Decoding bacterioplankton metabolism 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], 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: thrashc@lsu.edu

Short title: Decoding microbes of the Dead Zone

Abstract

Marine regions that experience low dissolved oxygen (DO) from seasonal to long-term time scales, a.k.a. dead zones, are increasing in number and severity around the globe with deleterious effects on ecology and economics. One of the largest of these occurs on the continental shelf of the northern Gulf of Mexico (nGOM) as a result of eutrophication-enhanced bacterioplankton respiration and strong stratification. The effects of this perturbation on microbial assemblages, and therefore the underlying potential for biogeochemical cycling, have only begun to be explored. Here we present predicted roles for multiple organisms in various phases of nitrogen, carbon, and sulfur cycling and evidence for electron donor-based niche partitioning based on more than 70 high-quality genomes (33 of which are > 70% complete) reconstructed from whole community metagenomic data in the 2013 nGOM dead zone and metatranscriptomic expression data. All but two genomes could be classified into 17 named bacterial and archaeal phyla, and many represented the most abundant organisms in the dead zone (e.g. Thaumarchaeota, *Synechococcus*), some of which include members of the uncultivated "microbial dark matter" (e.g. Marine group II Euryarchaeota, SAR406). Surprisingly, we also recovered near complete genomes belonging to Candidate Phyla that are usually associated with anoxic environments: Parcubacteria (OD1), Peregrinibacteria, Latescibacteria (WS3), and ACD39. This work provides an important biogeochemical map for multiple phyla that will help to resolve the impacts of hypoxia on nutrient flow in the dead zone.

Introduction

Dissolved oxygen (DO) below 2 mg L^{-1} (~62.5 μM- "hypoxia") becomes dangerous or lethal to a wide variety of marine life, including organisms of economic importance[1]. Hypoxia results from 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]. Dead zones are increasing in number and frequency as use of nitrogen-based fertilizers increases with a concomitant increase in transport to coastal oceans via runoff [2]. In the nGOM, nitrogen deposition from the Mississippi and Atchafalaya Rivers leads to one of the world's largest seasonal dead zones, with bottom water hypoxia that can extend over 20,000 km^2 [1]. Action plans to mitigate the nGOM dead zone 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 dead zone nutrient transformation: microorganisms.

Much of our current knowledge regarding microbial contributions to regions of low DO comes from numerous reports describing 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, the nGOM dead zone 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 dead 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 in the dead zone showed Thaumarchaea dominated[21] in low DO waters and could be highly active[22]. Of the thousands of operational taxonomic units (OTUs), only 15 OTU abundances correlated with DO (either positively or negatively), and many of the most abundant taxa were expected aerobes based on taxonomic affiliation[21]. Further, *amoA* and Thaumarchaeal 16S rRNA gene copy per L of seawater inversely correlated with DO. Consistently, aerobic ammonia oxidation, not nitrate reduction, was the dominant source of nitrite in the 2012 dead zone[22]. Nitrite accumulation in waters even with low DO was unusually high and possibly due to a decoupling of ammonia- and nitrite-oxidation[22]. Thus the nGOM dead zone has unique microbiological features that echo its physico-chemical distinctions.

However, despite being one of the largest zones of eutrophication-based hypoxia in the world and its relevance to comparisons with global oxygen minimum zones, we still know comparatively little about the microbial ecology and physiology within the nGOM dead zone. To improve our understanding of the metabolic roles played by dead zone taxa, we reconstructed nearly 80 bacterioplankton genomes from metagenomic sequences across six sampling sites

87 spanning DO concentrations of 2.2 - 128.8 μM[21], and mapped metatranscriptomic sequences to
88 measure gene and genome activity. We confirm that many of the most dominant and active taxa
89 have predicted aerobic metabolism, implicate many organisms in the nitrogen, sulfur, and carbon
90 cycles, recover members of Candidate Phyla (CP) that have gone previously unreported in the
91 system, and provide evidence for electron donor-based niche partitioning.
92
93 **Results**
94 ***Metagenomic assembly yielded high quality genomes from a wide range of taxa***
95 We recovered a total of 77 bacterioplankton genomes from two separate assemblies (Table 1). 33
96 had predicted completeness > 70%, and only six had predicted contamination > 6% (Table S1).
97 The genomes belonged to 17 bacterial and archaeal phyla, and in some cases, multiple subclades
98 within a phylum (Table 1). Multiple genomes represented dominant taxa previously observed
99 with 16S rRNA gene surveys[21]: two Thaumarchaeota genomes that grouped closely with
100 *Nitrosopumilus* taxa, six Marine Group II Euryarchaeota (MGII), three *Synechococcus*, five
101 *Nitrospina* and SAR406 (Marinimicrobia) genomes, and several species from the *Alpha*- and
102 *Gammaproteobacteria* (Figs S1, S2). Actinobacteria, Proteobacteria, and Planctomycetes were
103 represented by the largest number of genomes, although most in the latter two phyla had low
104 predicted completion percentages (Table S1). None of the Planctomycetes genomes grouped
105 with known anammox bacteria in the *Brocadiaceae* that play important roles in some OMZs[9,11],
106 but instead grouped with either the *Planctomycetaceae* or *Phycisphaerae* (Fig. S1). In general,
107 all of the most abundant phyla detected with 16S rRNA gene analysis in Gillies *et al.* (2015)
108 were represented by at least three genomes, and there were many from less abundant phyla,
109 including *Gemmatimonadetes* and *Ignavibacteriae* (Table 1). Notably, we recovered genomes
110 from several uncultivated microbial "dark matter" clades. In addition to the uncultivated
111 SAR406s and MGIIs, we reconstructed three SAR202 genomes and six genomes from three
112 different Candidate Phyla (CP): Parcubacteria (formerly OD1- specifically *Candidatus*
113 Uhrbacteria), Peregrinibacteria, Latescibacteria (WS3), and one putatively assigned to ACD39
114 (Fig. S1, S3).
115
116 ***Roles in nitrogen and sulfur cycling***
117 We identified potential bacterioplankton contributions to every phase of inorganic nitrogen
118 cycling except nitrogen fixation (Fig. 1). *Nitrosopumilus* and *Nitrospina* spp. exclusively
119 occupied roles in ammonia and nitrite oxidation, respectively, consistent with previous
120 observations in the dead zone[21,22]. The Bin 71 *Nitrosopumilus* genome has predicted *amoABC*
121 genes in a syntenic arrangement to that of *Nitrosopumilus maritimus* SCM1 (Walker, 2010), and
122 we observed transcripts from these genes in all samples, supporting the view of active
123 nitrification. *Nitrosopumilus* spp. dominated the lowest DO sites (Fig. 2), and showed
124 proportionate activity, based on rank-residuals (see methods) for RNA vs. DNA recruitment rank
125 (Fig. 3, Rows 21 & 28). In contrast, *Nitrospina* spp. displayed much lower abundances (Fig. 2),
126 but one genome (Bin 25) demonstrated high activity relative to other taxa across all samples

127 (Fig. 3, Row 75). Bin 25 was also the only genome with a nitrite oxidase *nxrA*[23], and we
128 observed expression across all samples with reliable data (Table S1). Active, but low abundance
129 *Nitrospina* corroborate Bristow *et al.* (2015), who found low overall expression of *nxrA*, with all
130 transcripts predicted to belong to *Nitrospina* spp.. Nitrite levels and temperatures in our samples
131 also support their suggestion that *Nitrospina* spp. may have been nitrite and/or temperature
132 limited[22,24], thus leading to low rates of nitrite- relative to ammonia-oxidation and an
133 accumulation of nitrite to micromolar concentrations- an unusual phenomenon in a system not
134 fully anoxic.
135 The 2013 dead zone data show a similar nitrite accumulation to micromolar
136 concentrations (Table S1). Abundant nitrite may have been reduced along two different paths.
137 Two each of Bacteriodetes, Latescibacteria, and SAR406 genomes had predicted *nrfAH* genes
138 for nitrite reduction to ammonia (Table S1). We observed activity for these taxa (Fig. 3) and the
139 *nrfAH* genes (Fig. 4), indicating a possible nitrite/ammonia metabolic loop in three of the four
140 hypoxic sites. Alternatively, multiple genomes contained predicted *nirK* and *nirS* genes for
141 reduction of nitrite to nitric oxide. Aggregate expression of *nirK* genes, found primarily in the
142 *Nitrospina* and *Nitrosopumilus* genomes, but also in one *Flavobacteriaceae* and one MGII
143 genome, exceeded that of all other reductive nitrogen cycle genes (Fig. 4). The presence of *nirK*
144 in the absence of other denitrification genes may indicate a nitrite detoxification method[25,26], or
145 simply an alternative terminal electron accepting process during low DO[27].
146 We observed roles for many organisms in other parts of the nitrogen cycle as well.
147 Although multiple genomes contained genes for nitrate reduction, and we observed expression of
148 *narGH* in most samples (Fig. 4), complete denitrification capability was only predicted for one
149 *Flavobacteriaceae* genome (Bin 10, *Polaribacter* sp.) (Fig. 1). Also, we only observed
150 expression of all required genes (*nar*, *nap*, *nir*, *nor*, *nos*) in the lowest DO sample (E2A- Fig. 4).
151 However, our findings suggest denitrification can occur via multiple taxa in concert, and many
152 organisms appeared poised to reduce nitrous oxide via *nosZ* genes (Fig. 1). The pervasive *nosZ*
153 genes in the absence of other parts of the denitrification pathway suggest that these taxa have
154 evolved to reduce nitrous oxide as a terminal electron acceptor. Given the paucity of genomes
155 with nitric oxide reductases, production of nitrous oxide via *Nitrosopumilus*[28,29] seems plausible
156 as the dominant source. We detected *nosZ* transcripts in nearly all samples (Fig. 4). A number of
157 genomes also had predicted capacities to convert organic nitrogen compounds to inorganic
158 sources of nitrogen (Fig. 1). Importantly, *Synechococcus* and one *Rhodobacteraceae* genome,
159 closely related to *Donghicola* spp., encoded cyanate lyases corroborating previous reports of
160 similar organisms using cyanate as a nitrogen source[30,31].
161 Fewer organisms had putative pathways for sulfur cycling (Fig. 1). *Chromatiales*,
162 *Rhodobacteraceae*, and *Rhodospirillales* genomes contained *dsrAB* genes (Table S1), predicted
163 by phylogenetic analysis[32] to operate oxidatively (Fig. S4). Genomes in these same clades also
164 contained *aprAB* and *sat* genes for sulfate/sulfite conversion, although based on the *dsrAB* data,
165 *aprAB* likely function in reverse as well. We observed putative thiosulfate oxidation genes in the
166 same groups and the *Gemmatimonadaceae*. Thus, sulfur oxidation was phylogenetically

167 restricted. *Chromatiales* taxa were abundant in hypoxic samples (Fig. 2) similarly to reports in
168 the ETNP OMZ[33]. However, these taxa contributed fewer transcripts than would be expected
169 based on DNA abundance (Fig. 3) suggesting a limited influence. Two SAR406 genomes had
170 predicted genes for thiosulfate reduction to sulfide (and/or polysulfide reduction[34])- we detected
171 transcripts for these genes only in samples E2A and E4 (Table S1).
172
173 ***Illuminating metabolism within microbial dark matter***
174 Several recovered genomes belonged to clades with no cultured representatives: Marinimicrobia,
175 MGII, SAR202, and the CP Parcubacteria (OD1), Peregrinibacteria, and Latescibacteria (WS3).
176 Marinimicrobia (a.k.a. SAR406 and Marine Group A) are widely distributed[35,36] and frequently
177 observed in oxygen minimum zones[5,37], but nevertheless remain poorly understood from a
178 functional perspective. Four Marinimicrobia genomes (estimated at > 73% complete) informed
179 our metabolic reconstruction of this group (Figs. 5, S5). SAR406 genomes have both high and
180 low-affinity cytochrome c oxidases[38] (Table S1) extending previous observations of adaptive
181 traits for a range of DO concentrations[37]. This is consistent with their predominance in lower DO
182 samples from the dead zone (Fig. 2). The genomes also encode potential for
183 thiosulfate/polysulfide reduction to sulfide as previously observed[37], however we additionally
184 predict capabilities in nitrate reduction, nitrous oxide reduction, and nitrite reduction to ammonia
185 (Fig. 1, Table S1). The genomes contain putative transporters for a variety of dissolved organic
186 matter (DOM) components including nucleosides, amino and fatty acids, and oligopeptides.
187 They also contain an enrichment of both glycoside hydrolases (GH) and carbohydrate binding
188 modules (CBM) (Table S1), the majority of which have secretion signals indicating they target
189 external carbohydrate nutrients. This supports previous predictions for organic matter
190 degradation by these taxa[39]. Thus, Marinimicrobia likely perform chemoorganoheterotrophic
191 metabolism with flexibility in terminal electron accepting processes that implicate them in
192 nitrogen and sulfur cycling.
193 Other well-known, but still as yet uncultivated, taxa include SAR202 and MGII.
194 Members of the SAR202 clade of Chloroflexi also inhabit a wide variety of marine
195 environments[40], frequently in deeper waters[40-42], but remain functionally obscure. We recovered
196 three SAR202 genomes (estimated > 83% complete) that, although present at low abundance
197 (Fig. 2), nevertheless showed relatively high activity in some samples (Fig. 3). These organisms
198 likely respire oxygen, however nitrate and nitrous oxide reductases were recovered in two
199 genomes, indicating a potential for utilization of alternative terminal electron acceptors (Table
200 S1). While transporter predictions support previous observations of DOM uptake[41], the SAR202
201 genomes also contained predicted copies of 2-oxogluterate synthase and citryl-CoA lyase,
202 suggesting a possible C-fixation route via the reductive TCA cycle. Our MGII reconstructions
203 largely corroborate previous descriptions[43,44] of these organisms as aerobic
204 chemoorganoheterotrophs with transporters for a variety of DOM compounds (Table S1).
205 Recovery of Parcubacteria from a coastal marine system is unusual, but not
206 unprecedented. Parcubacteria SAGs have been identified from marine and brackish sources[39].

207 However, Peregrinibacteria have thus far only been found in terrestrial subsurface aquifers[45-48].
208 Consistent with previous reports of obligate fermentative metabolism by Parcubacteria and
209 Peregrinibacteria[39,45,47,49], we identified no respiratory pathways for these taxa and they trended
210 towards greater abundances in the lowest DO samples (Fig. 2). Furthermore, we found no genes
211 linking them to substantial roles in nitrogen or sulfur cycling.
212 Latescibacteria have been proposed to have a particle-associated, strictly anaerobic
213 lifestyle, with many genes available for the breakdown of complex algal cell wall components
214 and utilization of the resultant monomers[39,50]. While our carbon substrate predictions and
215 flagellar gene identification largely substantiate these findings, the dead zone Latescibacteria
216 encode cytochrome c oxidases and other elements of aerobic electron transport chains (Table
217 S1). The discrepancy in predicted physiological roles between these and previously identified
218 Latescibacteria may be due to the incompleteness of the single-cell genomes[39] or perhaps
219 horizontal gene transfer causing differentiation within the phylum. Indeed, the best BLASTP hits
220 to the cytochrome c oxidase subunits in the Bin 50 genome matched an *Acidobacteria* strain
221 (Table S1). While the possibility also exists that the entire contig was inappropriately binned, the
222 remaining genes on the contig do not support its assignment to another specific taxon (Table S1).
223 The even distribution of the dead zone Latescibacteria across varied DO concentrations also
224 supports their predicted aerobic capability (Fig. 2), but this hypothesis needs further
225 investigation.
226 One of the genomes with the highest transcriptional activity across samples was Bin 13
227 (Fig. 3), which grouped with CP ACD39 taxa (Fig. S1). However, the phylogenetic position was
228 poorly supported, so we consider this designation putative. The genome for this organism
229 encodes a predicted obligate aerobic chemoheterotrophic metabolism, with no designated roles in
230 the nitrogen or sulfur cycles (Fig. S7). Its high activity indicated the importance of a DO
231 consuming, organic carbon-based lifestyle even in low DO waters.
232
233 ***Evidence for electron donor-based niche partitioning***
234 Although some of the recovered CP taxa likely subsist via an obligate anaerobic lifestyle, the
235 majority of groups in the dead zone had one or more genomes that encoded cytochrome c
236 oxidases (and other electron transport chain components) for respiring oxygen (Table S1).
237 Pervasive aerobic metabolism in an area of depleted oxygen seems counterintuitive, yet despite
238 DO being as low as 0.2 μM in the E2A sample, oxygen likely remained prevalent enough to
239 sustain aerobic microbes. As little as 0.2-0.3 μM oxygen inhibited denitrification in OMZ
240 populations by 50%[51], and Bristow *et al.* (2015) experimentally verified that at levels as low as
241 20 μM DO, aerobic ammonia oxidation by marine Thaumarchaea, rather than nitrate reduction,
242 led to nitrite accumulation in 2012 nGOM dead zone bottom waters. Most of the taxa responsible
243 for the largest number of transcripts (proportionally- Fig. 3, stars) were from groups of predicted
244 aerobes, and the only groups that could be assigned dissimilatory nitrate reduction genes (e.g.
245 Bin 10 (Flavobacteriaceae), Bin 51-1 (Marinimicrobia), etc.) also had cytochrome c oxidases
246 (Table S1), making these taxa likely only facultative anaerobes. Based on overall pathway

247 reconstruction and gene expression, the likely options for external terminal electron acceptors
248 appeared limited to oxygen, nitrate, nitrite, or nitrous oxide. Indeed, we observed expression of
249 cytochrome c oxidase genes at similar or greater levels than nitrate reductases (Fig. 4), indicating
250 active aerobic respiration in dead zone bacterioplankton, even in dysoxic (1-20 μM DO[5]) water.
251 Although we also observed *nirK* transcripts at higher levels than any of these other gene sets,
252 concurrent expression of all genes indicates possible co-reduction of multiple compounds, which
253 can occur in non-limiting electron donor circumstances[52].
254 In contrast to the limited number of terminal electron-accepting processes available,
255 electron-donating processes occurred in much greater variety, spanning different forms of both
256 lithotrophy and organotrophy. The dominant Thaumarchaea were predicted ammonia-oxidizing
257 chemolithotrophs and one of the most active genomes, Bin 25, represented nitrite-oxidizing
258 *Nitrospina* spp. We identified the potential for sulfur chemolithotrophy in both *Gamma-*
259 (*Chromatiales*) and *Alphaproteobacteria* (*Rhodobacteraceae* and *Rhodospirillales*) genomes.
260 Furthermore, these genomes encoded a multitude of chemoorganoheterotrophic options. We
261 identified transporters for import of many different DOM components, from sugars to amino and
262 carboxylic acids, differentially distributed throughout the various taxonomic groups (e.g.,
263 oligopeptide transporter family vs. p-aminobenzoyl-glutamate transporter family- Table S1).
264 Within carbohydrate metabolism, CAZy profiles indicated other differential preferences. For
265 example, while as expected the chemolithotrophic *Nitrosopumilus* genomes contained no
266 predicted GH or polysaccharide lyases (PL) (e.g., see the CAZy database), Marinimicrobia
267 genomes were enriched in the former, and Latescibacteria genomes showed evidence for
268 enrichment in both (similarly to previous observations[50]). SAR202 genomes contained similar
269 numbers of GH and carbohydrate esterases. The comparative cornucopia of metabolic pathways
270 for electron-donating relative to electron-accepting processes strongly supports the hypothesis
271 that niche differentiation within dead zone bacterioplankton occurs predominantly via
272 specialization for different oxidizable substrates rather than for different roles in the important
273 redox cascade in OMZs[4,5]. Indeed, the diversity of compounds within DOM supplies myriad
274 opportunities for microbial differentiation in the chemoorganoheterotrophic space[53-56].
275
276 **Discussion**
277 Our results provide the first system-level genomic, and some of the most comprehensive
278 metabolic, reconstructions and assessment of active functions to date for bacterioplankton in the
279 nGOM dead zone. These genomes comprise the most abundant and/or active taxa observed
280 throughout the region in both 2012[22] and 2013[21] and CP taxa that had gone previously
281 undetected. The pervasiveness of aerobic respiration pathways and the limited number of
282 alternative terminal electron accepting processes supports a view that bacterioplankton likely
283 occupy different niches based on specialized oxidation of different electron donors, including
284 various DOM components.
285 This has distinct implications for dead zone creation, maintenance, and dissolution
286 because the amount of oxygen consumed by a given taxon will depend greatly on the kinetics of

287 the terminal oxidase(s), the reducing equivalents available from the electron donor of choice, and
288 in the case of organic compounds, the relative use of that substrate for growth vs. energy
289 (bacterial growth efficiency[57]). How changing electron donor affects the relative abundance and
290 DO consumption rates of these organisms will be an important area of future research.
291 Unfortunately, the relative oxygen consumption rates for different taxa across different substrates
292 has been cataloged for only a small number of taxa[58] and many important to the dead zone
293 remain uncultured. Customized cultivation strategies for nGOM bacterioplankton are showing
294 some success in this area[59], and can provide valuable additional data on respiration rates for
295 individual organisms. We also know little about the seasonal variations in bacterioplankton
296 assemblages in the nGOM, or how these relate to electron donor availability. Previous work
297 indicates, for example, that Thaumarchaea occur in lower abundances during spring[60,61], before
298 DO decreases, than they did during the 2013 dead zone. This means that these taxa become
299 competitive at a later seasonal stage, but whether that is caused by low DO, changes in
300 competitiveness based on electron donor availability, or some other factor is unknown. Indeed,
301 the lack of microbial time series data in the nGOM leaves many additional open questions for
302 these and other taxa. Which taxa respire the most oxygen? Are they abundant year round, or, if
303 not, when do they become important and what selects for this? How do changing DOM sources
304 and components affect respiration rates at the individual taxon and whole community scale?
305 What is the role of particle-associated taxa in DO consumption? Answers will provide greater
306 understanding of the dynamics resulting in low DO and deoxygenation trends in the world's
307 oceans, as well as provide context for comparative studies between dead zones and oxygen
308 minimum zones around the globe.
309
310
311 **Methods**
312 *Sample selection and nucleic acid processing.* Six samples representing hypoxic (n=4) and oxic
313 (n=2) DO concentrations were picked from among those previously reported[21] at stations D1,
314 D2, D3, E2, E2A, and E4 (Table S1). At these six stations 10 L of seawater was collected and
315 filtered with a peristaltic pump. A 2.7 µM Whatman GF/D pre-filter was used and samples were
316 concentrated on 0.22 µM Sterivex filters (EMD Millipore). Sterivex filters were sparged and
317 filled with RNAlater. DNA and RNA were extracted directly off of the filter by placing half of
318 the Sterivex filter in a Lysing matrix E (LME) glass/zirconia/silica beads Tube (MP Biomedicals,
319 Santa Ana, CA) using the protocol described in Gillies *et al.* (2015) which combines
320 phenol:chloroform:isoamyalcohol (25:24:1) and bead beating. Genomic DNA and RNA were
321 stored at -80°C until purified. DNA and RNA were purified using QIAGEN (Valencia, CA)
322 AllPrep DNA/RNA Kit. DNA quantity was determined using a Qubit2.0 Fluorometer (Life
323 Technologies, Grand Island, NY). RNA with an RNA integrity number (RIN) (16S/23S rRNA
324 ratio determined with the Agilent TapeStation) ≥ 8 (on a scale of 1-10, with 1 being degraded
325 and 10 being undegraded RNA) was selected for metatranscriptomic sequencing. Using a Ribo-
326 Zero kit (Illumina) rRNA was subtracted from total RNA. Subsequently, mRNA was reverse

327 transcribed to cDNA as described in Mason *et al.* (2012)[62]. RNA (as cDNA) and DNA libraries
328 (six samples/lane) were linker barcoded and pooled for high-throughput Illumina HiSeq 2000
329 paired-end sequencing at Argonne National Laboratory.
330
331 *Sequencing, assembly, and binning.* DNA and RNA were sequenced separately, six samples per
332 lane, with Illumina HiSeq 2000 chemistry to generate 100 bp, paired-end reads at the Argonne
333 National Laboratory Next Generation Sequencing facility. The data are available at the NCBI
334 SRA repository under the BioSample accession numbers SAMN05791315, SAMN05791316,
335 SAMN05791317, SAMN05791318, SAMN05791319, SAMN05791320, SAMN05791321,
336 SAMN05791322, SAMN05791323, SAMN05791324, SAMN05791325, SAMN05791326 (the
337 first six refer to DNA samples, the latter six, RNA). DNA sequencing resulted in a total of
338 416,924,120 reads that were quality trimmed to 413,094,662 reads. Prior to assembly, adaptors
339 were removed using Scythe (https://github.com/vsbuffalo/scythe), and low-quality reads (Q <
340 30) were trimmed with Sickle (https://github.com/najoshi/sickle). Reads with three or more Ns or
341 with average quality score of less than Q20 and a length < 50 bps were removed. Genomes were
342 reconstructed using two rounds of assembly. Metagenomic reads from all six samples were
343 pooled, assembled, and binned using previously described methods[63,64]. Briefly, quality filtered
344 reads were assembled with IDBA-UD[65] on a 1TB RAM, 40-core node at the LSU High
345 Performance Computing cluster SuperMikeII, using the following settings: -mink 65 –maxk 105
346 –step 10 –pre_correction –seed_kmer 55. Initial binning of the assembled fragments was
347 performed using tetra-nucleotide frequency signatures using 5 kbp fragments of the contigs.
348 Emergent self-organizing maps (ESOM) were manually delineated and curated based on clusters
349 within the map (Fig. S8, S9). The primary assembly utilized all reads and produced 28,080
350 contigs ≥ 3 kb totaling 217,715,956 bp. Of these, 303 contigs were over 50 kb, 72 over 100 kb,
351 and the largest contig was just under 495 kb. Binning produced 76 genomes, but the
352 Thaumarchaeota bin was very large and predicted to have high contamination from other taxa.
353 Assuming expected genome sizes of ~1.6 Mb and average relative abundances within the
354 community of 20% or greater, our initial sequencing run would have produced at least 5000x
355 coverage for a hypothetical Thaumarchaeota genome. To reduce this burden on the assembly
356 algorithm, a second round of assembly using a 5% random sub-selection of the pooled DNA
357 reads was completed using the same settings. The subset of reads was generated randomly using
358 the split command to create sub-files that were 1/20th the size of the original the metagenomic
359 file. A sub-file was chosen at random and was assembled using the parameters previously stated.
360 The subassembly resulted in a mixed Thaumarcheaota bin (Fig. S9) that could be separated in
361 two based on comparative read mapping from different samples (below). The final genomes
362 were nearly complete (96 and 79%, respectively), and grouped with other *Nitrosopumilus*
363 genomes (Table 1, Figs. S1, S2).
364
365 *DNA and RNA mapping.* Metagenomic and metatranscriptomic sequencing reads from each
366 sample were separately mapped to binned contigs using BWA[66] to compare bin abundance

367 across samples and facilitate bin cleanup (below). Contigs within each bin were concatenated
368 into a single fasta sequence and BWA was used to map the reads from each sample to all bins.
369 All commands used for these steps are available in supplementary information.
370
371 *Bin QC.* Bins were examined for contamination and completeness with CheckM[67], and we
372 attempted to clean bins with > 10% estimated contamination using a combination of methods.
373 First, the CheckM modify command removed contigs determined to be outliers by GC content,
374 coding density, and tetranucleotide frequency. In one case (Bin 51_l), the outlying contigs were
375 sufficiently similar to be formed into a separate bin. Next, in bins that still showed > 10%
376 contamination, contigs were separated according to comparative relative abundance of mean
377 DNA read coverage by sample. In three cases (Bins 33, 40, 62), this was insufficient to reduce
378 contamination. Final bins were evaluated with CheckM again to generate the statistics in Table
379 S1 and final bin placements in the CheckM concatenated gene tree (Fig. S2).
380
381 *Ribosomal protein tree.* The concatenated ribosomal protein tree was generated using 16
382 syntenic genes that have been shown to undergo limited lateral gene transfer (rpL2, 3, 4, 5, 6, 14,
383 15, 16, 18, 22, 24 and rpS3, 8, 10, 17, 19)[68]. Ribosomal proteins for each bin were identified with
384 Phylosift[69]. Amino acid alignments of the individual ribosomal proteins were generated using
385 MUSCLE[70] and trimmed using BMGE[71] (with the following settings: -m BLOSUM30 –g 0.5).
386 The curated alignments were then concatenated for phylogenetic analyses and phylogeny
387 inferred via RAxML v 8.2.8[72] with 100 bootstrap runs (with the following settings: mpirun -np 4
388 -npernode 1 raxmlHPC-HYBRID-AVX -f a -m PROTCATLG -T 16 -p 12345 -x 12345 -# 100).
389 Note this is similar to the number utilized in a previous publication for this tree with automated
390 bootstrapping[73], and required just over 56 hours of wall clock time. The alignment is available in
391 SI.
392
393 *Taxonomic assignment*. Taxonomy for each bin was assigned primarily using the ribosomal
394 protein tree. However, for bins that did not have enough ribosomal proteins to be included in the
395 tree, or for which the placement within the tree was poorly supported, assignments were made
396 based on the concatenated marker gene tree as part of the CheckM analysis (Fig. S2), or via 16S
397 rRNA gene sequences, when available. 16S rRNA genes were identified via CheckM, and these
398 sequences were blasted against the NCBI nr database to corroborate CheckM assignments.
399 Several of the most abundant OTUs in Gillies *et al.* (2015) matched the elusive and important
400 OM1 clade of marine Actinobacteria[74], so we explicitly included genome sequences in our
401 ribosomal protein tree from taxa that have been placed near this clade previously[75,76]. While
402 several of our genomes group with Illumatobacter or the MedAcidi sequences, these are not
403 expected to be true OM1 organisms, and 16S rRNA gene blast results from bins 29-1, 58, and 60
404 did not support classifying these taxa as OM1. However, their proximity to a close relative of the
405 OM1 clade, *Candidatus* Actinomarina minuta, make them useful contributions to future studies
406 of the group. In the case of the SAR202 genomes, which did not have representative genomes in

407 either the ribosomal protein tree or the CheckM tree, the 16S rRNA gene sequences for two of
408 the three bins (43-1, 43-2) were available and aligned with the sequences used to define the
409 SAR202 clade[40] (Fig. S10). Alignment, culling, and inference were completed with MUSCLE[70],
410 Gblocks[77], and FastTree2[78], respectively, with the FT_pipe script. The script is provided in SI.
411 For putative CP genomes, taxonomy was also evaluated by examining the taxonomic
412 identification for each of the predicted protein sequences after a BLASTP search against the
413 NCBI nr database. Post-blast, the number of assignments to the dominant one or two taxonomic
414 names, along with the number of assignments to "uncultured bacterium," was plotted for each
415 genome according to the bit score quartile. Quartiles were determined in R using the summary
416 function. In the case of the Parcubacteria and Peregrinibacteria genomes, the majority of high
417 quality blast hits were consistent with the ribosomal protein tree taxonomy (Figs. S1, S3). For the
418 Bin 13 and WS3 genomes, blast hit identification did not return clear results- hits included a
419 wide variety of taxa. This is consistent with what would be expected from novel genomes that
420 have few or no members in the database, however this analysis did not directly corroborate the
421 ribosomal protein tree phylogeny. Bin 56 has two ribosomal protein operons on
422 scaffold_2719/Ga0113622_1153 and scaffold_21777/Ga0113622_1009. In the ribosomal protein
423 tree, the former placed the organism in the *Planctomycetaceae*, while the latter (which was much
424 smaller) placed the organism in CP WS3. The majority of BLASTP annotations to the nr
425 database matched *Planctomycetaceae* taxa, as did the 16S rRNA gene sequences found in the
426 genome, so this organism was designated a Planctomycetes.
427
428 *Metabolic reconstruction*. Post-binning, genomes were submitted individually to IMG[79] for
429 annotation. Genome accession numbers are in Table S1, and all are publically available.
430 Metabolic reconstruction found in Table S1 and Figs S5-7/ S11-13, came from these annotations
431 and inspection with IMG's analysis tools, including KEGG pathway assignments and transporter
432 predictions. Transporters highlighted for DOM uptake were identified based on information at
433 the Transporter Classification Database[80]. Carbohydrate-active enzymes (CAZymes) were
434 predicted using the same routines as those in operation for the updates of the carbohydrate-active
435 enzymes database (www.cazy.org)[81].
436
437 *RPKM abundance of taxa and genes.* Abundance of taxa within the sample was quantified by
438 evaluating mapped reads using Reads-Per-Kilobase-Per-Million (RPKM) normalization[82]
439 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
440 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
441 total number of reads in sample j divided by 10^6. N_{ij} was generated using the samtools[66] idxstats
442 function after mapping with BWA. The data in Fig. 2 were created by summing (N_{ij}/L_i) for
443 groups of taxa defined in Table S1 prior to multiplying by ($1/T_j$). RNA coverage was used to
444 evaluate both bin and gene activity. Mean coverage for each supercontig was calculated using
445 bedtools[83] and bins were assigned a rank from lowest mean recruitment (1) to highest mean
446 recruitment (2). Bins with particularly high or low activity (transcript abundance) relative to their

447 abundance (genome abundance) were identified using rank-residuals, calculated as follows: On a
448 plot of DNA coverage rank vs. RNA coverage rank, residuals for each bin or gene were
449 calculated from the identity. As the rank-residuals followed a Gaussian distribution, bins with a
450 residual that was > 1 s.d. from the rank-residual mean were classified as having higher-than-
451 expected transcriptional activity; bins with a residual that was < 1 s.d. from the mean were
452 classified as having lower-than-expected transcriptional activity. RPKM values were also
453 calculated for every gene in every bin analogously to that for bins, using RNA mapping values
454 extracted with the bedtools multicov function. Sample E2 was omitted from gene-specific
455 calculations as only 4588 transcriptomic reads mapped successfully from this sample, compared
456 to >100,000 from other samples. 17,827 of 140,347 genes had no evidence of expression in any
457 sample and so were removed from further analysis. 3,840 genes recruited reads in all remaining
458 samples. All calculations are available in Table S1 or the R markdown document
459 Per.gene.RPKM.Rmd in Supplemental Information. Note that RPKM values indicate abundance
460 measurements across a small number of samples. While we can evaluate the relative expression
461 of genes for those samples, our dataset lacks sufficient power to evaluate estimates of
462 significance in differential expression.
463
464 *nxrA and dsrAB phylogeny.* Initial annotation of our bins identified putative nitrate reductases,
465 however none with explicit assignment as nxr genes putatively involved in nitrite oxidation,
466 which was unusual considering that *Nitrospina* spp. are nitrite-oxidizing chemolithotrophs. Since
467 *nxrA*-type nitrite oxidases are homologous with *narG*-type nitrate reductases (and
468 polyphyletic)[23], annotation for these genes was curated with phylogenetic analysis using known
469 *nxr* genes. Reference sequences of known *nxrA*/*narG* homologs from Lüke *et al.* (2015) were
470 searched against a database of all the binned scaffolds (makeblastdb –dbtype nucl with default
471 settings except -parse_seqids -hash_index) using tblastn with default settings (except –
472 max_query_seqs 10000). Nine bins had hits to any sequences in the reference file with bit scores
473 > 100. All annotated protein sequences were downloaded for these bins and used to create a blast
474 protein database (makeblastdb –dbtype prot with default settings except -parse_seqids -
475 hash_index); the reference sequences were searched through this database using blastp with
476 default settings (except –max_query_seqs 10000). All blast suite commands used v. 2.2.28+. The
477 eight proteins returning matches with bit scores > 100 were selected for phylogenetic analysis
478 using the original reference sequences. Alignment, culling, and inference were completed with
479 the FT_pipe script. The protein from *Nitrospina* Bin 25 grouped closely with the *nxrA* gene from
480 *N. gracilis* (Fig. S15), providing strong evidence for this being a functional nitrite oxidase. To
481 evaluate whether the predicted *dsrAB* genes in bins 23, 35, and 61-2 functioned most likely as
482 oxidative or reductive, they aligned with a previously established reference dataset[32] that was
483 used for phylogenetic inference also using the FT_pipe script.
484
485 **Acknowledgements**

Funding for this work was provided to JCT through the Oak Ridge Associated Universities Ralph E. Powe Junior Faculty Enhancement Award and the Louisiana State University Department of Biological Sciences. A portion of the funding for this work was provided by a Planning Grant award to OUM from Florida State University. Funding for the research vessel and collection of oceanographic data was provided by the National Oceanic and Atmospheric Administration, Center for Sponsored Coastal Ocean Research Award Number NA09NOS4780204 to NNR. The authors thank the crew of the R/V *Pelican*. Portions of this research were conducted with high performance computing resources provided by Louisiana State University (http://www.hpc.lsu.edu).

Author contributions

JCT, BJB, and OUM designed the study. LEG, NNR, and JCT collected samples. NNR provided processed oceanographic data. LEG and OUM extracted, quantified, and determined quality of nucleic acids. JCT, KWS, and BJB reconstructed the genomes. KWS, BJB, BT, BH, and JCT conducted downstream analyses. JCT lead manuscript writing and all co-authors evaluated and contributed edits.

Competing financial interests

The authors declare no competing financial interests.

505 **Table 1.** Summary of recovered genomes.
506

Domain	Phylum	Bins	Total	Subclades	Completion range	Contamination range
Bacteria	Actinobacteria	19, 21, 21-1, 26, 29, 29-1, 52, 57, 57-1, 58, 59, 60	12	"Near OM1" Rubrobacteridae	15 – 92%	0 – 5.5%
	Bacteroidetes	10, 12, 37, 41, 49, 64	6	Flavobacteriaceae Sphingobacteriales Rhodothermaceae	17 – 96%	0 – 5.3%
	Chloroflexi	43, 43-1, 43-2	3	SAR202	83 – 90%	0 – 4%
	Cyanobacteria	27, 28, 28-1	3	Synechococcus	44 – 89%	0.5 – 3.7%
	Gemmatimonadetes	42	1	Gemmatimonas	86%	2.2%
	Ignavibacteria	51 l, 51-2	2	Ignavibacteria	5 – 57%	0%
	Marinimicrobia (SAR406)	45, 45-1, 45-2, 51, 51-1	5		21 – 90%	0 – 5.5%
	Nitrospinae	24, 25, 47, 47-1, 47-2	5	Nitrospina	11 – 94%	0.3 – 5.2%
	Planctomycetes	20, 22, 31, 32, 56, 56-1, 56-2, 63, 63-1, 63-2	10	Phycisphaerae Planctomycetaceae	9 – 95%	0 – 7.0%
	Proteobacteria	23, 33, 35, 36, 53, 54, 55, 61, 61-2	9	OM50/NOR5 Chromatiales Rhodbacteraceae Rhodospirillales	22 – 93%	0 – 47 %
	Verrucomicrobia	11, 30, 34, 46, 65	5	Verrucomicrobiales	11 – 52%	0 – 1.8%
	CP ACD39*	13	1		90%	5.1%
	CP Parcubacteria (OD1)	40	1	Cand. Uhrbacteria	72%	15%**
	CP Peregrinibacteria	16, 39	2		50 – 83%	0.3%
	CP Latescibacteria (WS3)	48, 50	2		52 – 85%	1.4 – 2.2%
Archaea	Euryarchaeota	14, 15, 17, 17-1, 18, 38	6	Marine Group II	17 – 83%	0 – 1.6%
	Thaumarchaeota	71, 72	2	Nitrosopumilus	79 – 96%	5.2 – 6.3%
Unclassified		61-1, 62	2	N/A	3 – 46%	0 – 25%

507 *Putative assignment. Placement in the ribosomal protein tree was poorly supported.
508 ** 75% estimated to be due to strain heterogeneity
509
510

511 Figure Legends
512
513 **Figure 1.** N, S, and C cycling by dead zone taxa. Compounds are organized roughly according to
514 oxidation state, with predicted genes and taxa encoding those genes indicated at the arrows
515 between compounds. Conversions with no predicted taxa have red dashed lines. NMO =
516 nitronate monooxygenase.
517
518 **Figure 2.** Read-per-kilobase-per-million (RPKM) abundances of phylogenetically cohesive
519 groups by site. Sites are colored by DO concentration, according to the legend.
520
521 **Figure 3.** Relative DNA to RNA recruitment rank for each genome, by sample. Colors indicate
522 the relative difference in the ratio of rank based on total RNA and DNA mapping. Red indicates
523 a higher RNA recruitment rank compared to DNA recruitment rank, and vice-versa for blue. The
524 heatplot compresses data in Fig. S14 Asterisks and circles indicate bins where the rank-residual
525 from the identity in RNA vs. DNA read mapping was more or less than one standard deviation
526 beyond 0, respectively. Dendrograms were calculated using an Unweighted Pair Group Method
527 with Arithmetic Mean (UPGMA) from Euclidian distances of rank residuals across all samples
528 and bins.
529
530 **Figure 4.** RPKM values for predicted cytochrome c oxidase and reductive nitrogen cycle genes
531 and subunits. Boxplots display distribution (median, 1st & 3rd quartile, 1.5 x interquartile range)
532 of RPKM values for genes assigned to a given functional category. Individual gene RPKMs are
533 represented by colored dots for each sample.
534
535 **Figure 5.** Metabolic reconstruction for important uncultivated dead zone taxa. Solid arrows and
536 outlines indicate the presence of a predicted pathway or gene with that function. Dotted lines
537 indicate genes/pathways that were not found.
538
539 Supplemental Table Legend
540
541 **Table S1.** Spreadsheet (Table_S1.xlsx) containing information on taxonomy, CheckM results,
542 IMG statistics, partial metabolic reconstruction, transporter classifications, CAZy predictions,
543 sample chemical data RPKM values and WS3 cytochrome c oxidase gene neighborhoods.
544
545 Supplemental Figure Legends
546
547 **Figure S1.** Maximum likelihood tree of concatenated ribosomal protein coding genes. Values at
548 internal nodes indicate bootstrap support (n=100). Scale bar indicates changes per position.
549
550 **Figure S2.** Phylogenetic placement of bins based on CheckM.

551
552 **Figure S3.** Annotations of protein-coding gene sequence best blastp hits in the nr database,
553 divided into quartiles by bit score.
554
555 **Figure S4.** Phylogenetic tree of *dsrAB* genes. Genes from our bins are highlighted in pink.
556 Clades are labeled according to (Müller, 2015). Values at nodes indicate Shimodaira-Hasegawa
557 "like" values (Price, 2010). Scale bar indicates changes per position.
558
559 **Figure S5.** Metabolic reconstruction of SAR406 based on the four most complete genomes (Bins
560 45, 45-1, 45-2, and 51-1). Colors indicate recovered pathways and the number of genomes
561 according to the scale in the upper right corner. Pathways/genes in black were not identified.
562
563 **Figure S6.** Metabolic reconstruction of SAR202 based on three genomes (Bins 43-1, 43-2, and
564 43). Colors indicate recovered pathways and the number of genomes according to the scale in the
565 upper right corner. Pathways/genes in black were not identified.
566
567 **Figure S7.** Metabolic reconstruction of the putative ACD39 organism based on the Bin 13
568 genome. Red indicates recovered pathways. Pathways/genes in black were not identified.
569
570 **Figure S8.** ESOM of the first assembly using all assembled contigs. Bins are indicated by
571 different coloration. Squares within bins indicate 5000 bp segments of assembled contigs.
572
573 **Figure S9.** ESOM of the second assembly using 5% of assembled contigs randomly sub-
574 selected. Bins are indicated by different coloration. Squares within bins indicate 5000 bp
575 segments of assembled contigs.
576
577 **Figure S10.** Phylogenetic tree of the SAR202 clade using the 16S rRNA gene. Genes recovered
578 from assembled bins are indicated as 43-*. Values at nodes indicate Shimodaira-Hasegawa
579 "like" values (Price, 2010). Scale bar indicates changes per position.
580
581 **Figure S11.** Metabolic reconstruction of the *Gemmatimonas* organism based on the Bin 42
582 genome. Red indicates recovered pathways. Pathways/genes in black were not identified.
583
584 **Figure S12.** Metabolic reconstruction of *Nitrosopumilus* based on two genomes (Bins 71 and
585 72). Colors indicate recovered pathways and the number of genomes according to the scale in the
586 upper right corner. Pathways/genes in black were not identified.
587
588 **Figure S13.** Metabolic reconstruction of *Nitrospina* based on the four most complete genomes
589 (Bins 24, 25, 47-2, and 47). Colors indicate recovered pathways and the number of genomes
590 according to the scale in the upper right corner. Pathways/genes in black were not identified.

591
592 **Figure S14.** RNA vs. DNA coverage rank, according to BWA mapping of reads, for each bin by
593 sample.
594
595 **Figure S15.** Phylogenetic tree of *nxrA* genes and near neighbors from Lüke et al. (2016).
596 Sequences from binned genomes are indicated with *Bin*. Values at nodes indicate Shimodaira-
597 Hasegawa "like" values (Price, 2010). Scale bar indicates changes per position.
598
599 Additional Supplemental Information such as scripts, workflows, and key files, are provided as a
600 link hosted at the Thrash Lab website: http://thethrashlab.com/publications.
601
602

References Cited

1 Rabalais, N. N. *et al.* Nutrient-enhanced productivity in the northern Gulf of Mexico: past, present and future. *Hydrobiologia* **475**, 39-63, (2002).

2 Diaz, R. J. & Rosenberg, R. Spreading Dead Zones and Consequences for Marine Ecosystems. *Science* **321**, 926-929, (2008).

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. 1-64 (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, D. E., DeLong, E. F., Letelier, R. M. & Stewart, F. J. Microbial oceanography of anoxic oxygen minimum zones. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15996-16003, (2012).

5 Wright, J. J., Konwar, K. M. & Hallam, S. J. Microbial ecology of expanding oxygen minimum zones. *Nat Rev Micro* **10**, 381-394, (2012).

6 Grote, J., Jost, G., Labrenz, M., Herndl, G. J. & Jurgens, K. Epsilonproteobacteria Represent the Major Portion of Chemoautotrophic Bacteria in Sulfidic Waters of Pelagic Redoxclines of the Baltic and Black Seas. *Appl. Environ. Microbiol.* **74**, 7546-7551, (2008).

7 Glaubitz, S., Kiesslich, K., Meeske, C., Labrenz, M. & Jurgens, K. SUP05 Dominates the Gammaproteobacterial Sulfur Oxidizer Assemblages in Pelagic Redoxclines of the Central Baltic and Black Seas. *Appl. Environ. Microbiol.* **79**, 2767-2776, (2013).

8 Friedrich, J. *et al.* Investigating hypoxia in aquatic environments: diverse approaches to addressing a complex phenomenon. *Biogeosciences* **11**, 1215-1259, (2014).

9 Lam, P. & Kuypers, M. Microbial Nitrogen Cycling Processes in Oxygen Minimum Zones. *Annu. Rev. Marine. Sci.* **3**, 317-345, (2011).

10 Beman, J. M., Popp, B. N. & Alford, S. E. 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, (2012).

11 Saad, E. M. *et al.* Understanding marine dissolved organic matter production: Compositional insights from axenic cultures of Thalassiosira pseudonana. *Limnology and Oceanography* **AOP. doi: 10.1002/lno.10367**, (2016).

12 Canfield, D. E. *et al.* A Cryptic Sulfur Cycle in Oxygen-Minimum-Zone Waters off the Chilean Coast. *Science* **330**, 1375-1378, (2010).

13 Stewart, F. J., Ulloa, O. & DeLong, E. F. Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environmental Microbiology* **14**, 23-40, (2012).

14 Roberts, B. J. & Doty, S. M. Spatial and Temporal Patterns of Benthic Respiration and Net Nutrient Fluxes in the Atchafalaya River Delta Estuary. *Estuaries and Coasts* **38**, 1918-1936, (2015).

15 McCarthy, M. J., Carini, S. A., Liu, Z., Ostrom, N. E. & Gardner, W. S. Oxygen consumption in the water column and sediments of the northern Gulf of Mexico hypoxic zone. *Estuarine, Coastal and Shelf Science* **123**, 46-53, (2013).

16 Schaeffer, B. A. *et al.* 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, (2011).

17 Rabalais, N. N. *et al.* 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, (2007).
18 Rabalais, N. N., Turner, R. E. & Wiseman Jr, W. J. Hypoxia in the Gulf of Mexico. *J. Environ. Qual.* **30**, 320-329, (2001).
19 Rabalais, N. N., Turner, R. E. & Wiseman, J. W. J. Gulf of Mexico hypoxia, AKA "The dead zone". *Annual Review of Ecology and Systematics* **33**, 235-263, (2002).
20 Wiseman Jr, W. J., Rabalais, N. N., Turner, R. E., Dinnel, S. P. & MacNaughton, A. Seasonal and interannual variability within the Louisiana coastal current: stratification and hypoxia. *Journal of Marine Systems* **12**, 237-248, (1997).
21 Gillies, L. E., Thrash, J. C., deRada, S., Rabalais, N. N. & Mason, O. U. Archaeal enrichment in the hypoxic zone in the northern Gulf of Mexico. *Environmental Microbiology* **17**, 3847, (2015).
22 Bristow, L. A. *et al.* Biogeochemical and metagenomic analysis of nitrite accumulation in the Gulf of Mexico hypoxic zone. *Limnology and Oceanography*, doi: 10.1002/lno.10130, (2015).
23 Lücker, S., Nowka, B., Rattei, T., Spieck, E. & Daims, H. The Genome of Nitrospina gracilis Illuminates the Metabolism and Evolution of the Major Marine Nitrite Oxidizer. *Front. Microbiol.* **4**, 27, (2013).
24 Spieck, E., Keuter, S., Wenzel, T., Bock, E. & Ludwig, W. Characterization of a new marine nitrite oxidizing bacterium, Nitrospina watsonii sp. nov., a member of the newly proposed phylum “Nitrospinae”. *Syst. Appl. Microbiol.* **37**, 170-176, (2014).
25 Beaumont, H. J. E., Lens, S. I., Reijnders, W. N. M., Westerhoff, H. V. & Spanning, R. J. M. Expression of nitrite reductase in Nitrosomonas europaea involves NsrR, a novel nitrite‐sensitive transcription repressor. *Molecular Microbiology* **54**, 148-158, (2004).
26 Cantera, J. J. L. & Stein, L. Y. Role of nitrite reductase in the ammonia-oxidizing pathway of Nitrosomonas europaea. *Archives of Microbiology* **188**, 349-354, (2007).
27 Walker, C. B. *et al.* Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8818-8823, (2010).
28 Santoro, A. E., Buchwald, C., McIlvin, M. R. & Casciotti, K. L. Isotopic signature of N2O produced by marine ammonia-oxidizing archaea. *Science* **333**, 1282-1285, (2011).
29 Löscher, C. R. *et al.* Production of oceanic nitrous oxide by ammonia-oxidizing archaea. *Biogeosciences* **9**, 2419-2429, (2012).
30 Kamennaya, N. A. & Post, A. F. Characterization of cyanate metabolism in marine Synechococcus and Prochlorococcus spp. *Appl. Environ. Microbiol.* **77**, 291-301, (2011).
31 Vollmers, J. *et al.* Poles apart: Arctic and Antarctic Octadecabacter strains share high genome plasticity and a new type of xanthorhodopsin. *PLOS ONE* **8**, (2013).
32 Müller, A., Kjeldsen, K., Rattei, T., Pester, M. & Loy, A. Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases. *ISME J.* **9**, 1152-1165, (2015).
33 Beman, J. M. & Carolan, M. T. Deoxygenation alters bacterial diversity and community composition in the ocean's largest oxygen minimum zone. *Nature Communications* **4**, 2705-2705, (2013).
34 Burns, J. L. & DiChristina, T. J. Anaerobic respiration of elemental sulfur and thiosulfate by Shewanella oneidensis MR-1 requires psrA, a homolog of the phsA gene of

Salmonella enterica serovar typhimurium LT2. *Appl. Environ. Microbiol.* **75**, 5209-5217, (2009).

35 Gordon, D. A. & Giovannoni, S. J. Detection of stratified microbial populations related to Chlorobium and Fibrobacter species in the Atlantic and Pacific oceans. *Appl. Environ. Microbiol.* **62**, (1996).

36 Schattenhofer, M. *et al.* Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environmental Microbiology* **11**, 2078-2093, (2009).

37 Wright, J. J. *et al.* Genomic properties of Marine Group A bacteria indicate a role in the marine sulfur cycle. *ISME J.* **8**, 455-468, (2014).

38 Pitcher, R. S. & Watmough, N. J. The bacterial cytochrome cbb3 oxidases. *Biochimica et biophysica acta* **1655**, 388-399, (2004).

39 Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431-437, (2013).

40 Cho, J.-C., Vergin, K. L., Morris, R. M. & Giovannoni, S. J. 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, (2004).

41 Varela, M. M., Aken, H. M. & Herndl, G. J. Abundance and activity of Chloroflexi‐type SAR202 bacterioplankton in the meso‐ and bathypelagic waters of the (sub)tropical Atlantic. *Environmental Microbiology* **10**, (2008).

42 Treusch, A. H. *et al.* Seasonality and vertical structure of microbial communities in an ocean gyre. *ISME J.* **3**, 1148-1163, (2009).

43 Li, M. *et al.* Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. *Nature Communications* **6**, 8933. DOI: 8910.1038/ncomms9933, (2015).

44 Iverson, V. *et al.* Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* **335**, 587-590, (2012).

45 Wrighton, K. C. *et al.* Fermentation, Hydrogen, and Sulfur Metabolism in Multiple Uncultivated Bacterial Phyla. *Science* **337**, 1661-1665, (2012).

46 Wrighton, K. C. *et al.* Metabolic interdependencies between phylogenetically novel fermenters and respiratory organisms in an unconfined aquifer. *ISME J.* **8**, 1452-1463, (2014).

47 Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature Communications* **7**, 13219, (2016).

48 Luef, B. *et al.* Diverse uncultivated ultra-small bacterial cells in groundwater. *Nature Communications* **6**, 6372, (2015).

49 Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature*, 1-18, (2015).

50 Youssef, N. H. *et al.* In Silico Analysis of the Metabolic Potential and Niche Specialization of Candidate Phylum "Latescibacteria" (WS3). *PLOS ONE* **10**, 10.1371/journal.pone.0127499, (2015).

51 Dalsgaard, T. *et al.* Oxygen at Nanomolar Levels Reversibly Suppresses Process Rates and Gene Expression in Anammox and Denitrification in the Oxygen Minimum Zone off Northern Chile. *mBio* **5**, e01966-01914-e01966-01914, (2014).

739 52 Alewell, C., Paul, S., Lischeid, G. & Storck, F. R. Co-regulation of redox processes in
740 freshwater wetlands as a function of organic matter availability? *Science of The Total*
741 *Environment* **404**, 335-342, (2008).
742 53 Moran, M. *et al.* Deciphering ocean carbon in a changing world. *Proc. Natl. Acad. Sci.*
743 *U.S.A.* **113**, 3143-3151, (2016).
744 54 Elizabeth, B. K. The Impact of Microbial Metabolism on Marine Dissolved Organic
745 Matter. *Annu. Rev. Marine. Sci.* **3**, 567-599, (2011).
746 55 Buchan, A., LeCleir, G. R., Gulvik, C. A. & González, J. M. Master recyclers: features
747 and functions of bacteria associated with phytoplankton blooms. *Nat Rev Micro* **12**, 686-
748 698, (2014).
749 56 Osterholz, H. *et al.* Deciphering associations between dissolved organic molecules and
750 bacterial communities in a pelagic marine system. *ISME J.* **AOP.**
751 **doi:10.1038/ismej.2015.231**, (2016).
752 57 del Giorgio, P. A. & Cole, J. J. Bacterial growth efficiency in natural aquatic systems.
753 *Annual Review of Ecology and Systematics* **29**, 503-541, (1998).
754 58 Romero-Kutzner, V. *et al.* Respiration quotient variability: bacterial evidence. *Mar. Ecol.*
755 *Prog. Ser.* **519**, 47-59, (2015).
756 59 Henson, M. W. *et al.* Artificial Seawater Media Facilitate Cultivating Members of the
757 Microbial Majority from the Gulf of Mexico. *mSphere* **1**, e00028-00016.
758 doi:00010.01128/mSphere.00028-00016., (2016).
759 60 King, G. M., Smith, C. B., Tolar, B. & Hollibaugh, J. T. Analysis of composition and
760 structure of coastal to mesopelagic bacterioplankton communities in the northern gulf of
761 Mexico. *Front. Microbiol.* **3**, doi: 10.3389/fmicb.2012.00438, (2013).
762 61 Tolar, B. B., King, G. M. & Hollibaugh, J. T. An analysis of thaumarchaeota populations
763 from the northern gulf of Mexico. *Front. Microbiol.* **4**, 72, (2013).
764 62 Mason, O. U. *et al.* Metagenome, metatranscriptome and single-cell sequencing reveal
765 microbial response to Deepwater Horizon oil spill. *ISME J.*, 1-13, (2012).
766 63 Dick, G. J. *et al.* Community-wide analysis of microbial genome sequence signatures.
767 *Genome Biol* **10**, R85, (2009).
768 64 Sharon, I. *et al.* Time series community genomics analysis reveals rapid shifts in bacterial
769 species, strains, and phage during infant gut colonization. *Genome Res* **23**, 111-120,
770 (2013).
771 65 Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a de novo assembler
772 for single-cell and metagenomic sequencing data with highly uneven depth.
773 *Bioinformatics* **28**, 1420-1428, (2012).
774 66 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler
775 transform. *Bioinformatics* **25**, 1754-1760, (2009).
776 67 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM:
777 assessing the quality of microbial genomes recovered from isolates, single cells, and
778 metagenomes. *Genome Research*, 1-14, (2015).
779 68 Sorek, R. *et al.* Genome-wide experimental determination of barriers to horizontal gene
780 transfer. *Science* **318**, 1449-1452, (2007).
781 69 Darling, A. E. *et al.* PhyloSift: phylogenetic analysis of genomes and metagenomes.
782 *PeerJ* **2**, e243, (2014).
783 70 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high
784 throughput. *Nucleic Acids Research* **32**, 1792-1797, (2004).

785 71 Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): a
786 new software for selection of phylogenetic informative regions from multiple sequence
787 alignments. *BMC evolutionary biology* **10**, 210, (2010).
788 72 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
789 large phylogenies. *Bioinformatics*, btu033, (2014).
790 73 Hug, L. A. *et al.* A new view of the tree of life. *Nature Microbiology* **1**, 16048, (2016).
791 74 Rappé, M. S., Gordon, D. A., Vergin, K. L. & Giovannoni, S. J. Phylogeny of
792 Actinobacteria Small Subunit (SSU) rRNA Gene Clones Recovered From Marine
793 Bacterioplankton. *Syst. Appl. Microbiol.* **22**, 106-112, (1999).
794 75 Ghai, R., Mizuno, C., Picazo, A., Camacho, A. & Rodriguez-Valera, F. Metagenomics
795 uncovers a new group of low GC and ultra-small marine Actinobacteria. *Sci. Rep.* **3**,
796 2471, (2013).
797 76 Mizuno, C., Rodriguez-Valera, F. & Ghai, R. Genomes of Planktonic Acidimicrobiales:
798 Widening Horizons for Marine Actinobacteria by Metagenomics. *mBio* **6**, 14, (2015).
799 77 Castresana, J. Selection of conserved blocks from multiple alignments for their use in
800 phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540-552, (2000).
801 78 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood
802 trees for large alignments. *PLOS ONE* **5**, e9490, (2010).
803 79 Markowitz, V. M. *et al.* IMG 4 version of the integrated microbial genomes comparative
804 analysis system. *Nucleic Acids Research* **42**, D560-567, (2014).
805 80 Saier, M. H. *et al.* The Transporter Classification Database (TCDB): recent advances.
806 *Nucleic Acids Research* **44**, (2016).
807 81 Lombard, V., Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-
808 active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **42**, (2014).
809 82 Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and
810 quantifying mammalian transcriptomes by RNA-Seq. *Nat Meth* **5**, 621-628, (2008).
811 83 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing
812 genomic features. *Bioinformatics* **26**, 841-842, (2010).
813

Figure 1

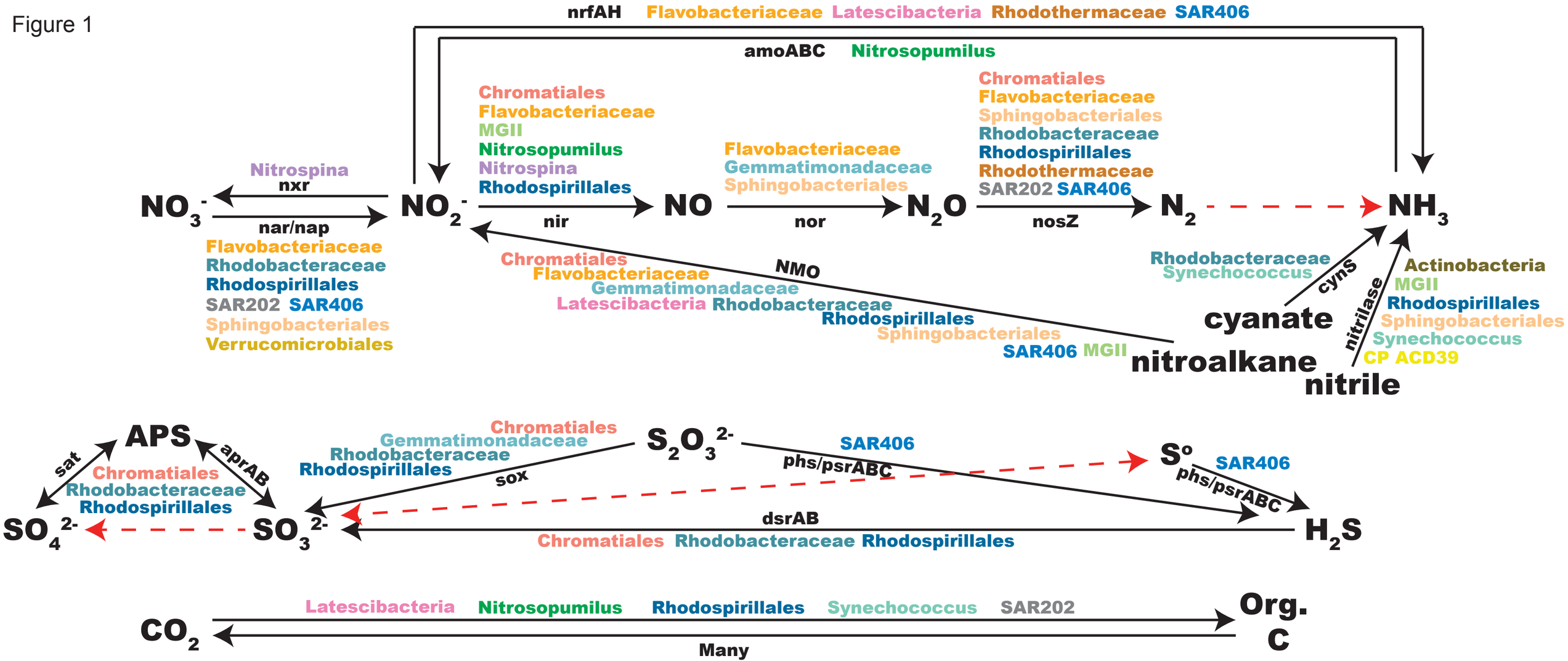

nrfAH
Flavobacteriaceae
Latescibacteria
Rhodothermaceae
SAR406
amoABC
Nitrosopumilus
Chromatiales
Flavobacteriaceae
MGII
Nitrosopumilus
Nitrospina
Rhodospirillales
Flavobacteriaceae
Gemmatimonadaceae
Sphingobacteriales
Chromatiales
Flavobacteriaceae
Sphingobacteriales
Rhodobacteraceae
Rhodospirillales
Rhodothermaceae
SAR202 SAR406
Nitrospina
nxr
NO3-
nar/nap
NO2-
nir
NO
nor
N2O
nosZ
N2
NH3
Flavobacteriaceae
Rhodobacteraceae
Rhodospirillales
SAR202 SAR406
Sphingobacteriales
Verrucomicrobiales
Chromatiales
Flavobacteriaceae
Gemmatimonadaceae
Latescibacteria Rhodobacteraceae
Rhodospirillales
Sphingobacteriales
SAR406 MGII
NMO
Rhodobacteraceae
Synechococcus
cyns
cyanate
nitrilase
nitroalkane
nitrile
Actinobacteria
MGII
Rhodospirillales
Sphingobacteriales
Synechococcus
CP ACD39
APS
sat
aprAB
Chromatiales
Rhodobacteraceae
Rhodospirillales
SO4 2-
SO3 2-
Chromatiales
Gemmatimonadaceae
Rhodobacteraceae
Rhodospirillales
sox
S2O3 2-
SAR406
phs/psrABC
S°
SAR406
phs/psrABC
H2S
dsrAB
Chromatiales Rhodobacteraceae Rhodospirillales
CO2
Latescibacteria Nitrosopumilus Rhodospirillales Synechococcus SAR202
Many
Org. C

Figure 2

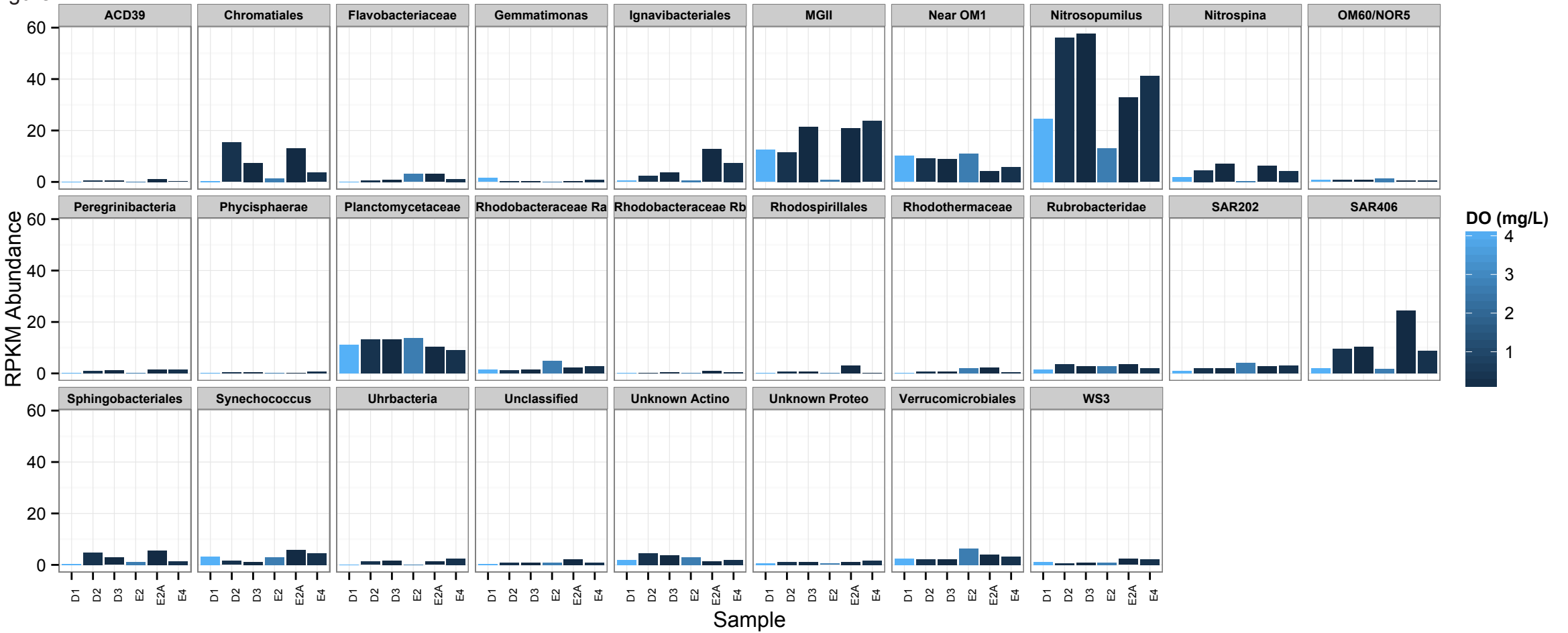

ACD39
Chromatiales
Flavobacteriaceae
Gemmatimonas
Ignavibacteriales
MGII
Near OM1
Nitrosopumilus
Nitrospina
OM60/NOR5
Peregrinibacteria
Phycisphaerae
Planctomycetaceae
Rhodobacteraceae Ra
Rhodobacteraceae Rb
Rhodospirillales
Rhodothermaceae
Rubrobacteridae
SAR202
SAR406
Sphingobacteriales
Synechococcus
Uhrbacteria
Unclassified
Unknown Actino
Unknown Proteo
Verrucomicrobiales
WS3
RPKM Abundance
0
20
40
60
D1
D2
D3
E2
E2A
E4
Sample
DO (mg/L)
1
2
3
4

Figure 3

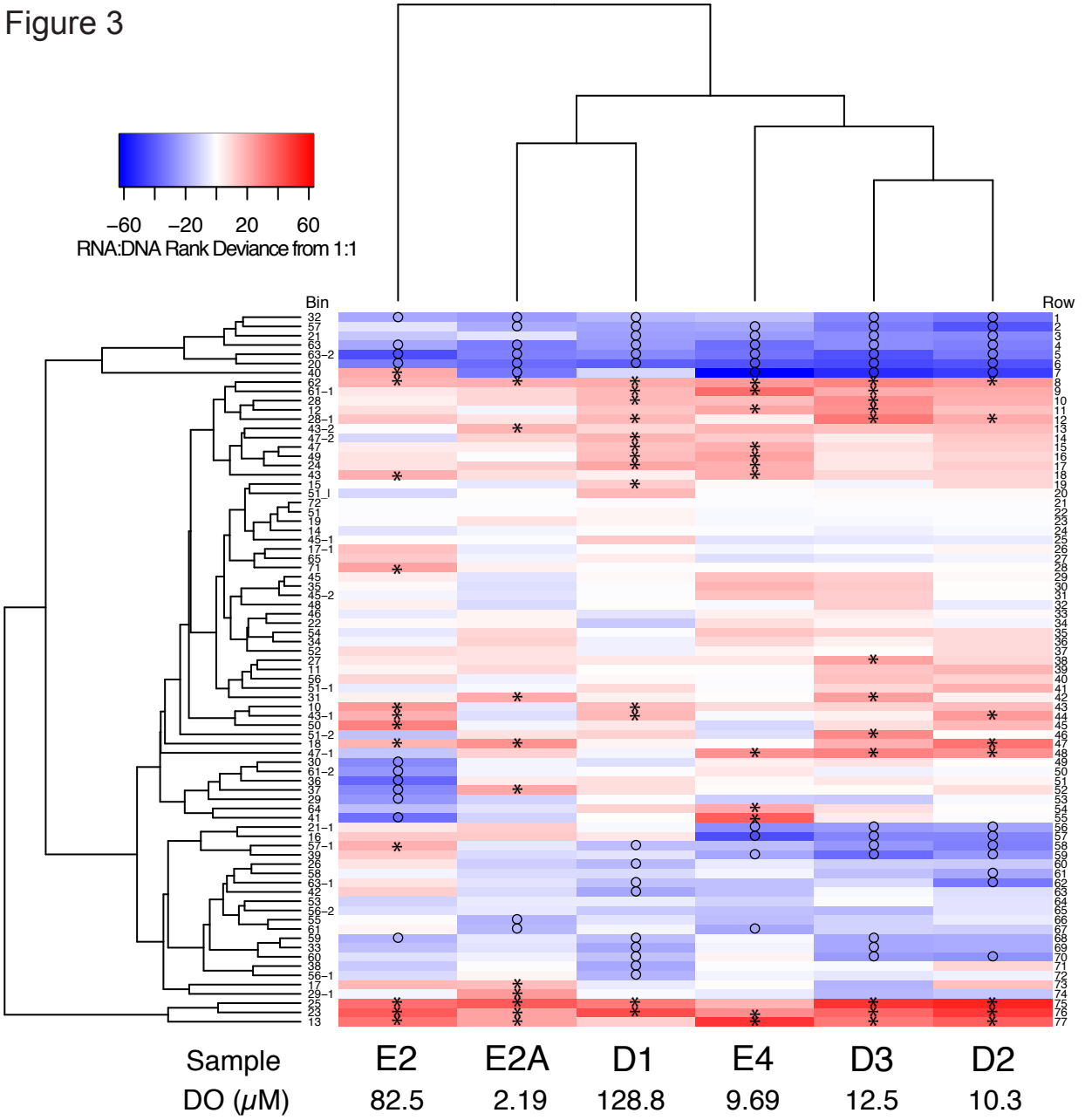
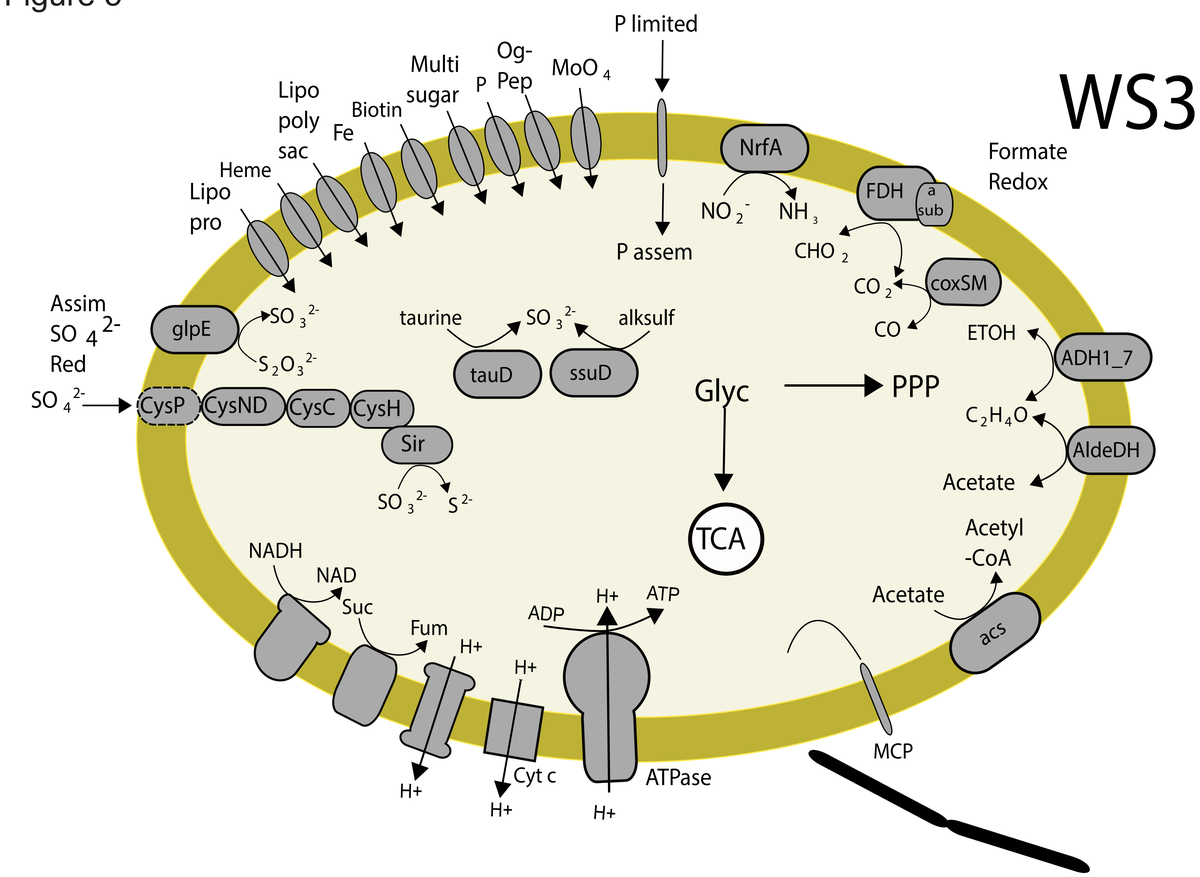
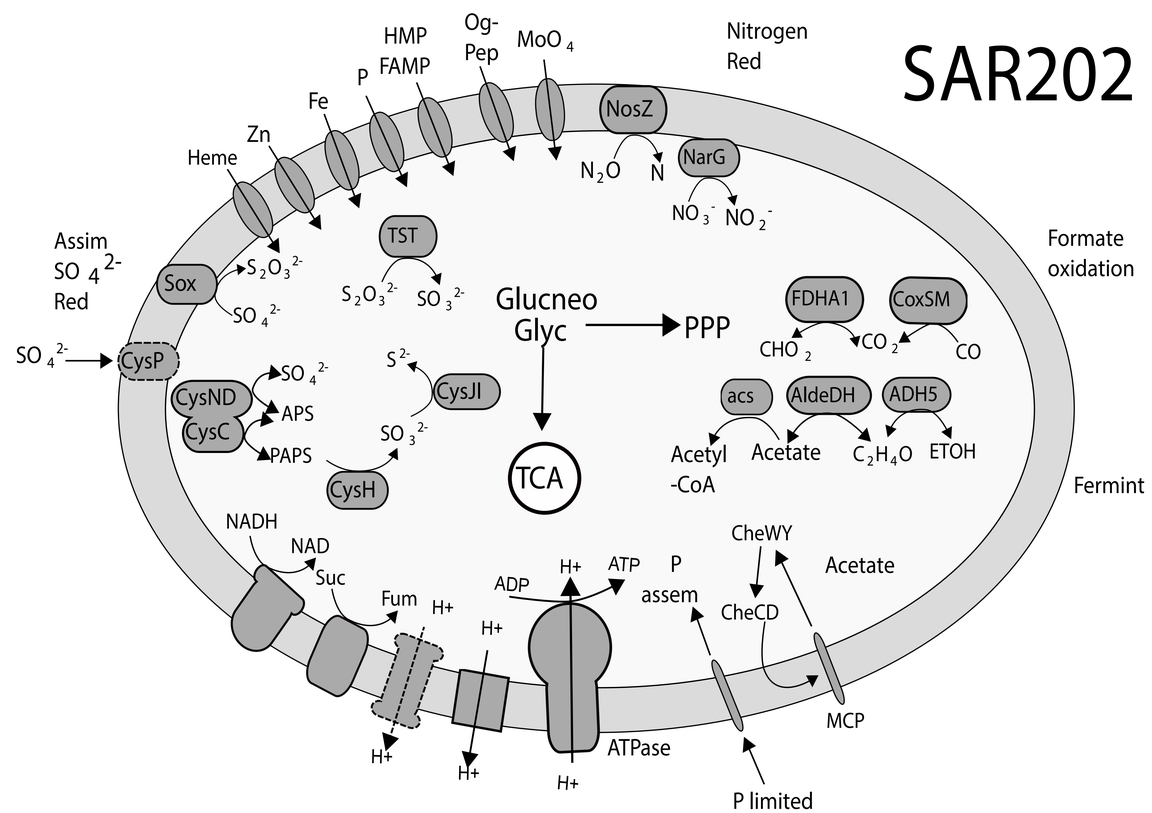

−60
−20
20
60
RNA:DNA Rank Deviance from 1:1
Bin
Row
Sample
E2
E2A
D1
E4
D3
D2
DO (μM)
82.5
2.19
128.8
9.69
12.5
10.3

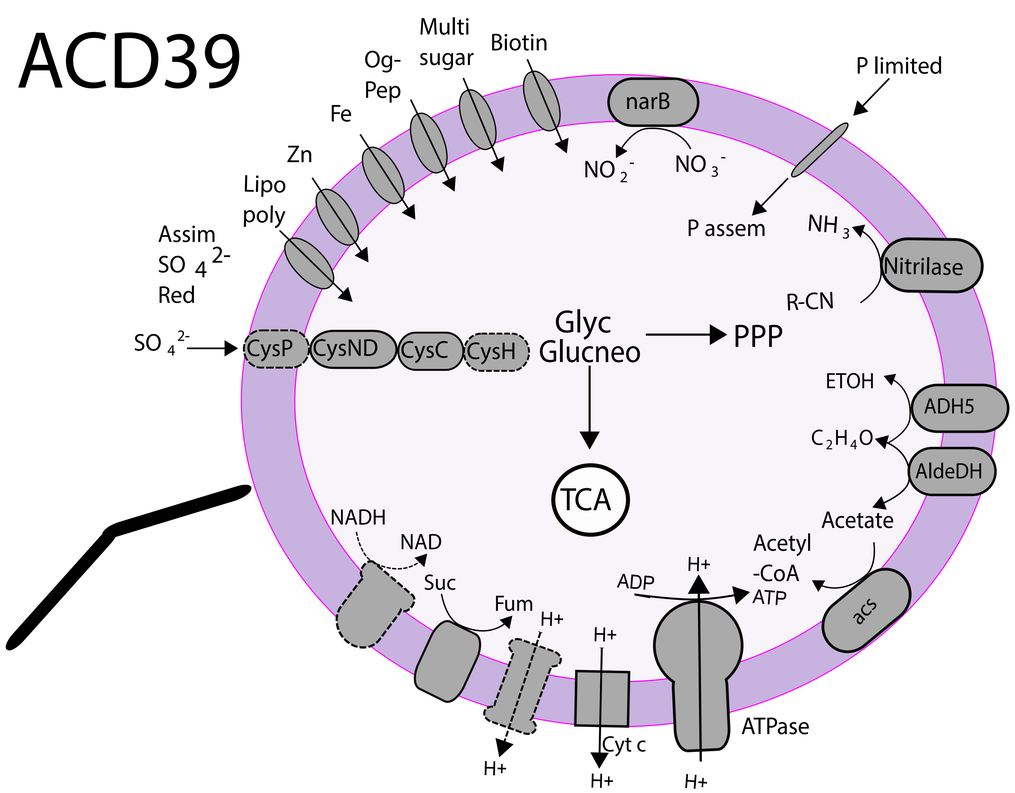
Figure 4

ccoN
ccoO
ccoP
ccoQ
coxA
coxB
coxC
coxD
cydA
cydB
cyoC
napA
napB
narG
narH
narI
narJ
nirB
nirD
nirK
nirS
norB
norC
nosZ
nrfA
nrfH
10000
100
1
RPKM
D1.RPKM
D2.RPKM
D3.RPKM
E2A.RPKM
E4.RPKM
sample
sample
D1.RPKM
D2.RPKM
D3.RPKM
E2A.RPKM
E4.RPKM

Figure 5


WS3
P limited
Lipo pro
Heme
Lipo poly sac
Fe
Biotin
Multi sugar
P
Og-Pep
MoO 4
NrfA
NO 2-
NH 3
FDH
a sub
Formate Redox
CHO 2
CO 2
coxSM
CO
P assem
Assim SO 4 2- Red
glpE
SO 3 2-
S 2O 3 2-
taurine
alksulf
tauD
ssuD
ETOH
ADH1_7
C 2H 4O
AldeDH
Acetate
SO 4 2-
CysP
CysND
CysC
CysH
Sir
S 2-
Glyc
PPP
TCA
Acetyl -CoA
Acetate
acs
NADH
NAD
Suc
Fum
H+
ADP
ATP
ATPase
Cyt c
MCP


SAR202
HMP FAMP
Og-Pep
MoO 4
Nitrogen Red
P
Fe
Zn
Heme
NosZ
N 2O
N
NarG
NO 3-
NO 2-
Assim SO 4 2- Red
Sox
S 2O 3 2-
SO 4 2-
TST
Formate oxidation
SO 4 2-
CysP
Glucneo
Glyc
PPP
FDHA1
CoxSM
CHO 2
CO 2
CO
CysND
CysC
APS
PAPS
S 2-
CysJI
SO 3 2-
CysH
acs
AldeDH
ADH5
Acetyl -CoA
Acetate
C 2H 4O
ETOH
TCA
Fermint
NADH
NAD
Suc
Fum
H+
ADP
ATP
P assem
CheWY
CheCD
Acetate
MCP
ATPase
P limited


ACD39
Multi sugar
Biotin
Og-Pep
Fe
Zn
Lipo poly
Assim SO 4 2- Red
narB
NO 2-
NO 3-
P limited
P assem
NH 3
Nitrilase
R-CN
SO 4 2-
CysP
CysND
CysC
CysH
Glyc
Glucneo
PPP
ETOH
ADH5
C 2H 4O
AldeDH
TCA
Acetate
Acetyl -CoA
acs
NADH
NAD
Suc
Fum
H+
ADP
ATP
ATPase
Cyt c

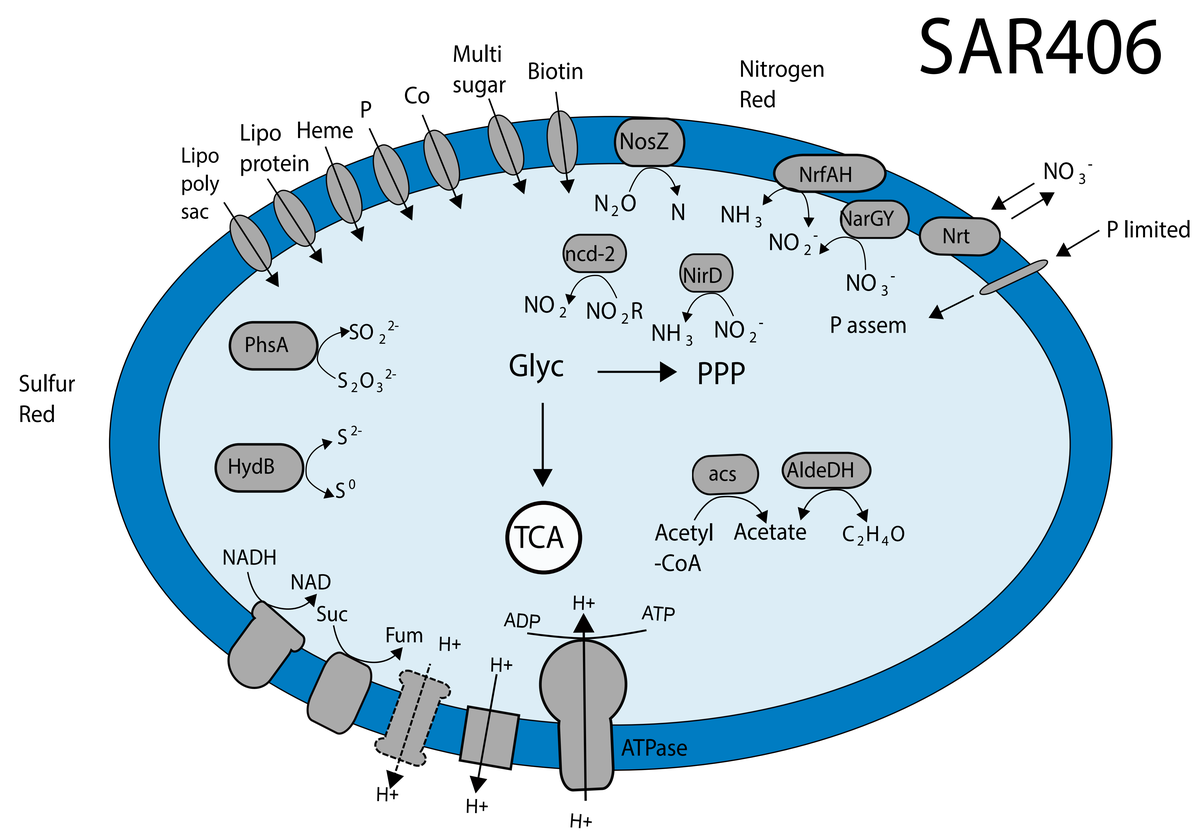

SAR406
Multi sugar
Biotin
Nitrogen Red
P
Co
Lipo protein
Heme
Lipo poly sac
NosZ
N 2O
N
NrfAH
NH 3
NO 2-
NarGY
NO 3-
Nrt
NO 3-
P limited
ncd-2
NO 2
NO 2R
NirD
NH 3
NO 2-
P assem
Sulfur Red
PhsA
SO 2 2-
S 2O 3 2-
Glyc
PPP
HydB
S 2-
S 0
acs
AldeDH
Acetyl -CoA
Acetate
C 2H 4O
TCA
NADH
NAD
Suc
Fum
H+
ADP
ATP
ATPase

Figure S3

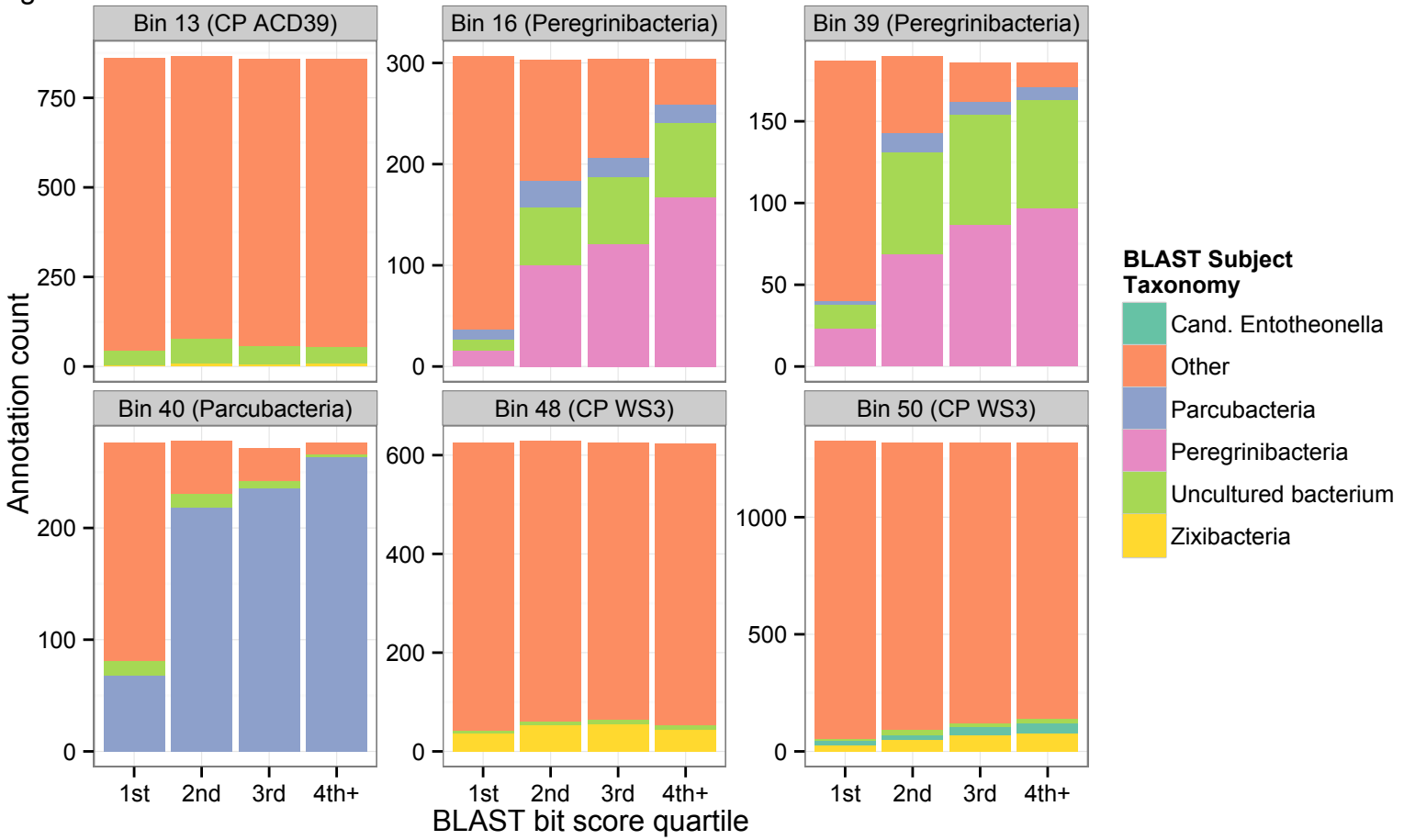

Bin 13 (CP ACD39)
Bin 16 (Peregrinibacteria)
Bin 39 (Peregrinibacteria)
Bin 40 (Parcubacteria)
Bin 48 (CP WS3)
Bin 50 (CP WS3)
Annotation count
BLAST bit score quartile
1st
2nd
3rd
4th+
BLAST Subject Taxonomy
Cand. Entotheonella
Other
Parcubacteria
Peregrinibacteria
Uncultured bacterium
Zixibacteria

Figure S4

"Reductive Bacterial"
"Oxidative Bacterial"
"Reductive Archaeal"
Methanopyrus kandleri
Methanocaldococcus jannaschii
Methanococcus maripaludis
Methanococcus vannielii
Methanococcus aeolicus
Methanococcus voltae
Methanococcus thermolithotrophicus
Methanococcus sp. 4484-2
Methanococcus sp. 4484-1
Methanococcus sp. 4484-3
Methanococcus sp. 4484-4
Methanococcus sp. 4484-5
Methanococcus sp. 4484-6
Methanococcus sp. 4484-7
Methanococcus sp. 4484-8
Methanococcus sp. 4484-9
Methanococcus sp. 4484-10
Methanococcus sp. 4484-11
Methanococcus sp. 4484-12
Methanococcus sp. 4484-13
Methanococcus sp. 4484-14
Methanococcus sp. 4484-15
Methanococcus sp.

Figure S5

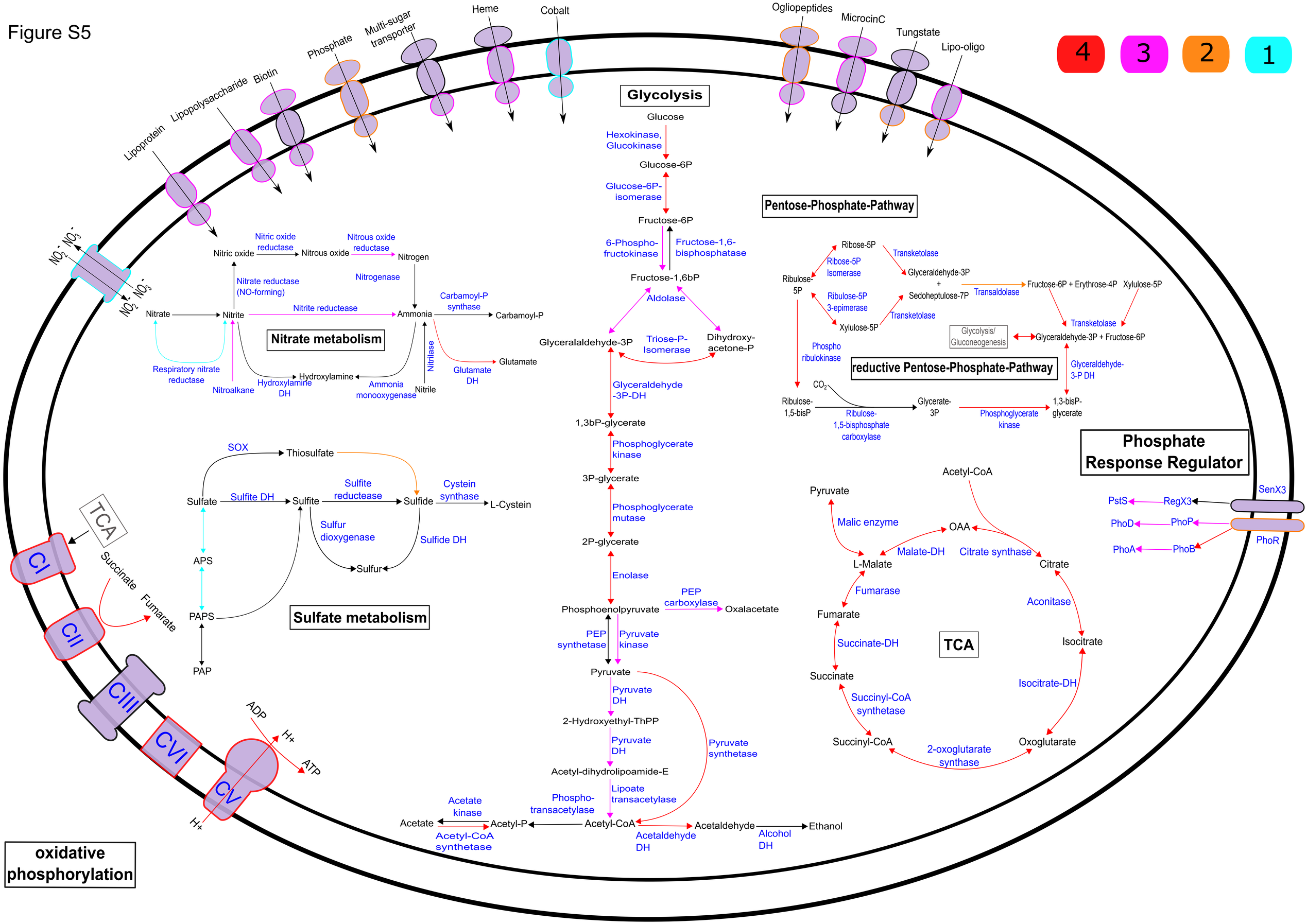

4
3
2
1
Lipoprotein
Lipopolysaccharide
Biotin
Phosphate
Multi-sugar transporter
Heme
Cobalt
Ogliopeptides
MicrocinC
Tungstate
Lipo-oligo
Glycolysis
Glucose
Hexokinase, Glucokinase
Glucose-6P
Glucose-6P-isomerase
Fructose-6P
6-Phospho-fructokinase
Fructose-1,6-bisphosphatase
Fructose-1,6bP
Aldolase
Glyceralaldehyde-3P
Triose-P-Isomerase
Dihydroxy-acetone-P
Glyceraldehyde -3P-DH
1,3bP-glycerate
Phosphoglycerate kinase
3P-glycerate
Phosphoglycerate mutase
2P-glycerate
Enolase
Phosphoenolpyruvate
PEP carboxylase
Oxalacetate
PEP synthetase
Pyruvate kinase
Pyruvate
Pyruvate DH
2-Hydroxyethyl-ThPP
Pyruvate DH
Acetyl-dihydrolipoamide-E
Lipoate transacetylase
Phospho-transacetylase
Acetyl-CoA
Pyruvate synthetase
Acetate kinase
Acetate
Acetyl-P
Acetyl-CoA synthetase
Acetaldehyde
Acetaldehyde DH
Ethanol
Alcohol DH
Pentose-Phosphate-Pathway
Ribose-5P
Ribose-5P Isomerase
Ribulose-5P
Ribulose-5P 3-epimerase
Xylulose-5P
Transketolase
Glyceraldehyde-3P + Sedoheptulose-7P
Transaldolase
Fructose-6P + Erythrose-4P
Xylulose-5P
Transketolase
Glycolysis/ Gluconeogenesis
Glyceraldehyde-3P + Fructose-6P
Glyceraldehyde-3-P DH
reductive Pentose-Phosphate-Pathway
Phospho ribulokinase
CO_2
Ribulose-1,5-bisP
Ribulose-1,5-bisphosphate carboxylase
Glycerate-3P
Phosphoglycerate kinase
1,3-bisP-glycerate
Phosphate Response Regulator
SenX3
RegX3
PhoR
PhoP
PhoB
PstS
PhoD
PhoA
Nitrate metabolism
NO_2^- NO_3^-
Nitric oxide
Nitric oxide reductase
Nitrous oxide
Nitrous oxide reductase
Nitrogen
Nitrogenase
Nitrate reductase (NO-forming)
Nitrate
Nitrite
Nitrite reductease
Ammonia
Carbamoyl-P synthase
Carbamoyl-P
Glutamate
Glutamate DH
Nitrilase
Nitrile
Ammonia monooxygenase
Hydroxylamine
Hydroxylamine DH
Nitroalkane
Respiratory nitrate reductase
Sulfate metabolism
SOX
Thiosulfate
Sulfate
Sulfite DH
Sulfite
Sulfite reductease
Sulfide
Cystein synthase
L-Cystein
Sulfur dioxygenase
Sulfur
Sulfide DH
APS
PAPS
PAP
TCA
Acetyl-CoA
Pyruvate
Malic enzyme
OAA
Malate-DH
Citrate synthase
Citrate
L-Malate
Fumarase
Fumarate
Aconitase
Isocitrate
Succinate-DH
Succinate
Isocitrate-DH
Succinyl-CoA synthetase
Succinyl-CoA
2-oxoglutarate synthase
Oxoglutarate
oxidative phosphorylation
CI
CII
CIII
CIV
CV
TCA
Succinate
Fumarate
ADP
ATP
H+
H+

Figure S6

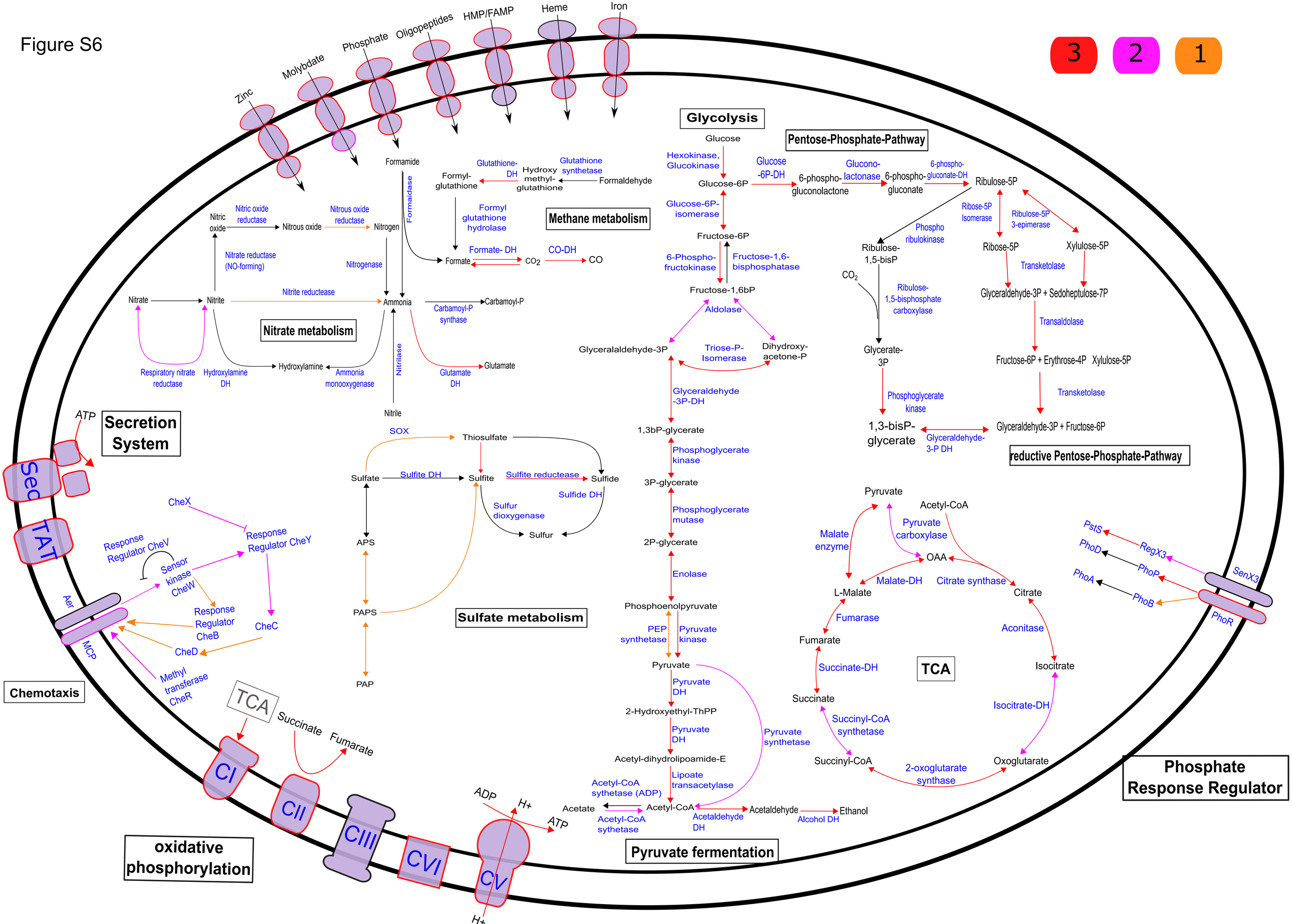

3
2
1
Zinc
Molybdate
Phosphate
Oligopeptides
HMP/FAMP
Heme
Iron
Glycolysis
Glucose
Hexokinase, Glucokinase
Glucose-6P
Glucose-6P-isomerase
Fructose-6P
6-Phospho-fructokinase
Fructose-1,6-bisphosphatase
Fructose-1,6bP
Aldolase
Glyceralaldehyde-3P
Triose-P-Isomerase
Dihydroxy-acetone-P
Glyceraldehyde-3P-DH
1,3bP-glycerate
Phosphoglycerate kinase
3P-glycerate
Phosphoglycerate mutase
2P-glycerate
Enolase
Phosphoenolpyruvate
PEP synthetase
Pyruvate kinase
Pyruvate
Pyruvate DH
2-Hydroxyethyl-ThPP
Pyruvate DH
Acetyl-dihydrolipoamide-E
Lipoate transacetylase
Acetyl-CoA
Pyruvate synthetase
Acetyl-CoA sythetase (ADP)
Acetate
Acetyl-CoA sythetase
Acetaldehyde
Acetaldehyde DH
Ethanol
Alcohol DH
Pyruvate fermentation
Pentose-Phosphate-Pathway
Glucose-6P
Glucose -6P-DH
6-phospho-gluconolactone
Glucono-lactonase
6-phospho-gluconate
6-phospho-gluconate-DH
Ribulose-5P
Ribose-5P Isomerase
Ribulose-5P 3-epimerase
Ribose-5P
Xylulose-5P
Transketolase
Glyceraldehyde-3P + Sedoheptulose-7P
Transaldolase
Fructose-6P + Erythrose-4P
Xylulose-5P
Transketolase
Glyceraldehyde-3P + Fructose-6P
Phospho ribulokinase
Ribulose-1,5-bisP
CO2
Ribulose-1,5-bisphosphate carboxylase
Glycerate-3P
Phosphoglycerate kinase
1,3-bisP-glycerate
Glyceraldehyde-3-P DH
reductive Pentose-Phosphate-Pathway
Methane metabolism
Formamide
Formaldehyde
Hydroxy methyl-glutathione
Glutathione synthetase
Glutathione-DH
Formyl-glutathione
Formyl glutathione hydrolase
Formaidase
Formate
Formate- DH
CO2
CO-DH
CO
Nitrate metabolism
Nitric oxide
Nitric oxide reductase
Nitrous oxide
Nitrous oxide reductase
Nitrogen
Nitrate reductase (NO-forming)
Nitrogenase
Nitrate
Nitrite
Nitrite reductease
Ammonia
Carbamoyl-P
Carbamoyl-P synthase
Glutamate
Glutamate DH
Respiratory nitrate reductase
Hydroxylamine DH
Hydroxylamine
Ammonia monooxygenase
Nitrilase
Nitrile
Sulfate metabolism
SOX
Thiosulfate
Sulfate
Sulfite DH
Sulfite
Sulfite reductease
Sulfide
Sulfide DH
Sulfur dioxygenase
Sulfur
APS
PAPS
PAP
TCA
Pyruvate
Acetyl-CoA
Pyruvate carboxylase
Malate enzyme
OAA
Malate-DH
Citrate synthase
L-Malate
Citrate
Fumarase
Aconitase
Fumarate
Isocitrate
Succinate-DH
Succinate
Isocitrate-DH
Succinyl-CoA synthetase
Succinyl-CoA
2-oxoglutarate synthase
Oxoglutarate
Succinate
Fumarate
Phosphate Response Regulator
PstS
PhoD
PhoA
RegX3
PhoP
PhoB
SenX3
PhoR
Secretion System
ATP
Sec
TAT
Chemotaxis
CheX
Response Regulator CheV
Response Regulator CheY
Sensor kinase CheW
Aer
MCP
Response Regulator CheB
CheC
CheD
Methyl transferase CheR
TCA
Succinate
Fumarate
oxidative phosphorylation
CI
CII
CIII
CVI
CV
ADP
H+
ATP
H+

Figure S7

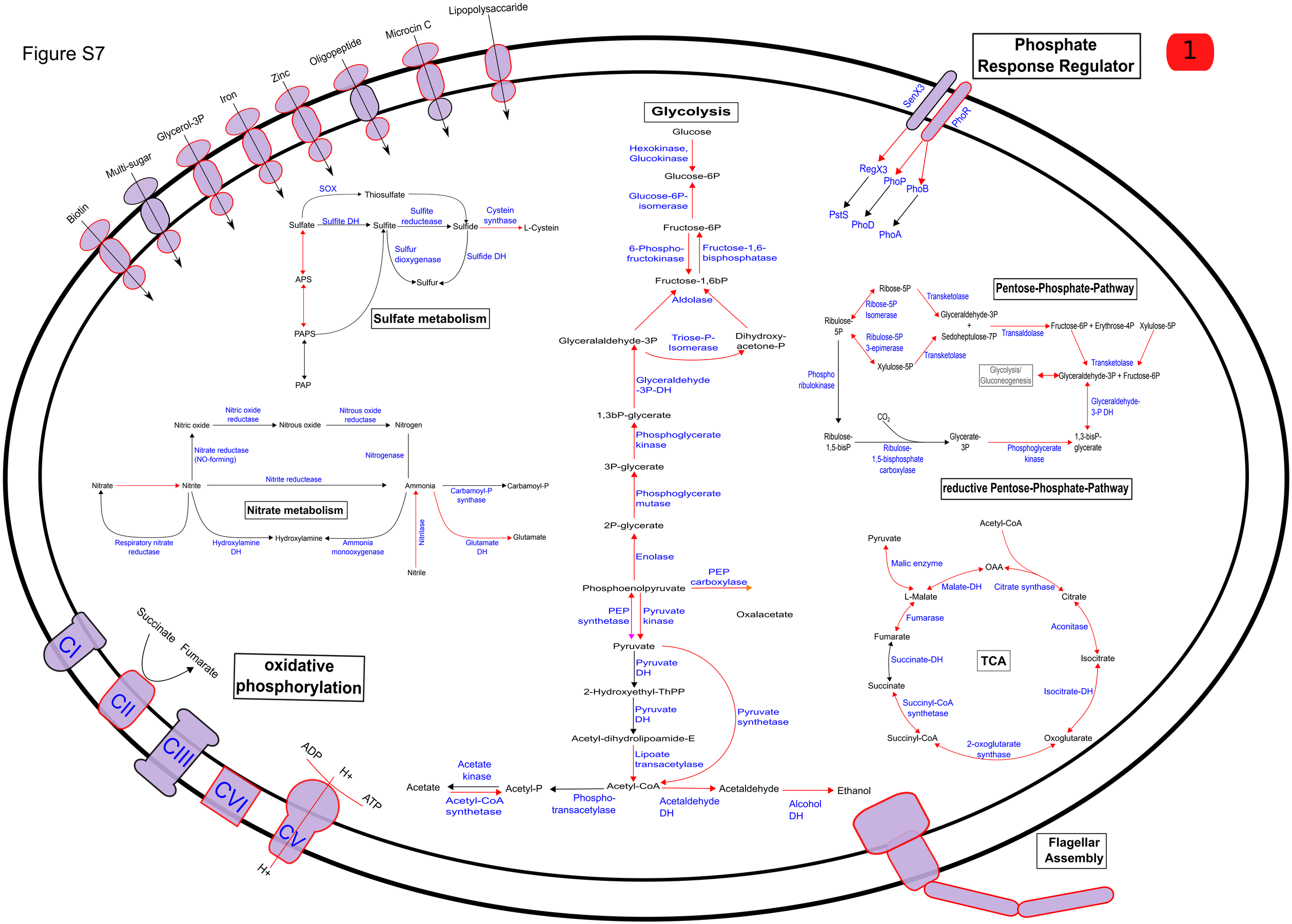
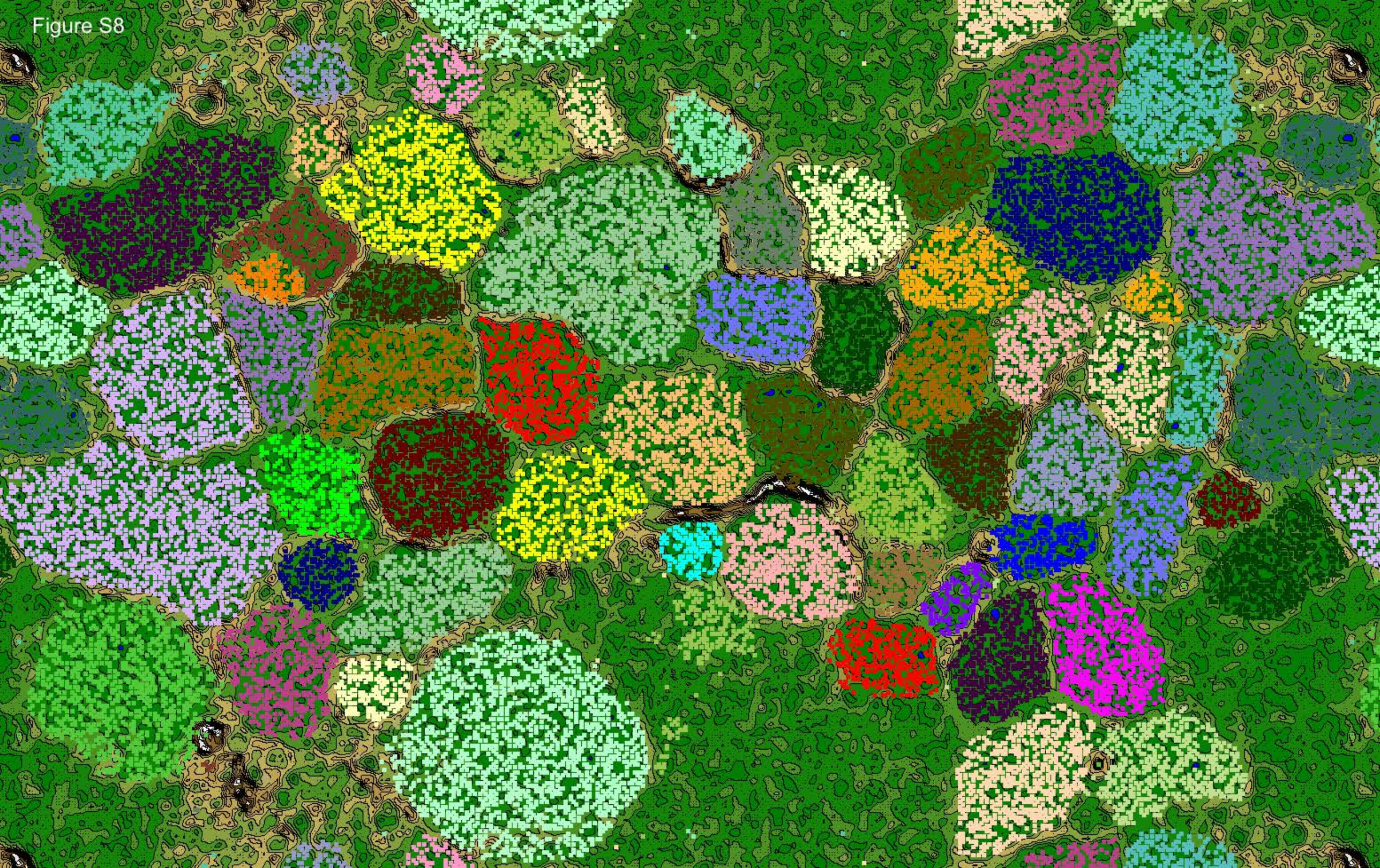

Lipopolysaccaride
Microcin C
Oligopeptide
Zinc
Iron
Glycerol-3P
Multi-sugar
Biotin
Phosphate Response Regulator
1
SenX3
PhoR
RegX3
PhoP
PhoB
PstS
PhoD
PhoA
Glycolysis
Glucose
Hexokinase, Glucokinase
Glucose-6P
Glucose-6P-isomerase
Fructose-6P
6-Phospho-fructokinase
Fructose-1,6-bisphosphatase
Fructose-1,6bP
Aldolase
Glyceralaldehyde-3P
Triose-P-Isomerase
Dihydroxy-acetone-P
Glyceraldehyde -3P-DH
1,3bP-glycerate
Phosphoglycerate kinase
3P-glycerate
Phosphoglycerate mutase
2P-glycerate
Enolase
Phosphoenolpyruvate
PEP carboxylase
Oxalacetate
PEP synthetase
Pyruvate kinase
Pyruvate
Pyruvate DH
2-Hydroxyethyl-ThPP
Pyruvate DH
Acetyl-dihydrolipoamide-E
Lipoate transacetylase
Pyruvate synthetase
Acetate
Acetate kinase
Acetyl-CoA synthetase
Acetyl-P
Phospho-transacetylase
Acetyl-CoA
Acetaldehyde DH
Acetaldehyde
Alcohol DH
Ethanol
Sulfate metabolism
SOX
Thiosulfate
Sulfate
Sulfite DH
Sulfite
Sulfite reductease
Sulfide
Cystein synthase
L-Cystein
Sulfur dioxygenase
Sulfide DH
Sulfur
APS
PAPS
PAP
Nitrate metabolism
Nitric oxide
Nitric oxide reductase
Nitrous oxide
Nitrous oxide reductase
Nitrogen
Nitrate reductase (NO-forming)
Nitrogenase
Nitrate
Nitrite
Nitrite reductease
Ammonia
Carbamoyl-P synthase
Carbamoyl-P
Respiratory nitrate reductase
Hydroxylamine DH
Hydroxylamine
Ammonia monooxygenase
Nitrilase
Glutamate DH
Glutamate
Nitrile
Pentose-Phosphate-Pathway
Ribose-5P
Ribose-5P Isomerase
Transketolase
Ribulose-5P
Glyceraldehyde-3P
+
Sedoheptulose-7P
Transaldolase
Fructose-6P + Erythrose-4P
Xylulose-5P
Ribulose-5P 3-epimerase
Xylulose-5P
Transketolase
Transketolase
Glycolysis/ Gluconeogenesis
Glyceraldehyde-3P + Fructose-6P
Phospho ribulokinase
Glyceraldehyde-3-P DH
CO_2
Ribulose-1,5-bisP
Ribulose-1,5-bisphosphate carboxylase
Glycerate-3P
Phosphoglycerate kinase
1,3-bisP-glycerate
reductive Pentose-Phosphate-Pathway
Acetyl-CoA
Pyruvate
Malic enzyme
OAA
L-Malate
Malate-DH
Citrate synthase
Citrate
Fumarase
Aconitase
Fumarate
Succinate-DH
TCA
Isocitrate
Succinate
Isocitrate-DH
Succinyl-CoA synthetase
Succinyl-CoA
2-oxoglutarate synthase
Oxoglutarate
oxidative phosphorylation
CI
CII
CIII
CVI
CV
Succinate
Fumarate
ADP
ATP
H+
H+
Flagellar Assembly

Figure S8



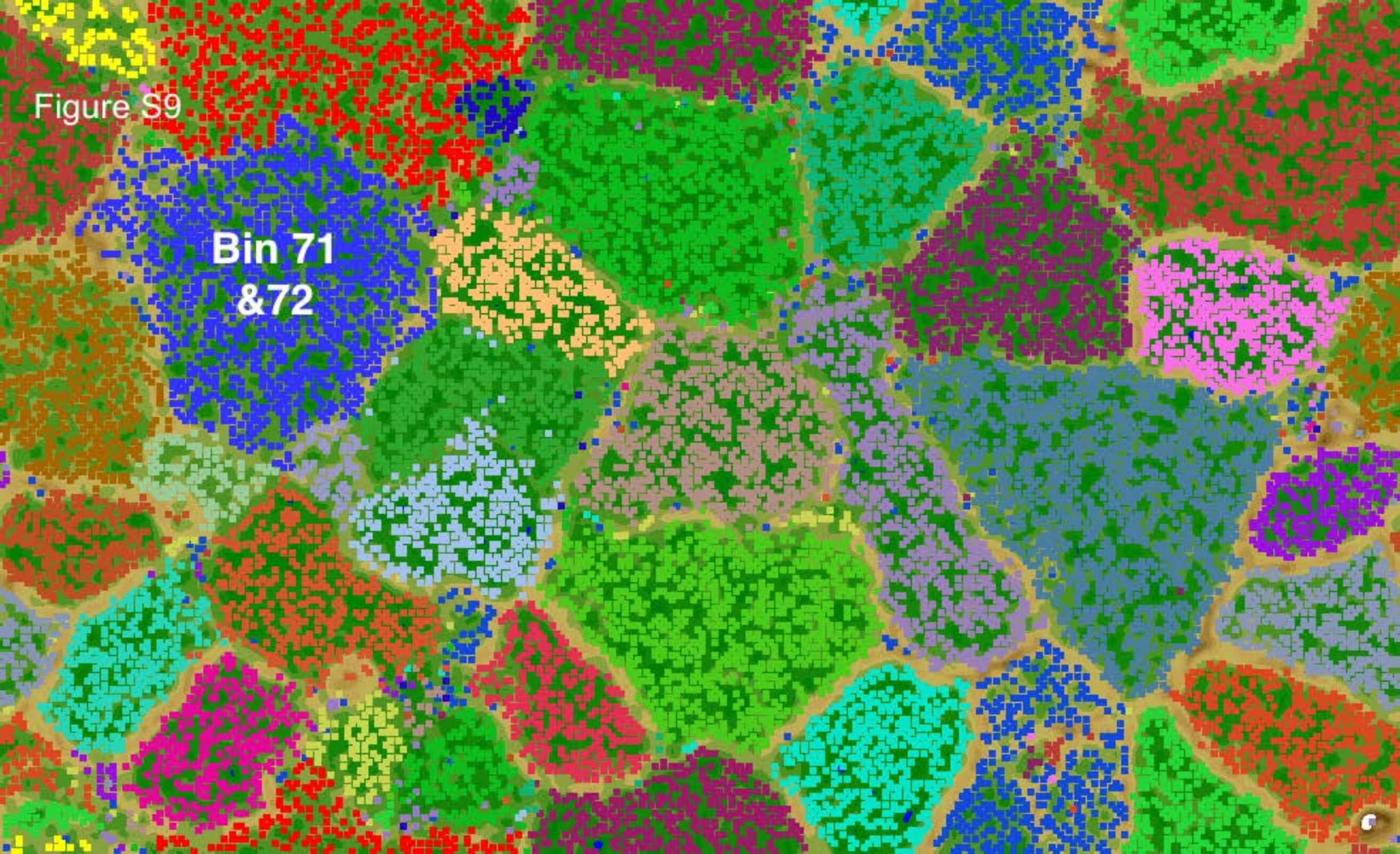


Figure S9
Bin 71
&72

Figure S10

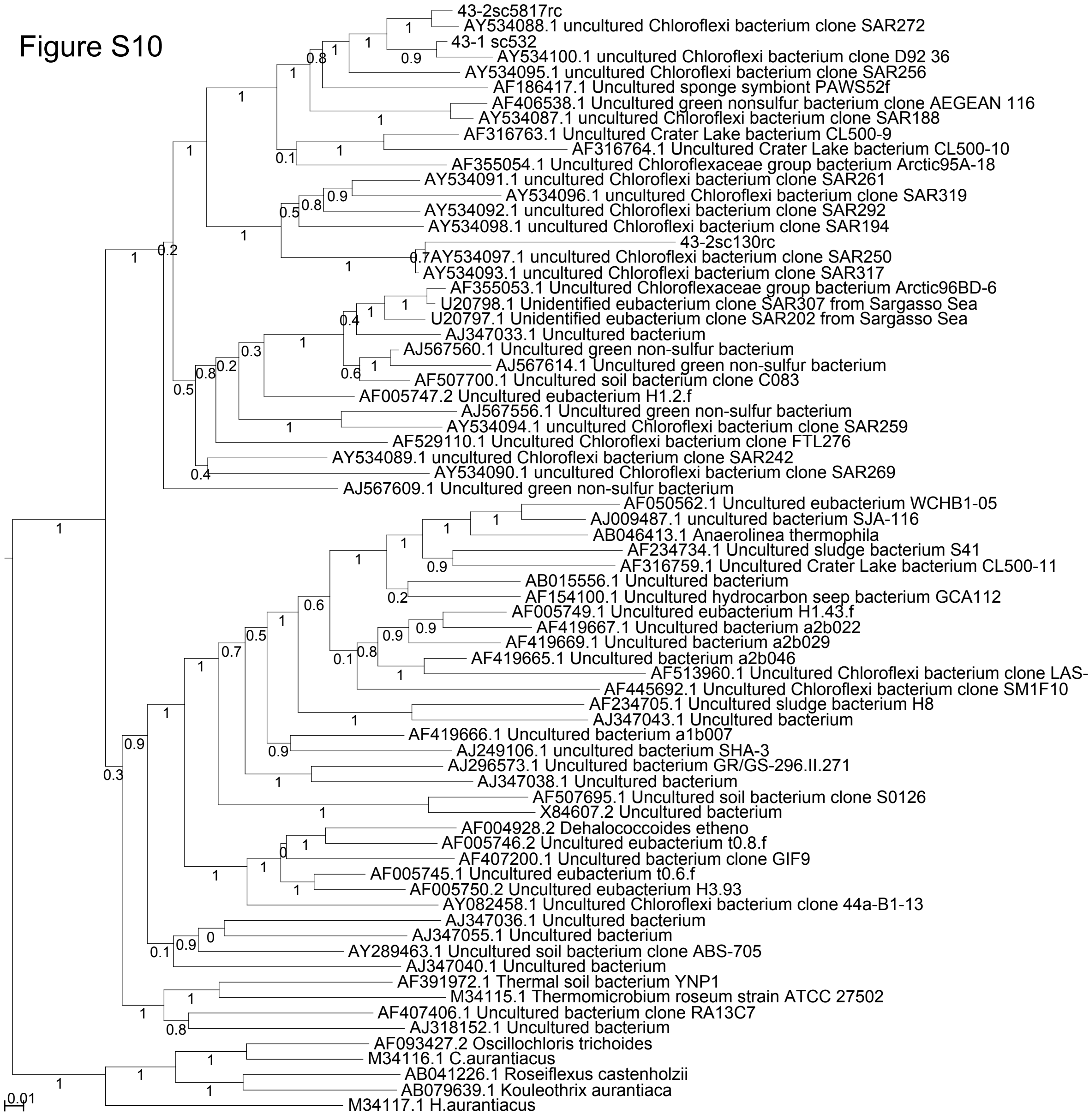

43-2sc5817rc
AY534088.1 uncultured Chloroflexi bacterium clone SAR272
43-1 sc532
AY534100.1 uncultured Chloroflexi bacterium clone D92 36
AY534095.1 uncultured Chloroflexi bacterium clone SAR256
AF186417.1 Uncultured sponge symbiont PAWS52f
AF406538.1 Uncultured green nonsulfur bacterium clone AEGEAN 116
AY534087.1 uncultured Chloroflexi bacterium clone SAR188
AF316763.1 Uncultured Crater Lake bacterium CL500-9
AF316764.1 Uncultured Crater Lake bacterium CL500-10
AF355054.1 Uncultured Chloroflexaceae group bacterium Arctic95A-18
AY534091.1 uncultured Chloroflexi bacterium clone SAR261
AY534096.1 uncultured Chloroflexi bacterium clone SAR319
AY534092.1 uncultured Chloroflexi bacterium clone SAR292
AY534098.1 uncultured Chloroflexi bacterium clone SAR194
43-2sc130rc
AY534097.1 uncultured Chloroflexi bacterium clone SAR250
AY534093.1 uncultured Chloroflexi bacterium clone SAR317
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
AY534089.1 uncultured Chloroflexi bacterium clone SAR242
AY534090.1 uncultured Chloroflexi bacterium clone SAR269
AJ567609.1 Uncultured green non-sulfur bacterium
AF050562.1 Uncultured eubacterium WCHB1-05
AJ009487.1 uncultured bacterium SJA-116
AB046413.1 Anaerolinea thermophila
AF234734.1 Uncultured sludge bacterium S41
AF316759.1 Uncultured Crater Lake bacterium CL500-11
AB015556.1 Uncultured bacterium
AF154100.1 Uncultured hydrocarbon seep bacterium GCA112
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-
AF445692.1 Uncultured Chloroflexi bacterium clone SM1F10
AF234705.1 Uncultured sludge bacterium H8
AJ347043.1 Uncultured bacterium
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
AF407200.1 Uncultured bacterium clone GIF9
AF005745.1 Uncultured eubacterium t0.6.f
AF005750.2 Uncultured eubacterium H3.93
AY082458.1 Uncultured Chloroflexi bacterium clone 44a-B1-13
AJ347036.1 Uncultured bacterium
AJ347055.1 Uncultured bacterium
AY289463.1 Uncultured soil bacterium clone ABS-705
AJ347040.1 Uncultured bacterium
AF391972.1 Thermal soil bacterium YNP1
M34115.1 Thermomicrobium roseum strain ATCC 27502
AF407406.1 Uncultured bacterium clone RA13C7
AJ318152.1 Uncultured bacterium
AF093427.2 Oscillochloris trichoides
M34116.1 C.aurantiacus
AB041226.1 Roseiflexus castenholzii
AB079639.1 Kouleothrix aurantiaca
M34117.1 H.aurantiacus
0.01

Figure S11

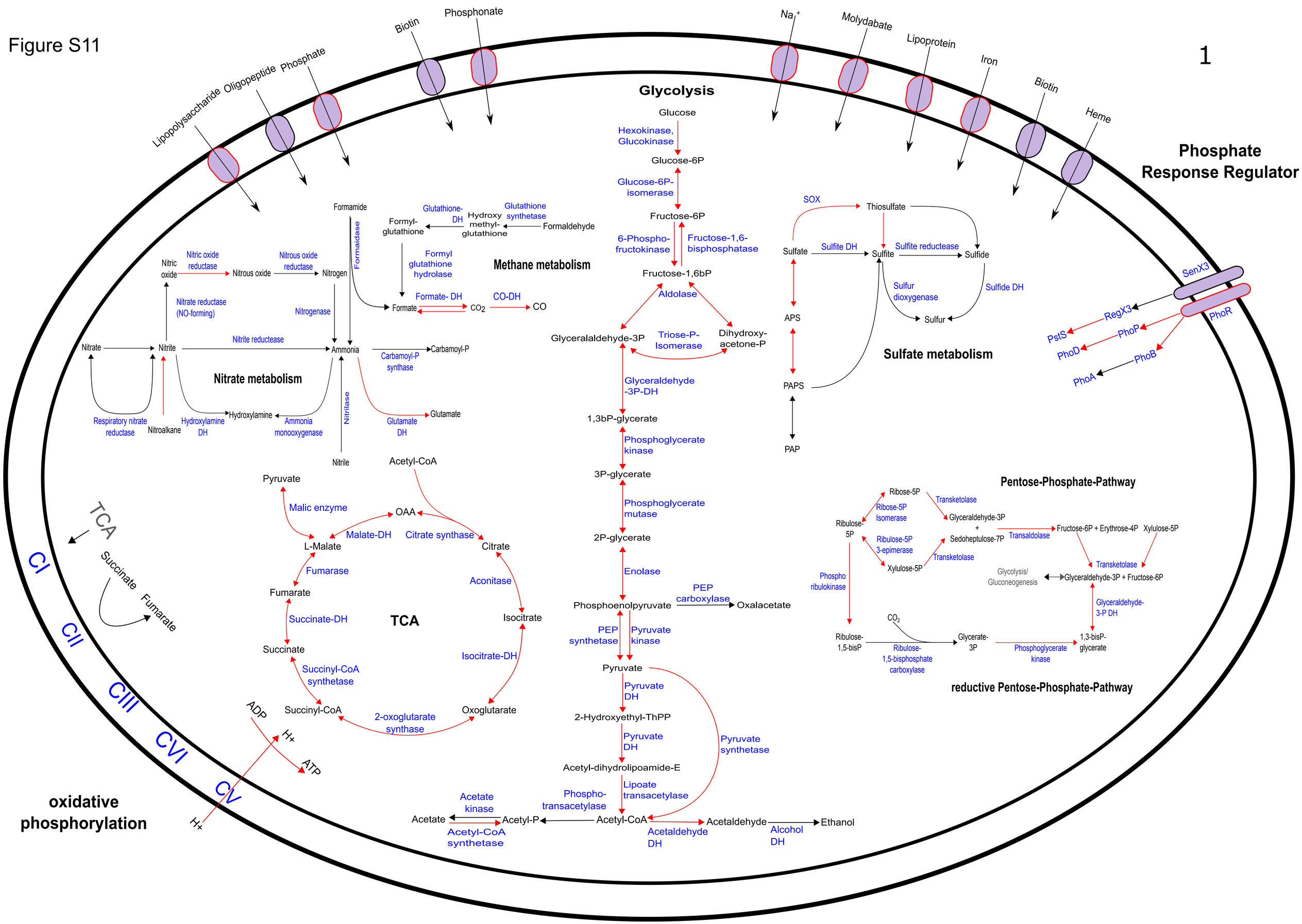

1
Lipopolysaccharide
Oligopeptide
Phosphate
Biotin
Phosphonate
Na+
Molydabate
Lipoprotein
Iron
Biotin
Heme
Phosphate Response Regulator
SenX3
RegX3
PhoR
PhoP
PhoB
PstS
PhoD
PhoA
Glycolysis
Glucose
Hexokinase, Glucokinase
Glucose-6P
Glucose-6P-isomerase
Fructose-6P
6-Phospho-fructokinase
Fructose-1,6-bisphosphatase
Fructose-1,6bP
Aldolase
Glyceralaldehyde-3P
Triose-P-Isomerase
Dihydroxy-acetone-P
Glyceraldehyde -3P-DH
1,3bP-glycerate
Phosphoglycerate kinase
3P-glycerate
Phosphoglycerate mutase
2P-glycerate
Enolase
Phosphoenolpyruvate
PEP carboxylase
Oxalacetate
PEP synthetase
Pyruvate kinase
Pyruvate
Pyruvate DH
2-Hydroxyethyl-ThPP
Pyruvate DH
Acetyl-dihydrolipoamide-E
Lipoate transacetylase
Phospho-transacetylase
Pyruvate synthetase
Acetyl-CoA
Acetaldehyde DH
Acetaldehyde
Alcohol DH
Ethanol
Acetyl-P
Acetyl-CoA synthetase
Acetate kinase
Acetate
Methane metabolism
Formamide
Formaidase
Formyl-glutathione
Glutathione-DH
Hydroxy methyl-glutathione
Glutathione synthetase
Formaldehyde
Formyl glutathione hydrolase
Formate
Formate- DH
CO2
CO-DH
CO
Nitric oxide
Nitric oxide reductase
Nitrous oxide
Nitrous oxide reductase
Nitrogen
Nitrogenase
Nitrate reductase (NO-forming)
Nitrate
Nitrite
Nitrite reductase
Ammonia
Carbamoyl-P synthase
Carbamoyl-P
Nitrate metabolism
Respiratory nitrate reductase
Nitroalkane
Hydroxylamine DH
Hydroxylamine
Ammonia monooxygenase
Nitrilase
Nitrile
Glutamate DH
Glutamate
Sulfate metabolism
SOX
Thiosulfate
Sulfate
Sulfite DH
Sulfite
Sulfite reductase
Sulfide
Sulfur dioxygenase
Sulfide DH
Sulfur
APS
PAPS
PAP
Pentose-Phosphate-Pathway
Ribose-5P
Ribose-5P Isomerase
Transketolase
Ribulose-5P
Ribulose-5P 3-epimerase
Xylulose-5P
Glyceraldehyde-3P + Sedoheptulose-7P
Transaldolase
Fructose-6P + Erythrose-4P
Xylulose-5P
Transketolase
Glycolysis/ Gluconeogenesis
Glyceraldehyde-3P + Fructose-6P
Phospho ribulokinase
Glyceraldehyde-3-P DH
Ribulose-1,5-bisP
CO2
Ribulose-1,5-bisphosphate carboxylase
Glycerate-3P
Phosphoglycerate kinase
1,3-bisP-glycerate
reductive Pentose-Phosphate-Pathway
TCA
Pyruvate
Malic enzyme
OAA
Acetyl-CoA
Citrate synthase
Citrate
Malate-DH
L-Malate
Fumarase
Fumarate
Succinate-DH
Succinate
Succinyl-CoA synthetase
Succinyl-CoA
2-oxoglutarate synthase
Oxoglutarate
Isocitrate-DH
Isocitrate
Aconitase
TCA
Succinate
Fumarate
CI
CII
CIII
CIV
CV
ADP
ATP
H+
oxidative phosphorylation

Figure S12

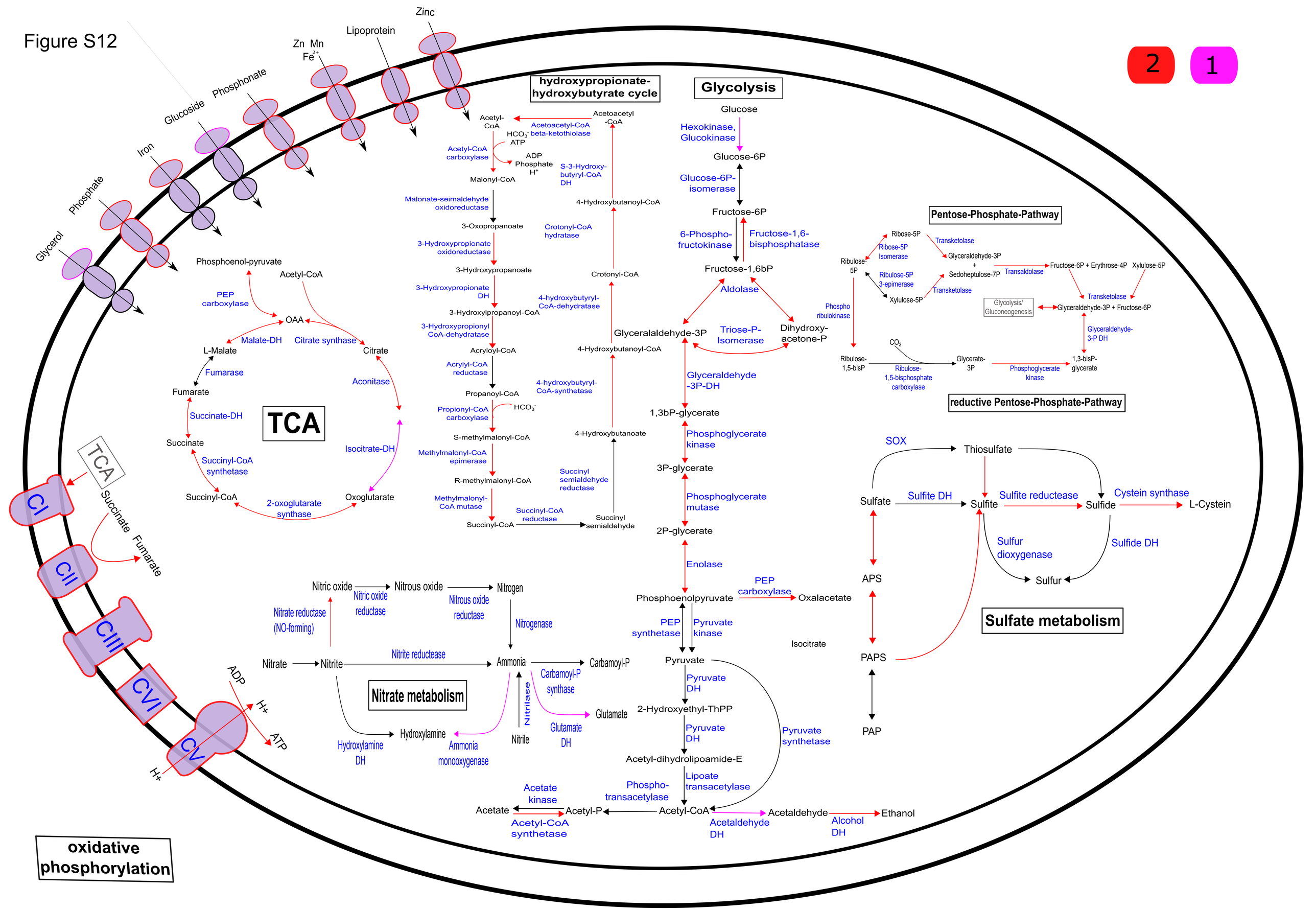

2
1
Zinc
Lipoprotein
Zn Mn
Fe2+
Phosphonate
Glucoside
Iron
Phosphate
Glycerol
hydroxypropionate-hydroxybutyrate cycle
Acetyl-CoA
Acetoacetyl-CoA
Acetoacetyl-CoA beta-ketothiolase
HCO3-
ATP
Acetyl-CoA carboxylase
ADP
Phosphate
H+
S-3-Hydroxy-butyryl-CoA DH
Malonyl-CoA
Malonate-seimaldehyde oxidoreductase
4-Hydroxybutanoyl-CoA
3-Oxopropanoate
Crotonyl-CoA hydratase
3-Hydroxypropionate oxidoreductase
3-Hydroxypropanoate
Crotonyl-CoA
3-Hydroxypropionate DH
4-hydroxybutyryl-CoA-dehydratase
3-Hydroxylpropanoyl-CoA
3-Hydroxypropionyl CoA-dehydratase
Acryloyl-CoA
4-Hydroxybutanoyl-CoA
Acrylyl-CoA reductase
4-hydroxybutyryl-CoA-synthetase
Propanoyl-CoA
Propionyl-CoA carboxylase
HCO3-
S-methylmalonyl-CoA
4-Hydroxybutanoate
Methylmalonyl-CoA epimerase
R-methylmalonyl-CoA
Succinyl semialdehyde reductase
Methylmalonyl-CoA mutase
Succinyl-CoA reductase
Succinyl-CoA
Succinyl semialdehyde
Glycolysis
Glucose
Hexokinase, Glucokinase
Glucose-6P
Glucose-6P-isomerase
Fructose-6P
6-Phospho-fructokinase
Fructose-1,6-bisphosphatase
Fructose-1,6bP
Aldolase
Glyceralaldehyde-3P
Triose-P-Isomerase
Dihydroxy-acetone-P
Glyceraldehyde -3P-DH
1,3bP-glycerate
Phosphoglycerate kinase
3P-glycerate
Phosphoglycerate mutase
2P-glycerate
Enolase
Phosphoenolpyruvate
PEP carboxylase
Oxalacetate
PEP synthetase
Pyruvate kinase
Pyruvate
Pyruvate DH
2-Hydroxyethyl-ThPP
Pyruvate DH
Acetyl-dihydrolipoamide-E
Lipoate transacetylase
Phospho-transacetylase
Acetyl-CoA
Pyruvate synthetase
Acetate kinase
Acetate
Acetyl-P
Acetyl-CoA synthetase
Acetaldehyde
Acetaldehyde DH
Alcohol DH
Ethanol
Pentose-Phosphate-Pathway
Ribose-5P
Ribose-5P Isomerase
Ribulose-5P
Transketolase
Glyceraldehyde-3P + Sedoheptulose-7P
Transaldolase
Fructose-6P + Erythrose-4P
Xylulose-5P
Ribulose-5P 3-epimerase
Xylulose-5P
Transketolase
Glycolysis/ Gluconeogenesis
Transketolase
Glyceraldehyde-3P + Fructose-6P
Phospho ribulokinase
Glyceraldehyde-3-P DH
CO2
Ribulose-1,5-bisP
Ribulose-1,5-bisphosphate carboxylase
Glycerate-3P
Phosphoglycerate kinase
1,3-bisP-glycerate
reductive Pentose-Phosphate-Pathway
TCA
Phosphoenol-pyruvate
Acetyl-CoA
PEP carboxylase
OAA
Malate-DH
Citrate synthase
L-Malate
Citrate
Fumarase
Aconitase
Fumarate
Succinate-DH
Succinate
Isocitrate-DH
Succinyl-CoA synthetase
Succinyl-CoA
2-oxoglutarate synthase
Oxoglutarate
TCA
Succinate
Fumarate
CI
CII
CIII
CVI
CV
ADP
H+
ATP
H+
oxidative phosphorylation
Nitric oxide
Nitrous oxide
Nitrogen
Nitrate reductase (NO-forming)
Nitric oxide reductase
Nitrous oxide reductase
Nitrogenase
Nitrate
Nitrite
Nitrite reductase
Ammonia
Carbamoyl-P
Carbamoyl-P synthase
Nitrate metabolism
Nitrilase
Glutamate
Glutamate DH
Hydroxylamine
Hydroxylamine DH
Ammonia monooxygenase
Nitrile
SOX
Thiosulfate
Sulfate
Sulfite DH
Sulfite
Sulfite reductase
Sulfide
Cystein synthase
L-Cystein
Sulfur dioxygenase
Sulfide DH
Sulfur
APS
PAPS
PAP
Sulfate metabolism

Figure S13

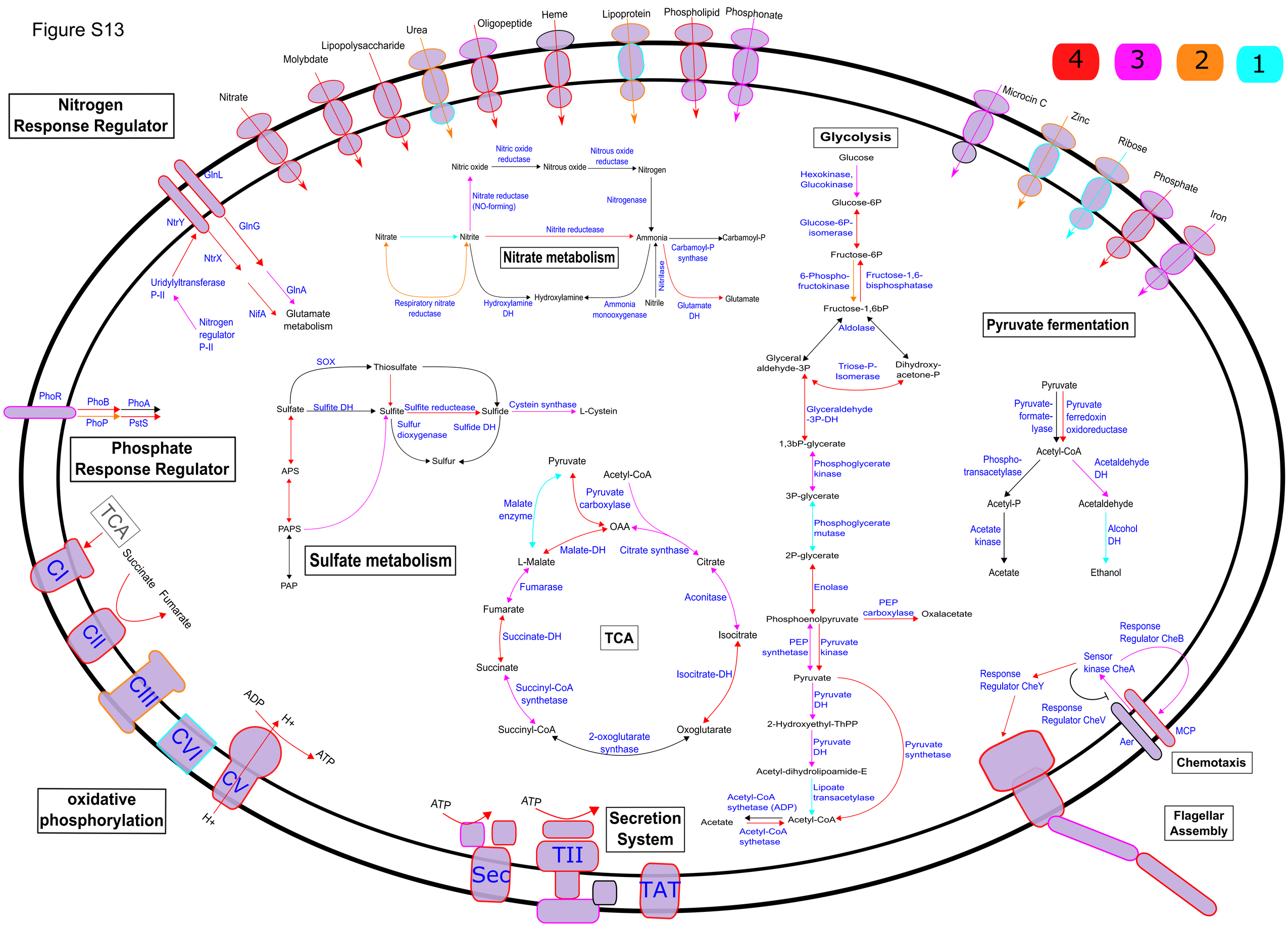
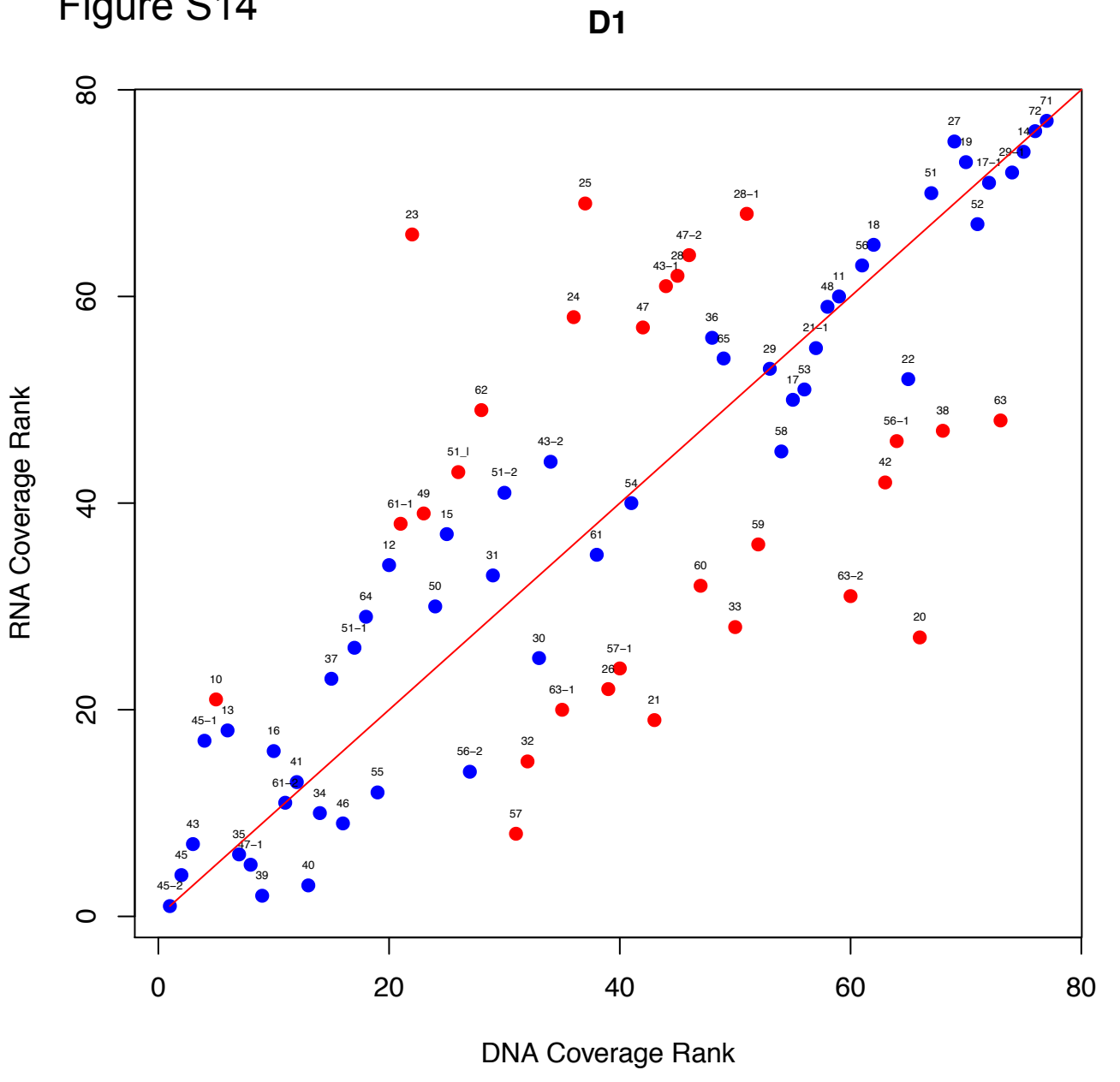
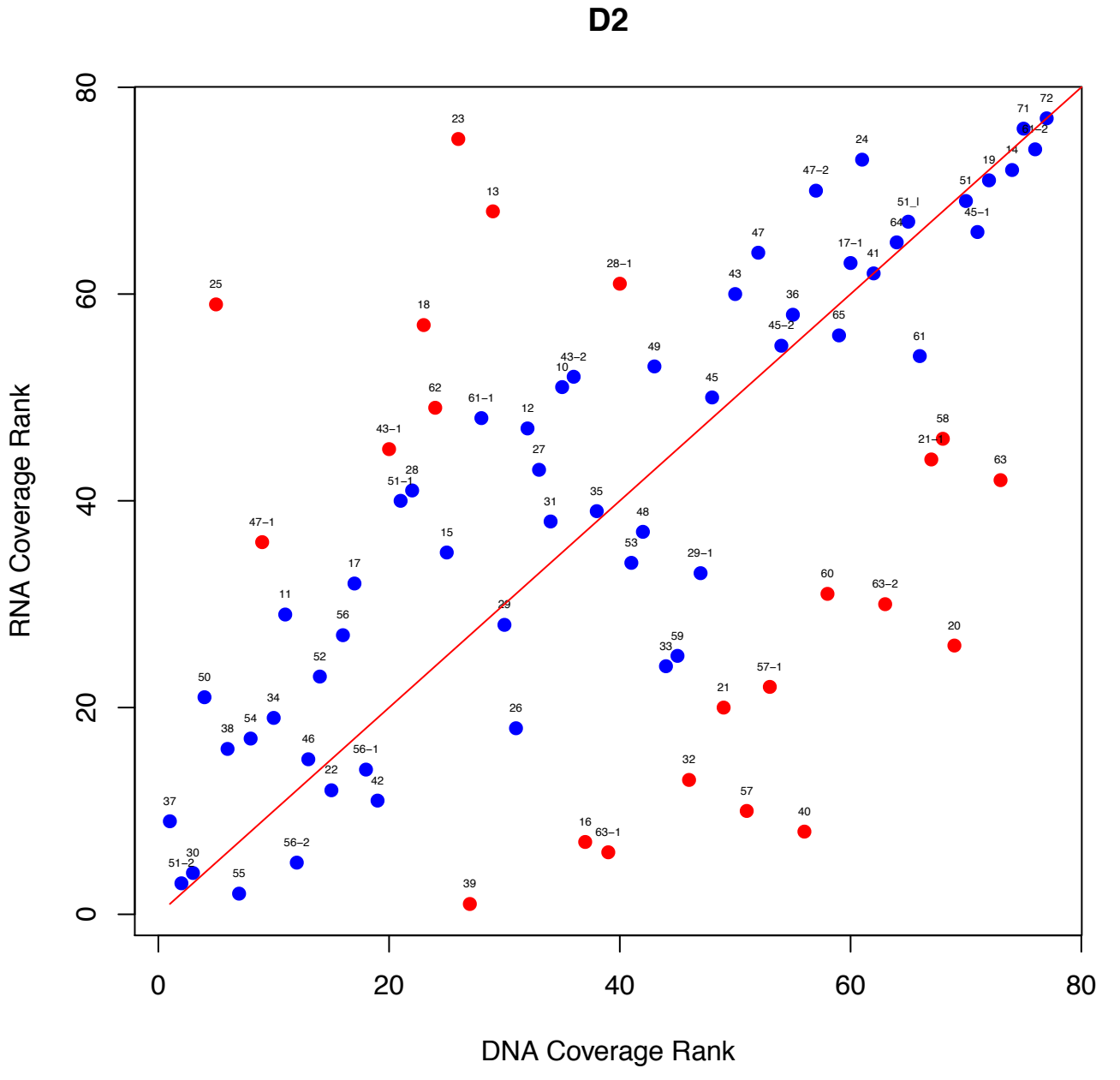
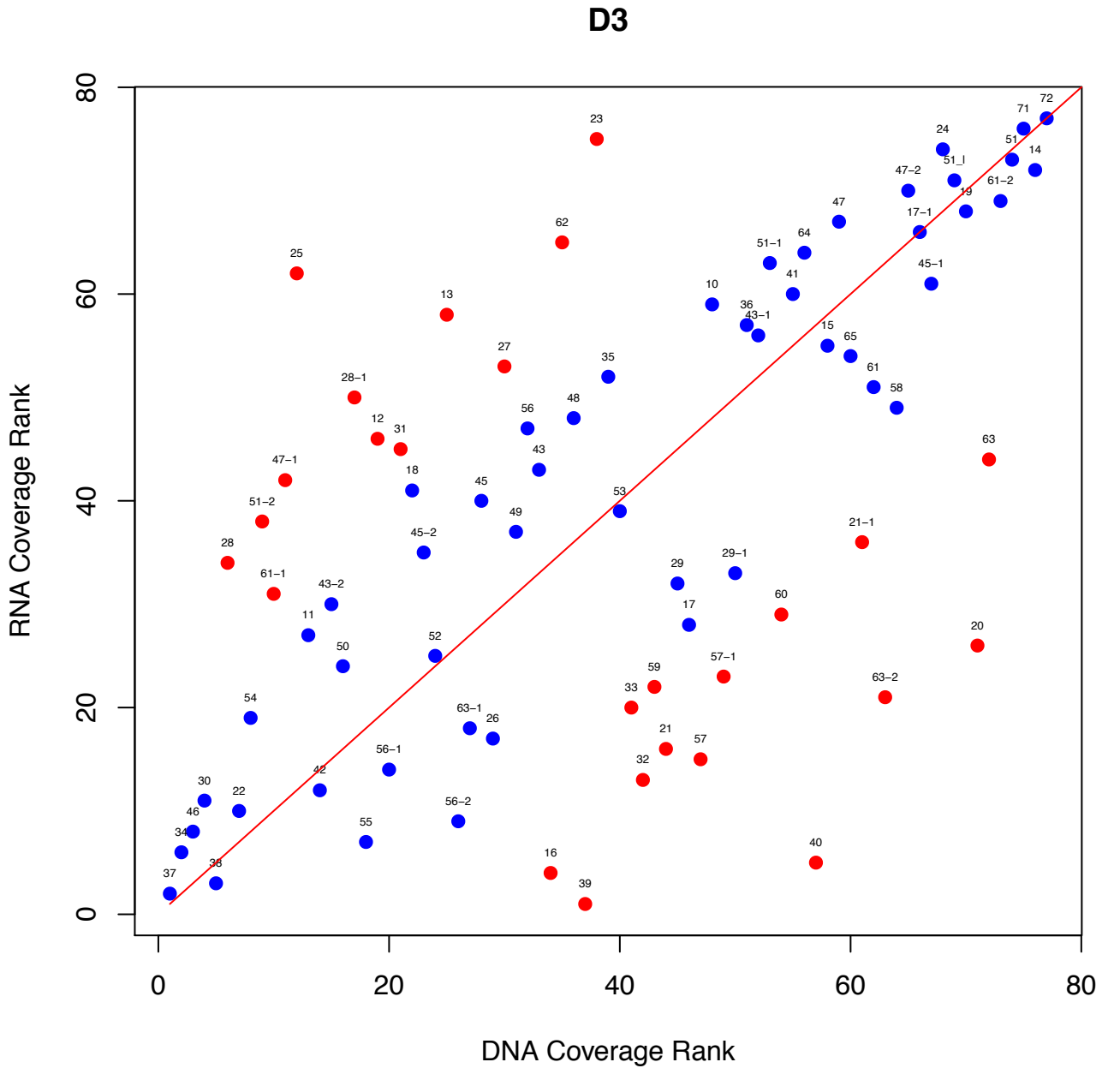
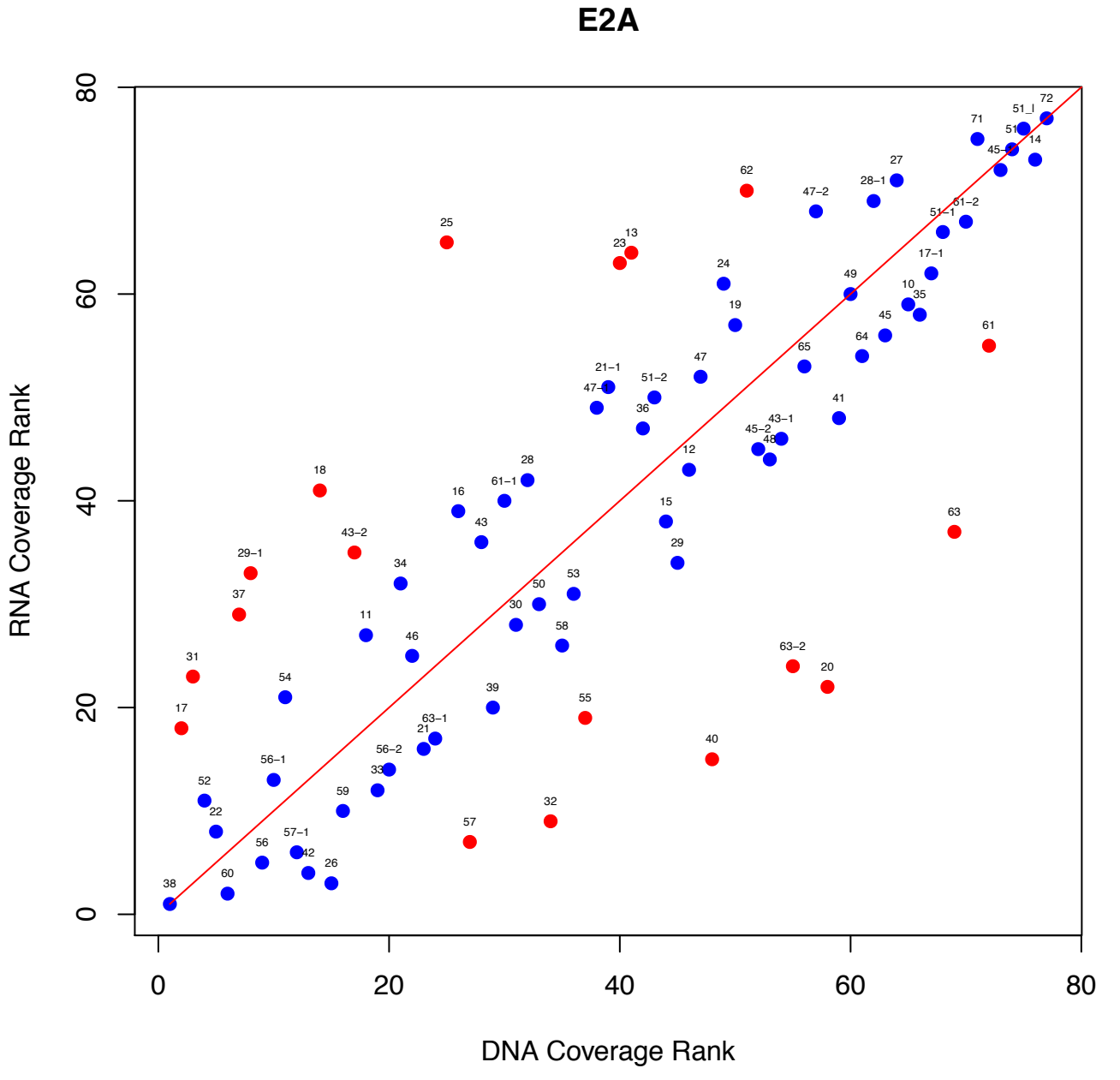

4
3
2
1
Nitrogen Response Regulator
Nitrate
Molybdate
Lipopolysaccharide
Urea
Oligopeptide
Heme
Lipoprotein
Phospholipid
Phosphonate
Microcin C
Zinc
Ribose
Phosphate
Iron
GlnL
NtrY
GlnG
NtrX
Uridylyltransferase P-II
Nitrogen regulator P-II
GlnA
NifA
Glutamate metabolism
Nitric oxide reductase
Nitrous oxide reductase
Nitric oxide
Nitrous oxide
Nitrogen
Nitrate reductase (NO-forming)
Nitrogenase
Nitrate
Nitrite
Nitrite reductase
Ammonia
Carbamoyl-P
Carbamoyl-P synthase
Nitrate metabolism
Nitrilase
Respiratory nitrate reductase
Hydroxylamine DH
Hydroxylamine
Ammonia monooxygenase
Nitrile
Glutamate DH
Glutamate
Glycolysis
Glucose
Hexokinase, Glucokinase
Glucose-6P
Glucose-6P-isomerase
Fructose-6P
6-Phospho-fructokinase
Fructose-1,6-bisphosphatase
Fructose-1,6bP
Aldolase
Glyceral aldehyde-3P
Triose-P-Isomerase
Dihydroxy-acetone-P
Glyceraldehyde -3P-DH
1,3bP-glycerate
Phosphoglycerate kinase
3P-glycerate
Phosphoglycerate mutase
2P-glycerate
Enolase
Phosphoenolpyruvate
PEP carboxylase
Oxalacetate
PEP synthetase
Pyruvate kinase
Pyruvate
Pyruvate DH
2-Hydroxyethyl-ThPP
Pyruvate DH
Acetyl-dihydrolipoamide-E
Lipoate transacetylase
Pyruvate synthetase
Acetyl-CoA sythetase (ADP)
Acetate
Acetyl-CoA
Acetyl-CoA sythetase
Pyruvate fermentation
Pyruvate
Pyruvate-formate-lyase
Pyruvate ferredoxin oxidoreductase
Acetyl-CoA
Phospho-transacetylase
Acetaldehyde DH
Acetyl-P
Acetaldehyde
Acetate kinase
Alcohol DH
Acetate
Ethanol
PhoR
PhoB
PhoA
PhoP
PstS
Phosphate Response Regulator
SOX
Thiosulfate
Sulfate
Sulfite DH
Sulfite
Sulfite reductase
Sulfide
Cystein synthase
L-Cystein
Sulfur dioxygenase
Sulfide DH
Sulfur
APS
PAPS
PAP
Sulfate metabolism
Pyruvate
Acetyl-CoA
Pyruvate carboxylase
Malate enzyme
OAA
Malate-DH
Citrate synthase
L-Malate
Citrate
Fumarase
Aconitase
Fumarate
Succinate-DH
TCA
Isocitrate
Succinate
Isocitrate-DH
Succinyl-CoA synthetase
Succinyl-CoA
2-oxoglutarate synthase
Oxoglutarate
TCA
Succinate
Fumarate
CI
CII
CIII
CVI
CV
ADP
ATP
H+
H+
oxidative phosphorylation
ATP
ATP
Sec
TII
TAT
Secretion System
Response Regulator CheB
Sensor kinase CheA
Response Regulator CheY
Response Regulator CheV
Aer
MCP
Chemotaxis
Flagellar Assembly

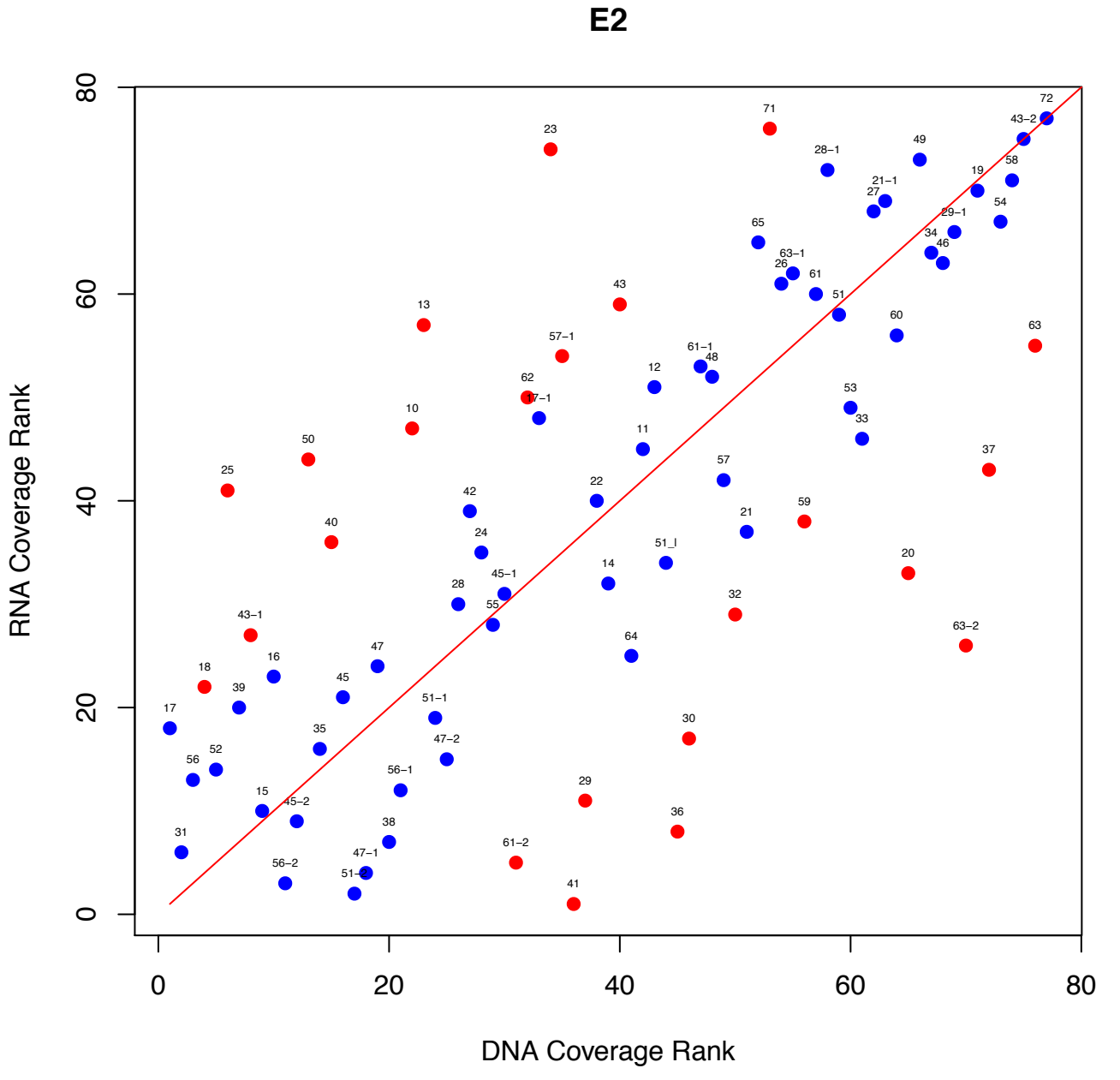
Figure S14


D1
RNA Coverage Rank
DNA Coverage Rank
0
20
40
60
80
0
20
40
60
80
71
72
14
27
19
29
17–1
51
52
25
28–1
23
18
56
47–2
28
43–1
11
48
24
47
36
65
21–1
29
22
53
17
62
63
38
56–1
58
43–2
51_l
51–2
42
54
49
61–1
15
59
61
12
31
60
63–2
50
64
33
20
51–1
30
57–1
37
26
10
63–1
21
13
45–1
16
32
56–2
41
55
61–2
34
46
57
43
35
47–1
45
39
40
45–2


D2
RNA Coverage Rank
DNA Coverage Rank
0
20
40
60
80
0
20
40
60
80
72
71
61–2
14
19
51
45–1
51_l
64
17–1
41
24
47–2
47
43
36
45–2
65
61
23
13
28–1
25
18
49
43–2
10
45
62
61–1
12
43–1
27
58
21–1
63
28
51–1
31
35
48
47–1
15
53
29–1
17
60
63–2
11
56
29
20
59
33
52
57–1
50
21
34
26
54
38
46
56–1
32
22
42
57
37
40
16
63–1
56–2
30
51–2
55
39


D3
RNA Coverage Rank
DNA Coverage Rank
0
20
40
60
80
0
20
40
60
80
72
71
24
51
14
51_l
47–2
61–2
19
47
17–1
23
62
64
51–1
25
41
45–1
10
13
36
43–1
15
65
27
35
61
28–1
58
48
56
12
31
63
43
47–1
18
45
53
51–2
49
21–1
45–2
28
29–1
29
61–1
43–2
60
17
11
20
52
50
57–1
59
63–2
33
54
63–1
26
21
57
56–1
32
42
30
22
56–2
46
55
34
40
16
38
37
39


E2A
RNA Coverage Rank
DNA Coverage Rank
0
20
40
60
80
0
20
40
60
80
72
51_I
71
51
14
45–
27
62
28–1
47–2
61–2
51–1
25
13
23
17–1
24
49
10
35
19
45
61
64
65
47
21–1
51–2
47–1
36
41
43–1
45–2
48
12
28
18
61–1
16
15
63
43
43–2
29
29–1
34
53
50
37
30
11
58
46
63–2
31
20
54
39
55
17
63–1
21
40
56–2
56–1
33
52
59
32
22
57
57–1
56
42
26
60
38


E2
RNA Coverage Rank
DNA Coverage Rank
0
20
40
60
80
72
43−2
71
23
49
28−1
58
19
21−1
27
54
29−1
65
34
46
63−1
26
61
43
13
51
60
63
57−1
61−1
48
12
62
17−1
53
33
10
11
50
57
37
25
22
42
21
59
40
24
51_l
20
14
45−1
28
32
43−1
55
63−2
64
47
16
18
45
39
51−1
17
30
35
47−2
52
56
56−1
29
15
45−2
36
38
31
61−2
47−1
56−2
51−2
41

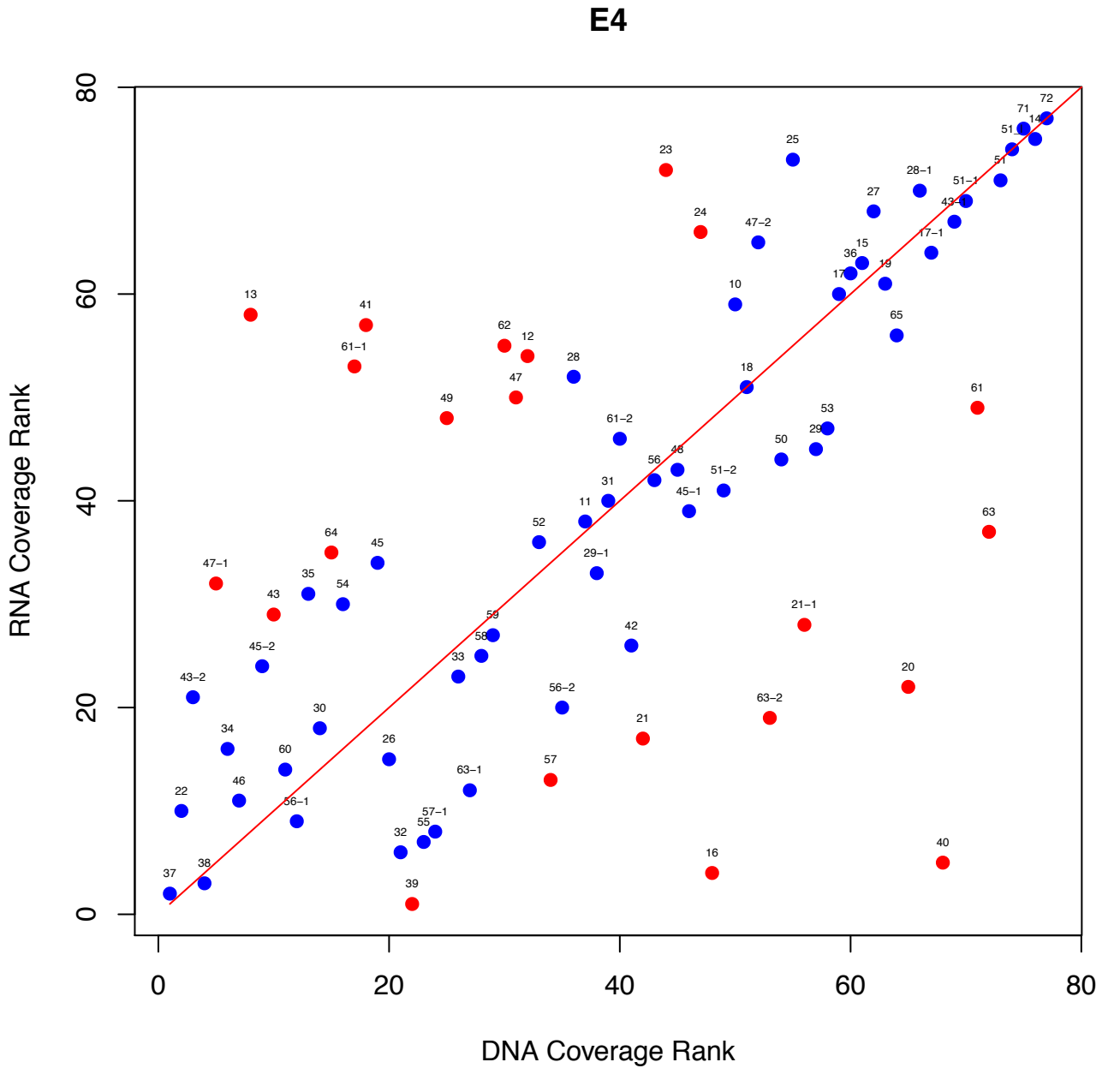
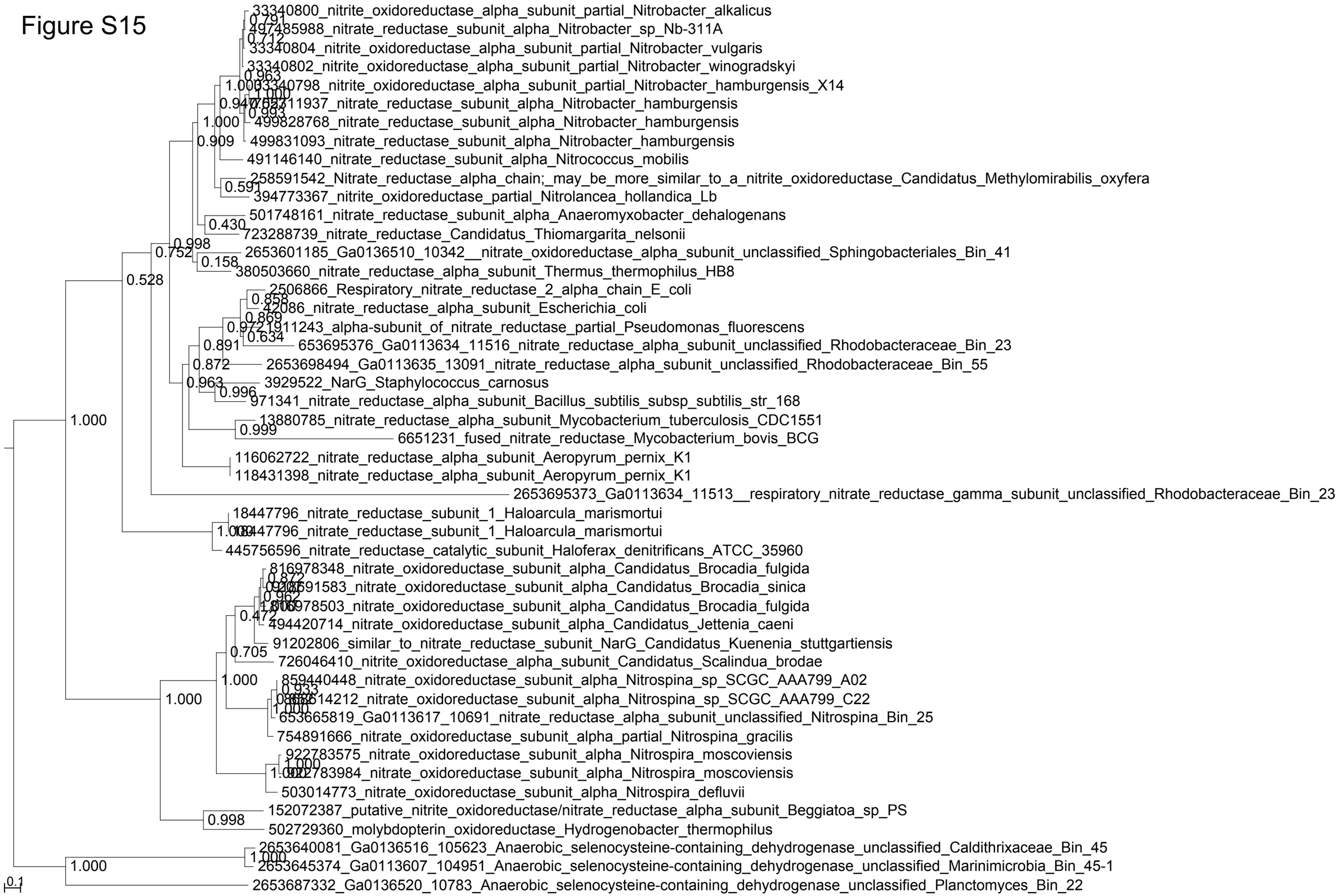

E4
RNA Coverage Rank
DNA Coverage Rank
0
20
40
60
80
0
20
40
60
80
72
71
14
51
51
23
25
28–1
51–1
27
43–1
24
47–2
17–1
15
36
19
17
10
13
41
65
62
12
61–1
28
18
47
61
49
53
61–2
29
50
48
56
51–2
31
45–1
11
63
52
64
45
29–1
47–1
35
54
43
21–1
59
42
58
45–2
33
20
43–2
56–2
63–2
30
21
34
26
60
57
63–1
46
22
56–1
57–1
55
32
40
16
38
37
39

Figure S15


33340800_nitrite_oxidoreductase_alpha_subunit_partial_Nitrobacter_alkalicus
497485988_nitrate_reductase_subunit_alpha_Nitrobacter_sp_Nb-311A
33340804_nitrite_oxidoreductase_alpha_subunit_partial_Nitrobacter_vulgaris
33340802_nitrite_oxidoreductase_alpha_subunit_partial_Nitrobacter_winogradskyi
33340798_nitrite_oxidoreductase_alpha_subunit_partial_Nitrobacter_hamburgensis_X14
499828768_nitrate_reductase_subunit_alpha_Nitrobacter_hamburgensis
499831093_nitrate_reductase_subunit_alpha_Nitrobacter_hamburgensis
491146140_nitrate_reductase_subunit_alpha_Nitrococcus_mobilis
258591542_Nitrate_reductase_alpha_chain;_may_be_more_similar_to_a_nitrite_oxidoreductase_Candidatus_Methylomirabilis_oxyfera
394773367_nitrite_oxidoreductase_partial_Nitrolancea_hollandica_Lb
501748161_nitrate_reductase_subunit_alpha_Anaeromyxobacter_dehalogenans
723288739_nitrate_reductase_Candidatus_Thiomargarita_nelsonii
2653601185_Ga0136510_10342__nitrate_oxidoreductase_alpha_subunit_unclassified_Sphingobacteriales_Bin_41
380503660_nitrate_reductase_alpha_subunit_Thermus_thermophilus_HB8
2506866_Respiratory_nitrate_reductase_2_alpha_chain_E_coli
42086_nitrate_reductase_alpha_subunit_Escherichia_coli
653695376_Ga0113634_11516_nitrate_reductase_alpha_subunit_unclassified_Rhodobacteraceae_Bin_23
2653698494_Ga0113635_13091_nitrate_reductase_alpha_subunit_unclassified_Rhodobacteraceae_Bin_55
3929522_NarG_Staphylococcus_carnosus
971341_nitrate_reductase_alpha_subunit_Bacillus_subtilis_subsp_subtilis_str_168
13880785_nitrate_reductase_alpha_subunit_Mycobacterium_tuberculosis_CDC1551
6651231_fused_nitrate_reductase_Mycobacterium_bovis_BCG
116062722_nitrate_reductase_alpha_subunit_Aeropyrum_pernix_K1
118431398_nitrate_reductase_alpha_subunit_Aeropyrum_pernix_K1
2653695373_Ga0113634_11513__respiratory_nitrate_reductase_gamma_subunit_unclassified_Rhodobacteraceae_Bin_23
18447796_nitrate_reductase_subunit_1_Haloarcula_marismortui
445756596_nitrate_reductase_catalytic_subunit_Haloferax_denitrificans_ATCC_35960
816978348_nitrate_oxidoreductase_subunit_alpha_Candidatus_Brocadia_fulgida
494420714_nitrate_oxidoreductase_subunit_alpha_Candidatus_Jettenia_caeni
91202806_similar_to_nitrate_reductase_subunit_NarG_Candidatus_Kuenenia_stuttgartiensis
726046410_nitrite_oxidoreductase_alpha_subunit_Candidatus_Scalindua_brodae
859440448_nitrate_oxidoreductase_subunit_alpha_Nitrospina_sp_SCGC_AAA799_A02
653665819_Ga0113617_10691_nitrate_reductase_alpha_subunit_unclassified_Nitrospina_Bin_25
754891666_nitrate_oxidoreductase_subunit_alpha_partial_Nitrospina_gracilis
922783575_nitrate_oxidoreductase_subunit_alpha_Nitrospira_moscoviensis
503014773_nitrate_oxidoreductase_subunit_alpha_Nitrospira_defluvii
152072387_putative_nitrite_oxidoreductase/nitrate_reductase_alpha_subunit_Beggiatoa_sp_PS
502729360_molybdopterin_oxidoreductase_Hydrogenobacter_thermophilus
2653640081_Ga0136516_105623_Anaerobic_selenocysteine-containing_dehydrogenase_unclassified_Caldithrixaceae_Bin_45
2653645374_Ga0113607_104951_Anaerobic_selenocysteine-containing_dehydrogenase_unclassified_Marinimicrobia_Bin_45-1
2653687332_Ga0136520_10783_Anaerobic_selenocysteine-containing_dehydrogenase_unclassified_Planctomyces_Bin_22
0.1