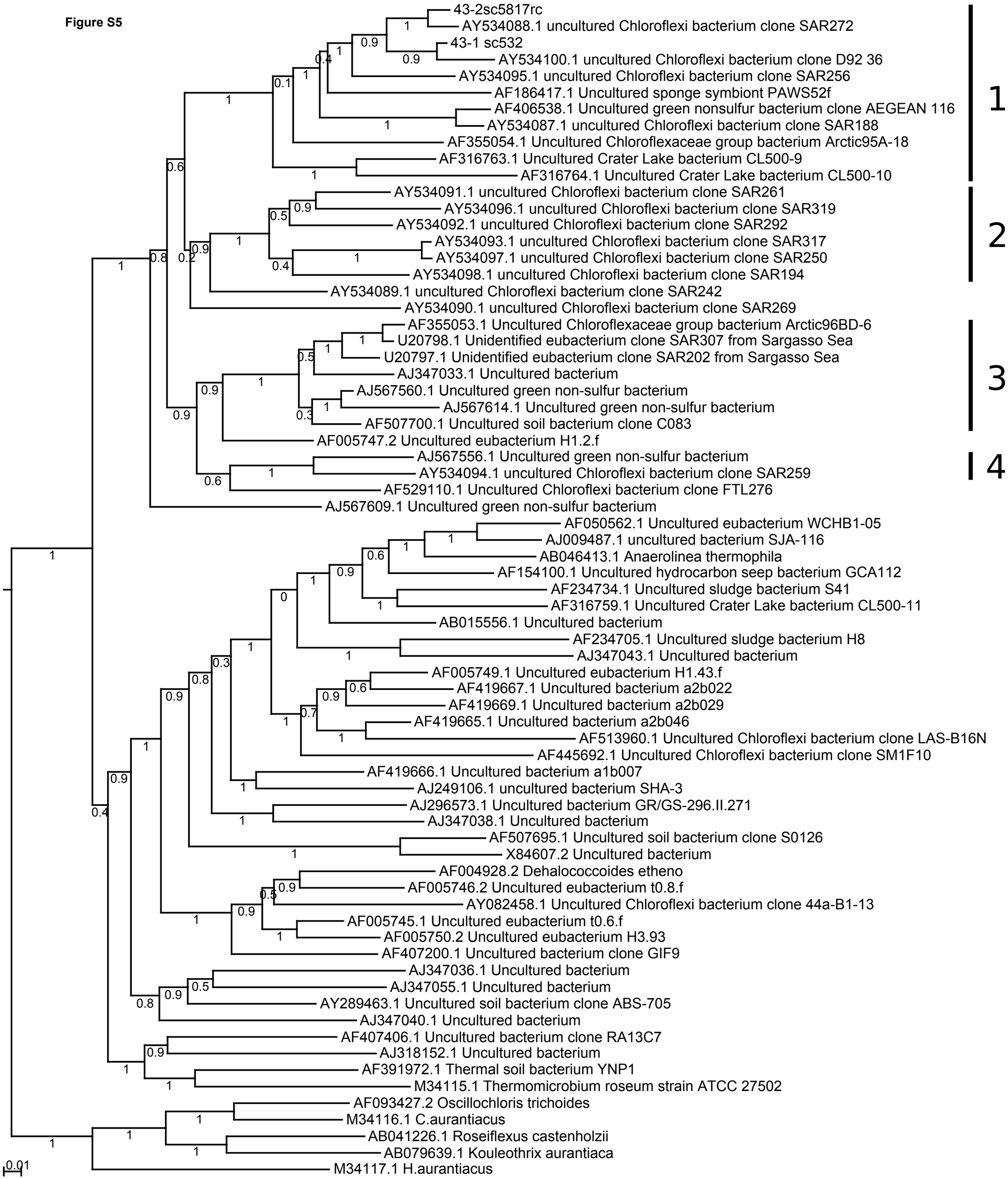

43-2sc5817rc
AY534088.1 uncultured Chloroflexi bacterium clone SAR272
43-1 sc532
AY534100.1 uncultured Chloroflexi bacterium clone D92 36
AY534095.1 uncultured Chloroflexi bacterium clone SAR256
AF186417.1 Uncultured sponge symbiont PAWS52f
AF406538.1 Uncultured green nonsulfur bacterium clone AEGEAN 116
AY534087.1 uncultured Chloroflexi bacterium clone SAR188
AF355054.1 Uncultured Chloroflexaceae group bacterium Arctic95A-18
AF316763.1 Uncultured Crater Lake bacterium CL500-9
AF316764.1 Uncultured Crater Lake bacterium CL500-10
AY534091.1 uncultured Chloroflexi bacterium clone SAR261
AY534096.1 uncultured Chloroflexi bacterium clone SAR319
AY534092.1 uncultured Chloroflexi bacterium clone SAR292
AY534093.1 uncultured Chloroflexi bacterium clone SAR317
AY534097.1 uncultured Chloroflexi bacterium clone SAR250
AY534098.1 uncultured Chloroflexi bacterium clone SAR194
AY534089.1 uncultured Chloroflexi bacterium clone SAR242
AY534090.1 uncultured Chloroflexi bacterium clone SAR269
AF355053.1 Uncultured Chloroflexaceae group bacterium Arctic96BD-6
U20798.1 Unidentified eubacterium clone SAR307 from Sargasso Sea
U20797.1 Unidentified eubacterium clone SAR202 from Sargasso Sea
AJ347033.1 Uncultured bacterium
AJ567560.1 Uncultured green non-sulfur bacterium
AJ567614.1 Uncultured green non-sulfur bacterium
AF507700.1 Uncultured soil bacterium clone C083
AF005747.2 Uncultured eubacterium H1.2.f
AJ567556.1 Uncultured green non-sulfur bacterium
AY534094.1 uncultured Chloroflexi bacterium clone SAR259
AF529110.1 Uncultured Chloroflexi bacterium clone FTL276
AJ567609.1 Uncultured green non-sulfur bacterium
AF050562.1 Uncultured eubacterium WCHB1-05
AJ009487.1 uncultured bacterium SJA-116
AB046413.1 Anaerolinea thermophila
AF154100.1 Uncultured hydrocarbon seep bacterium GCA112
AF234734.1 Uncultured sludge bacterium S41
AF316759.1 Uncultured Crater Lake bacterium CL500-11
AB015556.1 Uncultured bacterium
AF234705.1 Uncultured sludge bacterium H8
AJ347043.1 Uncultured bacterium
AF005749.1 Uncultured eubacterium H1.43.f
AF419667.1 Uncultured bacterium a2b022
AF419669.1 Uncultured bacterium a2b029
AF419665.1 Uncultured bacterium a2b046
AF513960.1 Uncultured Chloroflexi bacterium clone LAS-B16N
AF445692.1 Uncultured Chloroflexi bacterium clone SM1F10
AF419666.1 Uncultured bacterium a1b007
AJ249106.1 uncultured bacterium SHA-3
AJ296573.1 Uncultured bacterium GR/GS-296.II.271
AJ347038.1 Uncultured bacterium
AF507695.1 Uncultured soil bacterium clone S0126
X84607.2 Uncultured bacterium
AF004928.2 Dehalococcoides etheno
AF005746.2 Uncultured eubacterium t0.8.f
AY082458.1 Uncultured Chloroflexi bacterium clone 44a-B1-13
AF005745.1 Uncultured eubacterium t0.6.f
AF005750.2 Uncultured eubacterium H3.93
AF407200.1 Uncultured bacterium clone GIF9
AJ347036.1 Uncultured bacterium
AJ347055.1 Uncultured bacterium
AY289463.1 Uncultured soil bacterium clone ABS-705
AJ347040.1 Uncultured bacterium
AF407406.1 Uncultured bacterium clone RA13C7
AJ318152.1 Uncultured bacterium
AF391972.1 Thermal soil bacterium YNP1
M34115.1 Thermomicrobium roseum strain ATCC 27502
AF093427.2 Oscillochloris trichoides
M34116.1 C.aurantiacus
AB041226.1 Roseiflexus castenholzii
AB079639.1 Kouleothrix aurantiaca
M34117.1 H.aurantiacus
0.01
1
2
3
4

**Figure S6**

PRR10 4 11429
PRR10 5 11429
PRR10 3 11429
PRR10 11 11429
PRR10 6 11429
PRR12 1 6 1945
PRR10 1 11372
PRR10 2 11420
PRR10 8 1901
PRR10 9 1899
PRR10 7 1908
PRR10 10 1895
PRR12 3 1 11403
PRR12 3 3 11383
PRR12 3 2 11353
PRR12 2 1 11429
PRR12 2 4 11429
PRR12 2 3 11429
PRR12 2 2 11429
PRR12 2 5 11429
PRR12 2 6 11279
PRR12 1 2 11429
PRR12 1 3 11429
PRR12 1 4 11429
PRR12 1 5 11429
PRR12 1 1 1957
MSB4E2 4 11372
MSB4E2 5 11384
MSB4E2 1 11429
MSB4E2 2 11429
MSB4E2 3 11429
CV106 1 11429
CV106 4 11429
CV106 6 11429
CV106 3 11429
CV106 5 11429
PRR1211350
PRR1211379
PRR1211359
PRR1211387
wb1 H11 1 11429
wb1 H11 5 11322
wb1 H11 2 11429
SSS58A 2 11408
SSS58A 7 11412
SSS58A 3 11426
SSS58A 4 11312
SSS58A 5 11429
SSS58A 6 11429
SSS58A 9 11429
SSS58A 1 11429
SSS58A 10 11429
SSS58A 8 11290
LD1PA13 3 11407
LD1PA13 4 11390
LD1PA13 5 11393
LD1PA13 7 11429
LD1PA13 2 11429
LD1PA13 6 11429
GN03 16 11392
GN03 18 11428
GN03 15 11429
GN03 6 11429
GN03 19 11429
GN03 11 11415
GN03 13 11422
GN03 12 11429
GN03 14 11429
GN03 17 11429
GN03 7 11429
GN03 8 11429
GN03 5 11429
GN03 10 11429
GN03 9 11429
GN03 20 11410
GN03 4 11410
RII 11429
RII AN061 11429 2
RII 11386
SAW1 B6 1 11429
SAW1 B6 5 11429
SAW1 B6 4 11072
SAW1 B6 3 11409
SAW1 B6 6 11395
PBSIII9 11 11236
PBSIII9 23 11259
PBSIII9 2 1916
PBSIII9 16 11403
PBSIII9 15 11182
PBSIII9 20 11230
PBSIII9 18 11075
PBSIII9 13 11398
PBSIII9 14 11429
PBSIII9 12 11429
PBSIII9 19 11429
PBSIII9 17 11429
PBSIII9 7 11088
PBSIII9 6 11389
PBSIII9 8 11402
PBSIII9 10 11402
PBSIII9 9 11429
PBSIII9 1 11412
PBSIII9 24 11429
PBSIII9 21 11420
PBSIII9 22 11370
PBSIII9 3 11400
PBSIII9 5 1988
PBSIII9 4 1874
105967 AY485294 p PAUC34f otu 2404
106276 AY485296 p PAUC34f otu 2404
106433 AY485297 p PAUC34f otu 2404
106114 AY485295 p PAUC34f otu 2404
592392 GU118564 p PAUC34f otu 2404
160190 DQ513022 p PAUC34f otu 2404
536025 GQ351144 p PAUC34f otu 2404
539443 FJ746128 p PAUC34f otu 2404
586109 GQ346983 PAUC34f otu 2404
238478 EF076099 p PAUC34f otu 2404
242733 EF076082 p PAUC34f otu 2404
588254 FJ529325 p PAUC34f otu 2404
1674 AF186412 p PAUC34f otu 2404
50&&scaffold 2726
Bacillus acidogenesis 71698 11429
Planococcus citreus 15025 11429
Lactobacillus acidophilus 108734 11429
Clostridium acetobutylicum 129255 11429
Desulfitobacter alkalitolerans 98963 11387
M62693 1 Fibrobacter succinogenes strain MC1
M62696 2 Fibrobacter succinogenes subsp succinogenes S85
M62691 1 Fibrobacter intestinalis strain LH1
Edaphobacter modestum 160117 11420
Terriglobus roseus 162763 11420
Acidobacterium capsulatum 3704 11422
AJ290824 1 Chlorobium limicola strain DSM246
Y10651 2 Chlorobium phaeobacteroides strain UdG 6047
EU770420 1 Chlorobium sp sy9
NR 119242 1 Chlorobium phaeovibrioides strain DSM 269
C ferrooxidans 171289 11429
NR 119288 2 Chlorobium ferrooxidans strain KofoX
AM086645 1 Pelodictyon clathratiforme strain 4DE
DQ383316 1 Uncultured Chlorobium sp clone i9-10
C tepidum 137177 11429
GQ200196 2 Bacteroidetes bacterium DQHS-15
KP214057 1 Bacteroidetes bacterium FA423
AM179866 1 Sporocytophaga sp A61
Sulcia muelleri 257992 11429
Leptospirillum 11429
Nitro 11426
Nitrospiraes Thermodesulfovibri11429
Chroococcus minutus 570499 11429
Oscillatoria duplisecta 166338 11354
Synechococcus elongatus 110298 11429
Gloeobacter violaceus 108670 11429
Anaerolinea thermophila 44630 11429
Caldilinea aerophila 26613 11427
Dehalococcoides BAV1 76593 11420
Kosmotoga 376719 11429
Thermotoga 62474 11429
Aminobacterium 11429
Synergistes 11377
Armatimonas 11429
Chthonomonas 11404
P brasiliensis 2857 11429
P limnophilus 2845 11429
agg27 OM190 137332 11429
Methylacidiphilum 11429
Verrucomicrobium 11376
OP11 GN01 11354
OP11 WCHB107 11417
OP11 OP11 3 11336
hungatei 11429

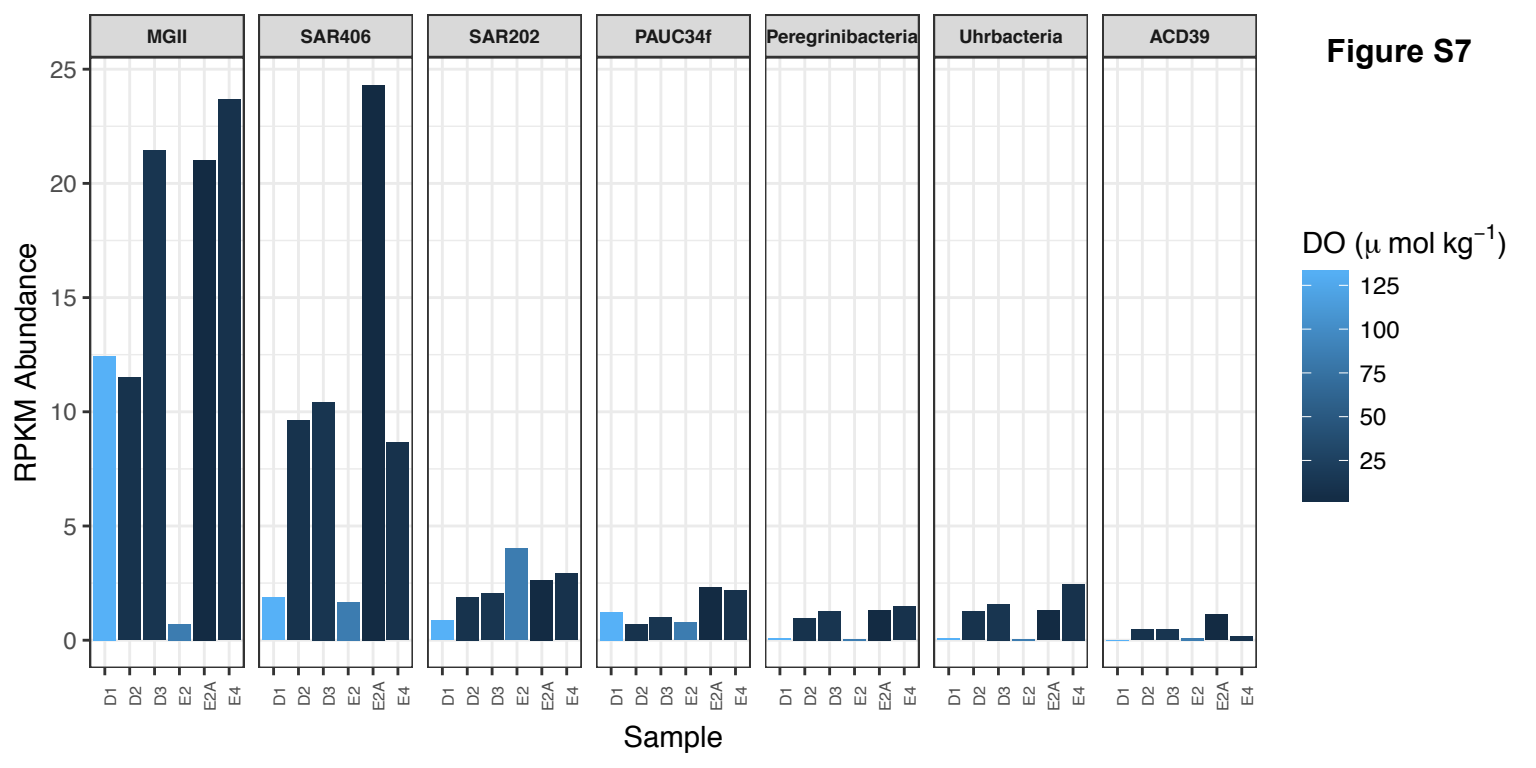


Figure S7
MGII
SAR406
SAR202
PAUC34f
Peregrinibacteria
Uhrbacteria
ACD39
RPKM Abundance
0
5
10
15
20
25
D1
D2
D3
E2
E2A
E4
Sample
DO (μ mol kg$^{-1}$)
125
100
75
50
25

**Figure S8**

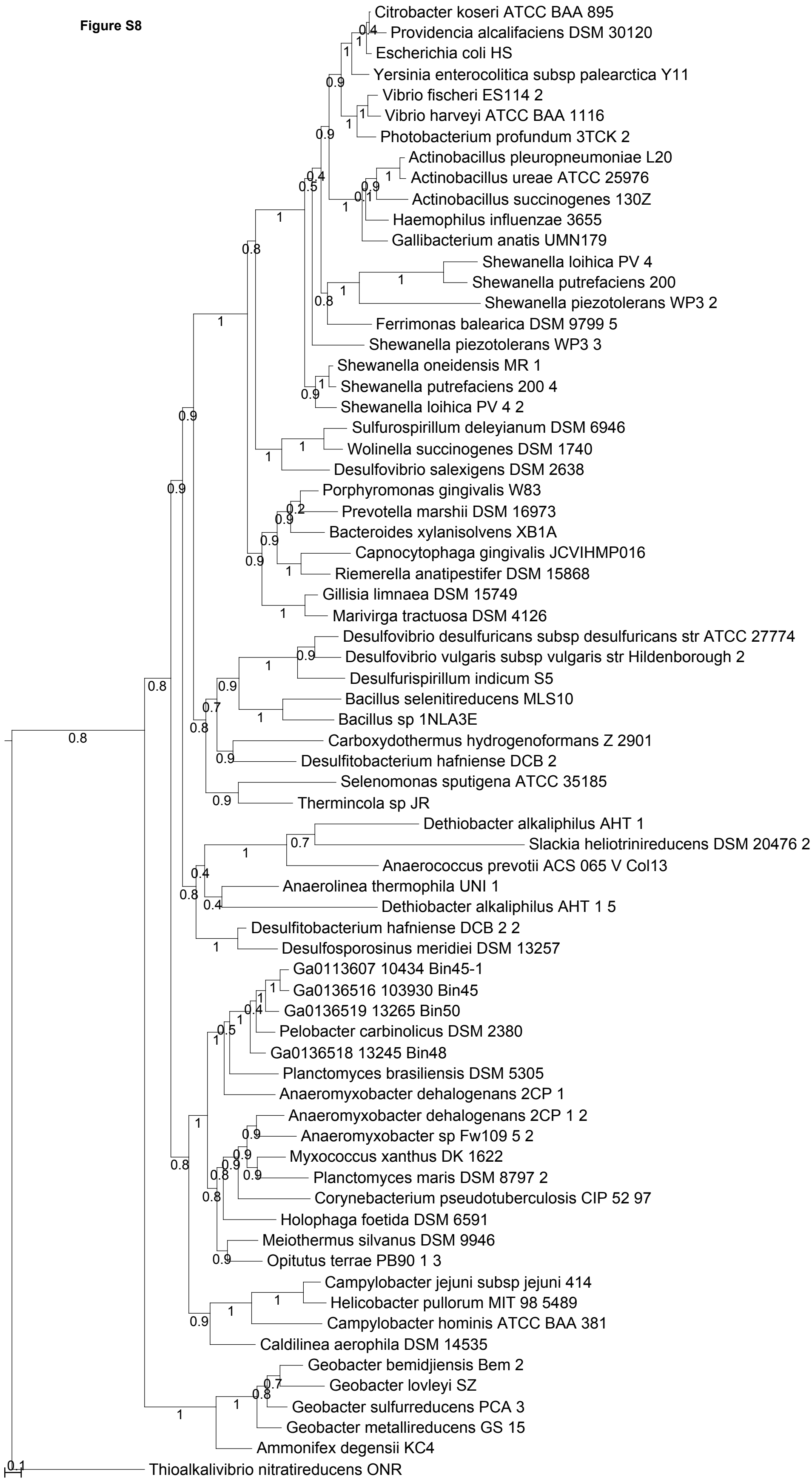

Citrobacter koseri ATCC BAA 895
Providencia alcalifaciens DSM 30120
Escherichia coli HS
Yersinia enterocolitica subsp palearctica Y11
Vibrio fischeri ES114 2
Vibrio harveyi ATCC BAA 1116
Photobacterium profundum 3TCK 2
Actinobacillus pleuropneumoniae L20
Actinobacillus ureae ATCC 25976
Actinobacillus succinogenes 130Z
Haemophilus influenzae 3655
Gallibacterium anatis UMN179
Shewanella loihica PV 4
Shewanella putrefaciens 200
Shewanella piezotolerans WP3 2
Ferrimonas balearica DSM 9799 5
Shewanella piezotolerans WP3 3
Shewanella oneidensis MR 1
Shewanella putrefaciens 200 4
Shewanella loihica PV 4 2
Sulfurospirillum deleyianum DSM 6946
Wolinella succinogenes DSM 1740
Desulfovibrio salexigens DSM 2638
Porphyromonas gingivalis W83
Prevotella marshii DSM 16973
Bacteroides xylanisolvens XB1A
Capnocytophaga gingivalis JCVIHMP016
Riemerella anatipestifer DSM 15868
Gillisia limnaea DSM 15749
Marivirga tractuosa DSM 4126
Desulfovibrio desulfuricans subsp desulfuricans str ATCC 27774
Desulfovibrio vulgaris subsp vulgaris str Hildenborough 2
Desulfurispirillum indicum S5
Bacillus selenitireducens MLS10
Bacillus sp 1NLA3E
Carboxydothermus hydrogenoformans Z 2901
Desulfitobacterium hafniense DCB 2
Selenomonas sputigena ATCC 35185
Thermincola sp JR
Dethiobacter alkaliphilus AHT 1
Slackia heliotrinireducens DSM 20476 2
Anaerococcus prevotii ACS 065 V Col13
Anaerolinea thermophila UNI 1
Dethiobacter alkaliphilus AHT 1 5
Desulfitobacterium hafniense DCB 2 2
Desulfosporosinus meridiei DSM 13257
Ga0113607 10434 Bin45-1
Ga0136516 103930 Bin45
Ga0136519 13265 Bin50
Pelobacter carbinolicus DSM 2380
Ga0136518 13245 Bin48
Planctomyces brasiliensis DSM 5305
Anaeromyxobacter dehalogenans 2CP 1
Anaeromyxobacter dehalogenans 2CP 1 2
Anaeromyxobacter sp Fw109 5 2
Myxococcus xanthus DK 1622
Planctomyces maris DSM 8797 2
Corynebacterium pseudotuberculosis CIP 52 97
Holophaga foetida DSM 6591
Meiothermus silvanus DSM 9946
Opitutus terrae PB90 1 3
Campylobacter jejuni subsp jejuni 414
Helicobacter pullorum MIT 98 5489
Campylobacter hominis ATCC BAA 381
Caldilinea aerophila DSM 14535
Geobacter bemidjiensis Bem 2
Geobacter lovleyi SZ
Geobacter sulfurreducens PCA 3
Geobacter metallireducens GS 15
Ammonifex degensii KC4
Thioalkalivibrio nitratireducens ONR
0.1

**Table S2.** Gene annotations associated with Figures 1 and 2. Numbers correspond to those in the figures.

| **Enzyme Number** | Gene | Product | EC Number |
|---|---|---|---|
| **Glycolosis** | | | |
| **1** | HK | hexokinase | 2.7.1.1 |
| **2** | pmm-pgm | phosphoglucomutase | 5.4.2.2 |
| **3** | pgi-pmi | glucose-6-phosphate isomerase | 5.3.1.9 |
| **4** | pfkA | 6-phosphofructokinase 1 | 2.7.1.11 |
| **5** | fbp | fructose-1,6-bisphosphatase I | 3.1.3.11 |
| **6** | K01622 | fructose 1,6-bisphosphate aldolase/phosphatase | 4.1.2.13 |
| **7** | tpiA | triosephosphate isomerase | 5.3.1.1 |
| **8** | gapA | glyceraldehyde 3-phosphate dehydrogenase | 1.2.1.12 |
| **9** | pgk | phosphoglycerate kinase | 2.7.2.3 |
| **10** | apgM | 2,3-bisphosphoglycerate-indipendent phosphoglycerate mutase | 5.4.2.12 |
| **11** | eno | enolase | 4.2.1.11 |
| **12** | ppsA | Phosphoenolpyruvate synthase/pyruvate phosphate dikinase | 2.7.9.2 |
| **13** | pyk | pyruvate kinase | 2.7.1.40 |
| **14** | porA | pyruvate ferredoxin oxidoreductase, alpha subunit | 1.2.7.1 |
| **15** | aceE | pyruvate dehydrogenase E1 component alpha subunit | 1.2.4.1 |
| **16** | aceF | pyruvate dehydrogenase E2 component | 2.3.1.12 |
| **17** | E6.2.1.13 | acetyl-CoA synthetase (ADP-forming) | 6.2.1.12 |
| **18** | acs | acetyl-CoA synthetase | 6.2.1.1 |
| **19** | ALDH | aldehyde dehydrogenase | 1.2.1.3 |
| **20** | ADH1 | alcohol dehydrogenase | 1.1.1.1 |
| **21** | pckA | phosphoenolpyruvate carboxykinase | 4.1.1.32 |
| **22** | LDH | lactate dehydrogenase | 1.1.1.27 |
| **23** | G6PD | glucose-6-phosphate 1-dehydrogenase | 1.1.1.49 |
| **24** | PGLS | 6-phosphogluconolactonase | 3.1.1.31 |
| **25** | PGD | 6-phosphogluconate dehydrogenase | 1.1.1.44 |
| **26** | rpiA | ribose 5-phosphate isomerase | 5.3.1.6 |
| **27** | rpe | ribulose-phosphate 3-epimerase | 5.1.3.1 |
| **28** | tktA,tktB | transketolase | 2.2.1.1 |
| **29** | talA,talB | transaldolase | 2.2.1.2 |
| **30** | PRK | phosphoribulokinase | 2.7.1.19 |
| **31** | rbcL | ribulose-bisphosphate carboxylase | 4.1.1.39 |
| **Citrate Acid Cycle** | | | |
| **32** | pyc | pyruvate carboxylase | 6.4.1.1 |
| **33** | CS,gltA | citrate synthase | 2.3.3.1 |
| **34** | acnA | aconitate hydratase | 4.1.2.3 |

| | | | |
|---|---|---|---|
| **35** | icd | isocitrate dehydrogenase | 1.1.1.42 |
| **36** | OGDH,DLST,korA | 2-oxoglutarate synthase | |
| **37** | LSC1 | succinyl-CoA synthetase | 6.2.1.4 |
| **38** | SDHA,SDH1 | succinate dehydrogenase | 1.3.5.1 |
| **39** | fumA, fumb | fumarate hydratase | 4.2.1.2 |
| **40** | MDH1,mgo | malate dehydrogenase | 1.1.1.37, 1.1.5.4 |
| **41** | maeB | malic enzyme | 1.1.1.40 |
| **Oxidative phosphorylation** | | | |
| **42** | | NADH dehydrogenase | 1.6.5.3 |
| **38** | SDHA,SDH1 | succinate dehydrogenase | 1.3.5.1 |
| **43** | fbcH | cytochrome c reductase | 1.10.2.2. |
| **44** | cco/cyo | cytochrome c oxidase | 1.9.3.1 |
| **45** | cox | cytochrome c oxidase | |
| **46** | ATP | H+-transporting ATPase | 3.6.3.6 |
| **Sulfur Metabolism** | | | |
| **47** | PAPS/Sat | 3'-phosphoadenosine 5'-phosphosulfate synthase | 2.7.7.4 |
| **48** | | phosphoadenylylsulfatase | 3.6.2.2 |
| **49** | cysH | phosphoadenosine phosphosulfate reductase | |
| **50** | | sulfite dehydrogenase | 1.8.2.1 |
| **51** | | sulfite reductase | |
| **52** | SOX | sulfur-oxidizing protein | |
| **53** | TST | thiosulfate/3-mercaptopyruvate sulfurtransferase | 2.8.1.1 |
| **54** | phsA | thiosulfate reductase | 1.8.5.5 |
| **55** | fccB | sulfide dehydrogenase | 1.8.2.3 |
| **56** | ETHE1 | sulfur dioxygenase | 1.13.11.18 |
| **57** | ssuD | alkanesulfonate monooxygenase | 1.14.14.5 |
| **58** | doxA | thiosulfate dehydrogenase | 1.8.5.2 |
| **59** | cysM | cysteine synthase | 2.5.1.47 |
| **Nitrogen Metabolism** | | | |
| **60** | narB | assimilatory nitrate reductase | 1.7.7.2 |
| **61** | narHI, napAB | dissimilatory nitrate reductase | 1.7.99.4 |
| **62** | nirB/nirD/nirK | assimilatory nitrite reductase | 1.7.7.1 |
| **63** | nrfAH | dissimilatory nitrite reductase | 1.7.2.2 |
| **64** | nirK | nitrite reductase (NO-forming) | 1.7.2.1 |
| **65** | norB | nitric oxide reductase | 1.7.2.5 |
| **66** | nosZ | nitrous-oxide reductase | 1.7.2.4 |
| **67** | nifD | nitrogenase | 1.18.6.1 |
| **68** | hao | hydroxylamine dehydrogenase | 1.7.2.6 |
| **69** | pmoA-amoA | methane/ammonia monooxygenase | 1.14.18.3 1. 14.99.39 |

| | | | |
|---|---|---|---|
| **70** | NAO | nitroalkane oxidase | 1.7.3.1 |
| **71** | | nitrilase | 3.5.5.1 |
| **72** | gudB | glutamate dehydrogenase | 1.4.1.2 |
| **73** | CPS1 | carbamoyl-phosphate synthase | 6.3.4.16 |
| **Methane Metabolism** | | | |
| **74** | | formamidase | 3.5.1.49 |
| **75** | FDH | formate dehydrogenase | 1.2.1.2 |
| **76** | fdhA | formate dehydrogenase alpha subunit | 1.2.1.43 |
| **77** | cdhA | acetyl-CoA decarbonylase/synthase complex | 1.2.7.4 |
| **78** | coxMS | carbon-monoxide dehydrogenase | 1.2.7.4 |
| **Wood-Ljungdahl Pathway** | | | |
| **78** | coxMS | carbon-monoxide dehydrogenase | 1.2.7.4 |
| **79** | fdhA | formate dehydrogenase alpha subunit | 1.2.1.43 |
| **80** | fhs | formate--tetrahydrofolate ligase | 6.3.4.3 |
| **81** | folD | methylenetetrahydrofolate dehydrogenase (NADP+) / methenyltetrahydrofolate cyclohydrolase | 1.5.1.5<br>3.5.4.9 |
| **82** | metF | methylenetetrahydrofolate reductase (NADPH) | 1.5.1.20 |
| **83** | acsE | 5-methyltetrahydrofolate corrinoid/iron sulfur protein methyltransferase | 2.1.1.258 |
| **84** | acsB | acetyl-CoA synthase | 2.3.1.169 |
| **Chemotaxis** | | | |
| **85** | mcp | methyl-accepting chemotaxis protein | |
| **86** | cheA | sensor kinase CheA | 2.7.13.3 |
| **87** | chW | purine-binding chemotaxis protein CheW | |
| **88** | cheV | response regulator CheV | |
| **89** | cheB | response regulator CheB | 3.1.1.61 |
| **90** | cheR | chemotaxis protein methyltransferase CheR | 2.1.1.80 |
| **91** | cheY | response regulator CheY | |
| **92** | cheC | chemotaxis protein CheC | |
| **93** | chex | chemotaxis protein CheX | |
| **94** | cheD | chemotaxis protein CheD | 3.5.1.44 |
| **Flagellar assembly** | | | |
| **95** | | Motor/Switch | |
| **96** | | Basal body | |
| **97** | | Hook | |
| **98** | | Filament | |
| **Phosphate response regulator** | | | |
| **99** | SenX3 | sensor histidine kinase SenX3 | 2.7.13.3 |
| **100** | PhoR | phosphate regulon sensor histidine kinase PhoR | 2.7.13.3 |
| **101** | RegX3 | response regulator RegX3 | |

| | | | |
|---|---|---|---|
| **102** | PhoB | phosphate regulon response regulator PhoB | |
| **103** | PhoP | alkaline phosphatase synthesis response regulator PhoP | |
| **104** | PstS | phosphate transport system substrate-binding protein | |
| **105** | PhoA | alkaline phosphatase | 3.1.3.1 |
| **106** | PhoD | alkaline phosphatase D | 3.1.3.1 |
| **Secretion System** | | | |
| **107** | SecDFEGY | protein translocase subunit | |
| **108** | SecA | protein translocase subunit secA | |
| **109** | FtsY | signal recognition particle-docking protein FtsY | |
| **110** | TatAC | sec-independent protein translocase protein | |
| **Transport** | | | |
| **111** | pstS | phosphate transport system substrate-binding protein | |
| **112** | pstC | phosphate transport system permease protein | |
| **113** | pstB | phosphate transport system ATP-binding protein | 3.6.3.27 |
| **114** | | Specific sugar transport system substrate-binding protein | |
| **115** | | specific sugar transport system permease protein | |
| **116** | msmX, msmK, malK, sugC, ggtA, msiK, mtlK, thuK | multiple sugar transport system ATP-binding protein | 3.6.3.- |
| **117** | OpuBC | osmoprotectant transport system substrate-binding protein | |
| **118** | OpuBB | osmoprotectant transport system permease protein | |
| **119** | OpuBA | osmoprotectant transport system ATP-binding protein | |
| **120** | LolC/LolE | lipoprotein-releasing system permease protein | |
| **121** | LolD | lipoprotein-releasing system ATP-binding protein | 3.6.3.- |
| **122** | AfuA | iron(III) transport system substrate-binding protein | |
| **123** | AfuB | iron(III) transport system permease protein | |
| **124** | AfuC | iron(III) transport system ATP-binding protein | 3.6.3.30 |
| **125** | BioY | biotin transport system substrate-specific component | |
| **126** | BioN | biotin transport system permease protein | |
| **127** | BioM | biotin transport system ATP-binding protein | 3.6.3.- |
| **128** | NatB | Sodium transport system permease protein | |
| **129** | NatA | sodium transport system ATP-binding protein | |
| **130** | CcmD | heme exporter protein D | |
| **131** | CcmC/CcmB | heme exporter protein C/B | |
| **132** | CcmA | heme exporter protein A | 3.6.3.41 |
| **133** | modA | molybdate transport system substrate-binding protein | |
| **134** | modB | molybdate transport system permease protein | |
| **135** | modC/modF | molybdate transport system ATP-binding protein | 3.6.3.29 |
| **136** | ZnuA | Zinc transport system substrate-binding protein | |
| **137** | ZnuB | Zinc transport system permease protein | |
| **138** | ZnuC | Zinc transport system ATP-binding protein | |
| **139** | LivK | branched-chain amino acid transport system substrate-binding protein | |

| | | | |
|---|---|---|---|
| **140** | LivH/LivM | branched-chain amino acid transport system permease protein | |
| **141** | LivG/LivF | branched-chain amino acid transport system ATP-binding protein | |
| **142** | PotD | spermidine/putrescine transport system substrate-binding protein | |
| **143** | potC, potB | spermidine/putrescine transport system permease protein | |
| **144** | PotA | spermidine/putrescine transport system ATP-binding protein | 3.6.3.31 |
| **145** | TupA | tungstate transport system substrate-binding protein | |
| **146** | TupB | tungstate transport system permease protein | |
| **147** | TupC | tungstate transport system ATP-binding protein | 3.6.3.55 |
| **148** | GlnH | glutamine transport system substrate-binding protein | |
| **149** | GlnP | glutamine transport system permease protein | |
| **150** | GlnQ | glutamine transport system ATP-binding protein | 3.6.3.- |
| **151** | tbpA | thiamine transport system substrate-binding protein | |
| **152** | thiP | thiamine transport system permease protein | |
| **153** | thiQ | thiamine transport system ATP-binding protein | |
| **154** | PhnD | phosphonate transport system substrate-binding protein | |
| **155** | PhnE | phosphonate transport system permease protein | |
| **156** | PhnC | phosphonate transport system ATP-binding protein | |
| **157** | TagG | teichoic acid transport system permease protein | |
| **158** | TagH | teichoic acid transport system ATP-binding protein | |
| **159** | RfbA | lipopolysaccharide transport system permease protein | |
| **160** | RfbB | lipopolysaccharide transport system ATP-binding protein | |
| **161** | OppA | oligopeptide transport system substrate-binding protein | |
| **162** | OppB/ OPPC | oligopeptide transport system permease protein | |
| **163** | OppD/ OppF | oligopeptide transport system ATP-binding protein | |
| **164** | thiY | putative hydroxymethylpyrimidine transport system substrate-binding protein | |
| **165** | thiX | putative hydroxymethylpyrimidine transport system permease protein | |
| **166** | thiZ | putative hydroxymethylpyrimidine transport system ATP-binding protein | |
| **167** | cbiN/cbiM/cibQ | cobalt/nickel transport system permease protein | |
| **168** | cbiO | cobalt/nickel transport system permease protein | |
| **169** | YejA | microcin C transport system substrate-binding protein | |
| **170** | YejB/YejE/YejF | microcin C transport system permease protein | |
| **171** | YejF | microcin C transport system ATP-binding protein | |
| **172** | NodJ | lipooligosaccharide transport system permease protein | |
| **173** | NodI | lipooligosaccharide transport system ATP-binding protein | |
| **174** | UgpB | sn-glycerol 3-phosphate transport system substrate-binding protein | |
| **175** | UgpA/UgpE | sn-glycerol 3-phosphate transport system permease protein | |
| **176** | UgpC | sn-glycerol 3-phosphate transport system ATP-binding protein | 3.6.3.20 |

| | | | |
|---|---|---|---|
| **177** | NNP | MFS transporter, NNP family, nitrate/nitrite transporter | |
| **Pilus Assembly** | | | |
| **178** | PilQ | type IV pilus assembly protein PilQ | |
| **179** | PilC | type IV pilus assembly protein PilC | |
| **180** | PilB/PilT/PilU | type IV pilus assembly protein PilB/PilT/PilU | |
| **181** | Pil F | type IV pilus assembly protein PilF | |
| **182** | PilP/PilN/PilO | type IV pilus assembly protein PilP/PilN/PilO | |
| **183** | PilM | type IV pilus assembly protein PilM | |
| **184** | PilA/FimV | type IV pilus assembly protein PilA/FimV | |