Depth-discrete eco-genomics of Lake Tanganyika reveals roles of diverse microbes, including candidate phyla, in tropical freshwater nutrient cycling

Patricia Q. Tran[1,2], Peter B. McIntyre[3], Benjamin M. Kraemer[4], Yvonne Vadeboncoeur[5], Ismael A. Kimirei[6], Rashid Tamatamah[6], Katherine D. McMahon[1,7], Karthik Anantharaman[1]

1. Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA
2. Department of Integrative Biology, University of Wisconsin-Madison, Madison, WI, USA
3. Department of Ecology and Evolution, Cornell University, Ithaca, NY, USA
4. Department of Ecosystem Research, Leibniz Institute for Freshwater Ecology and Inland Fisheries, Berlin, Germany
5. Department of Biological Sciences, Wright State University, Dayton, OH, USA
6. Tanzania Fisheries Research Institute (TAFIRI), Dar es Salaam, Tanzania
7. Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI, USA

Corresponding author: Karthik Anantharaman (karthik@bact.wisc.edu)

Abstract

Lake Tanganyika, the largest tropical freshwater lake and the second largest by volume on Earth is characterized by strong oxygen and redox gradients. In spite of the majority of its water column being anoxic, Tanganyika hosts some of the most diverse and prolific fisheries and ecosystems on Earth. Yet, little is known about microorganisms inhabiting this lake, and their impacts on biogeochemistry and nutrient cycling underlying ecosystem structure and productivity. Here, we apply depth-discrete metagenomics, single-cell genomics, and environmental analyses to reconstruct and characterize 3996 microbial genomes representing 802 non-redundant organisms from 81 bacterial and archaeal phyla, including two novel bacterial candidate phyla, Tanganyikabacteria and Ziwabacteria. We found sharp contrasts in community composition and metabolism between the oxygenated mixed upper layer compared to deep anoxic waters, with core freshwater taxa in the former, and Archaea and uncultured Candidate Phyla in the latter. Microbially-driven nitrogen cycling increased in the anoxic zone, highlighting microbial contribution to the productive surface layers, via production of upwelled nutrients, and greenhouse gases such as nitrous oxide. Overall, we provide a window into how oxygen gradients shape microbial community metabolism in widespread anoxic tropical freshwaters, and advocate for the importance of anoxic freshwater habitats in the context of global biogeochemical cycles and nutrient cycling.

Introduction

Located in the East African Rift Valley, Lake Tanganyika (LT) holds 16% of the Earth's freshwater, and is the second largest lake by volume[1]. By its sheer size and magnitude, LT exerts a major influence on biogeochemical cycling on regional and global scales[2,3]. Over the past centuries, LT has been extensively studied as a model system for its rich animal biodiversity[4,5] which have revealed important insights on species radiation and evolution. In contrast, the microbial community of LT that drives the productivity of this ecosystem largely remains a mystery.

LT provides a unique ecosystem to study microbial diversity and function in freshwater lakes, specifically tropical lakes. This ancient lake harbors some of the most spectacular and well-studied adaptive radiations of metazoan species diversity on Earth[5], but these organisms exist in a thin layer of oxygenated surface water. Lake Tanganyika is estimated to be at least 10 million years old and is considered oligotrophic (low nutrient concentrations). Among Lake Tanganyika's defining features is that approximately 80% of the 1890 km^3 of water is anoxic. Being meromictic, its water column is permanently stratified, with a large volume of anoxic and nutrient-rich bottom waters (hypolimnion), separated from the upper ~70m of well-lit, nutrient-depleted water (epilimnion). Despite stratification, periodic pulses of phosphorus, nitrogen and iron from upwelling of deep-waters replenish the epilimnion and sustain its productivity. Lake Tanganyika's influence in biogeochemical cycling regionally and globally is valuable, for instance it stores over 23 TG of methane below the oxycline[3], and stores about 14,000,000Tg of C in its sediments[2].

Only a handful of previous work has been done on the lake's microbial ecology[6,7]. Previous studies have found microbial community composition to be heterogeneous along both vertical and horizontal spans[7]. The differences were primarily related to thermal stratification, which leads to strong gradients in oxygen and nutrient concentrations. However, the contribution of microbes to nutrient cycling in Lake Tanganyika remains to be studied. Here we investigated the microbial community composition, metabolic interactions, and biogeochemical contributions along ecological gradients from high light, oxygenated surface waters to dark, oxygen-free and nutrient-rich bottom waters of Lake Tanganyika. Our comprehensive analyses include genome-resolved metagenomics and single-cell genomics to generate thousands of bacterial and archaeal genomes, metabolic reconstructions at the resolution of individual organisms and the entire lake ecosystem, and measurements of lake biogeochemistry to infer the roles of microorganisms in nutrient cycling across different layers in the water column. Our work offers a window into the understudied microbial diversity of LT and serves as a case study for investigating microbial roles and links to biogeochemistry in globally distributed anoxic freshwater lakes.

Results

We collected 24 samples from the LT water column spanning 0 to 1200m, at two stations (Kigoma: 13 samples, Mahale: 11 samples) in July and October 2015 (Table S1). Detailed environmental profiles from 2010-2013 were used to guide sampling location and depth (Figure

1A). Water column temperatures ranged from 24 to 28°C, and changes in dissolved oxygen (DO) were greatest at depths ranging from ~50 to 100 m during the time of sampling (Figure 1A, Figure 1B), dropping to 0% DO around 100m. A consistent chlorophyll-a peak was detected at a depth of ~120m throughout the years (Figure S1). Nitrate concentrations increased up to ~100 µg/L at a depth of 100m, followed by rapid depletion with the onset of anoxia (Figure 1D). Secchi depth, a measure of how deep light penetrates through the water column was on average 12.2m in July and 12.0m in October (Figure S2). Despite not having paired environmental data in 2015, environmental data from 2010 to 2013 was highly consistent and showed minimal interannual variability, particularly during our sampling period (July and October) (Figure S1, Figure S2). Additionally, the environmental profiles are similar to those previously collected (see references [1,8,9]).

Metagenomic sequencing and binning of 24 individual samples resulted in 3948 draft-quality metagenome-assembled genomes (MAGs) that were dereplicated (using a threshold of 98% average nucleotide identity (ANI)) into a set of 802 non-redundant genomes for downstream analyses (Table S2, Table S7). We also sequenced 48 single-cell amplified genomes (SAGs) from a depth of 1200m to complement the MAGs. To assign taxonomic classifications to the organisms represented by the genomes, we combined two complimentary genome-based phylogenetic approaches: a manual 16 ribosomal protein concatenated gene phylogeny (16RP) and GTDB-tk[10], an automated program which uses 120 concatenated protein coding genes. Overall, we observed congruence between the two approaches (Table S2). Six bacterial genomes clustered away from known genomes and likely represent two monophyletic lineages at the phylum-level. On this basis, we propose the candidate phyla (CP) *Candidatus Tanganyikabacteria* (named after the lake) and *Candidatus Ziwabacteria* (from the Swahili word for "lake" which is "ziwa"). Overall, our genomes represented 34 Archaea and 769 Bacteria from 81 phyla, including our two proposed phyla (Figure 2, Table S2).

To elucidate the stratification of microbial populations and metabolic processes in the water column of LT, we identified three zones referred to as "oxic", "sub-oxic", and "anoxic" which we operationally define based on general oxygen percent saturation (Figure 1C). These zones coincidently correspond broadly to the epilimnion, metalimnion and hypolimnion respectively. To relate microbial community composition to relative abundance in our samples, we mapped metagenomic reads from each depth-discrete metagenome against all 802 MAGs and computed the average normalized coverage (abundance) (Table S3).

The microbiome of Lake Tanganyika is shaped by oxygen and redox gradients

Broadly speaking, the tropical microbiome of LT was similar to the epilimnions of temperate worldwide[11]. Major phylogenetic groups found in high abundance included *Actinobacteria*, *Bacteroidetes*, *Betaproteobacteria*, *Cyanobacteria*, *Gammaproteobacteria*, *Planctomycetes* and *Verrucomicrobia* (Figure 2), many of which are cosmopolitan and canonical freshwater lineages. However, LT also hosts organisms less frequently observed elsewhere in freshwater systems such as *Ignavibacteria* and *Thaumarchaeota*. *Ignavibacteria* were among the

117 most abundant organisms in 23 out of 24 samples while *Thaumarchaeota* were present at all
118 depths but peaked in the sub-oxic zone.
119 *Cyanobacteria* constituted the most abundant phylum across all our samples (Figure 3,
120 Table S2). Cyanobacterial MAGs were also abundant in the hypolimnion, including the deepest
121 samples. While these *Cyanobacteria* encoded both photosystems I and II, we postulate that these
122 organisms likely conduct dark fermentative metabolism below the photic zone similar to that
123 observed in *Cyanobacteria* from deep terrestrial ecosystems[12]. Analysis of replication rates[13]
124 suggests that *Cyanobacteria* are indeed growing, but we cannot quantify absolute rates without
125 calibration using cultures. Cells were estimated to be replicating more quickly in the photic zone
126 as compared to the deep hypolimnion (Figure S3).
127 The dominant community members became more distinct with depth, as oxygen and light
128 limitations increased (Figure 1, Figure 3, Figure S1, Figure S2). We noticed some differences
129 between stations, but hesitate to attribute these to local factors without more extensive spatial
130 sampling. The presence of *Euryarchaeota* in the epilimnion in Mahale might be a result from the
131 upwelling, though more temporally resolved sampling is required to properly examine this.
132 Green sulfur bacteria (phylum *Chlorobi*) were the sixteenth most abundant group in samples just
133 below the sub-oxic zone (Figure S1). *Chlorobi* are observed to be highly abundant in other
134 freshwater lakes where strong light and sulfide gradients exist, such as the meromictic Lake
135 Cadagno[14,15]. We hypothesize that the lower abundance of *Chlorobi* observed in LT at and in the
136 anoxic zone may be attributed to the presence of anoxygenic *Cyanobacteria* and competition
137 between these groups (Figure S4). Nitrifying *Nitrospira* were present but only below 100m. We
138 identified a total of ninety-five organisms from CP with Eisenbacteria, Kaiserbacteria and
139 Rokubacteria comprising the most abundant lineages below 100m. Among CP, only
140 Tanganyikabacteria, Tectomicrobia and TM6 were present in both the oxic and anoxic zones.
141 Much remains to be learnt about microbial biodiversity inhabiting lakes' hypolimnions
142 and deep lakes worldwide. With few microbial ecology studies focused on lake hypolimnions[16–
143 19], we cannot speculate if microbes in the hypolimnion of Lake Tanganyika are truly endemic.
144 Nevertheless, we note the contrast between typical freshwater microbes dominating the
145 epilimnion, and the high prominence of Archaea and CP, including Candidate Phyla Radiation
146 (CPR) in the anoxic zones.
147
148 ***Shared microbiome between tropical and temperate freshwater lakes***
149 We examined MAGs from the well-sampled and globally distributed bacterial lineage,
150 *Actinobacteria* to ask whether endemic strains were present in Lake Tanganyika. Two specific
151 lineages (acI, followed by acIV) within the Actinobacteria are typically constrained to freshwater
152 systems and are cosmopolitan[11]. Previously, acI lineages were found in Lake Tanganyika by 16S
153 rRNA gene sequencing[7]. In comparison to other available acI from lakes in Japan, Wisconsin
154 and Switzerland[20,21], Lake Tanganyika MAGs only cover a small fraction of the Actinobacteria
155 phylogenetic diversity, dominated by freshwater groups acI-B1, acI-C, acIII, acIV, acSTL, acTH
156 and Iluma-A (Table S2). Overall, *Actinobacteria* lineages show possible niche partitioning with

157 depth in Lake Tanganyika (Figure S5). All from LT were most abundant in the epilimnion,
158 except acIII, acIV and two Iluma-A2. Non-canonical freshwater groups were present throughout
159 the water column. Several cosmopolitan clade members were virtually indistinguishable in the
160 16RP phylogeny from genomes recovered in US and European lakes (acI lineage), while others
161 constituted unique groups previously unsampled in other lakes (Figure S5).
162
163 ***Archaea and candidate phyla are prominent in the anoxic zone of Lake Tanganyika***
164 Little is known about the distribution and ecology of organisms belonging to CP and the
165 CPR in freshwater lakes. The CPR consists of mostly uncultivated bacteria and are
166 phylogenetically distinct from major bacterial lineages[13,22]. In a hypersaline soda lake,
167 uncultured CPR were found to be a major component of the bacterial community. Similarly,
168 Parcubacteria dominated the anaerobic bacterial community of boreal lakes[23].
169 Amongst non-CPR candidate phyla, Candidatus Aegiribacteria was first described below the
170 chemocline of Lake Mahoney (British-Columbia, Canada) in the euxinic hypolimnion, which is
171 anoxic, saline and sulfur-rich.
172 Here, we reconstructed 95 CP genomes (40 phyla, including CP Tanganyikabacteria and
173 CP Ziwabacteria) (Figure 2A) including 22 genomes from CPR organisms. Thirty-seven archaeal
174 MAGs accounted for 3.4% of microbial abundance (Figure 2A). We reconstructed a full circular
175 genome of a CP Saccharibacteria. Only *Euryarchaeota* and *Thaumarchaeota* were present
176 throughout the water column while most of archaea, including members of the DPANN lineages,
177 resided in the anoxic zone. *Verstrarchaeota*, *Bathyarchaeota*, *Woesearchaeota* and
178 *Micrarchaeota* genomes were recovered for the first time from tropical freshwaters, allowing for
179 future comparative genomics within these groups.
180
181 ***The novel candidate phyla Tanganyikabacteria and Ziwabacteria***
182 Recently, the discovery of non-photosynthetic relatives of *Cyanobacteria* and their
183 metabolic diversity[24,25] has changed our understanding of metabolic repertoire of these
184 organisms. A total of fifteen MAGs (3 dereplicated MAGs) from Tanganyika were distinct from
185 *Cyanobacteria*, and closely related to other non-photosynthetic *Cyanobacteria*-like organisms.
186 We placed these MAGs in the context of published reference genomes[26,27] to ascertain their
187 phylogeny. All 3 MAGs clustered away from *Cyanobacteria*, CP Melainabacteria, CP
188 Blackallbacteria, CP Saganbacteria and CP Margulisbacteria (Figure 4A). The 3 MAGs formed a
189 distinct monophyletic clade closely related to the CP, yet their 16S rRNA sequences shared less
190 than 80% ANI to them. Therefore, we propose that these MAGs belong to a new candidate
191 phylum that we designate as CP Tanganyikabacteria.
192 To assess CP Tanganyikabacteria's global distribution, we searched the 16S rRNA
193 sequences from the CP Tanganyikabacteria MAGs against publicly available datasets. Sequences
194 from (>75% 16S rRNA sequence identity phylum cutoff[28]) CP Tanganyikabacteria were widely
195 distributed across environments, but dominant in freshwater, marine and soil natural systems
196 (Figure 4B).

Organisms from CP Tanganyikabacteria are facultative denitrifiers with heterotrophic metabolism supplemented with the ability to oxidize sulfide (Figure 4C). All CP Tanganyikabacteria in Lake Tanganyika had flagellar machinery and pili, chemotaxis proteins, and several transporters including iron complex outer membrane receptor and transport proteins, ferrous iron, sugar and biopolymer transporters. Like other CP Melainabacteria, they generally do not have a TCA cycle, cytochrome c for the electron transfer chain, nor glycolysis pathways[25]. However, the one denitrifying CP Tanganyikabacteria had a TCA cycle, and cyt c cbb3 type and certain genes in the glycolysis pathway were found. None had Ni-Fe hydrogenases, unlike its CP relatives which have a hydrogen-dependent metabolism[26].

Another group of organisms were distantly related to the Acidobacteria group, CP Raymondbacteria and Candidate division TG3 (Figure S6). Outside of Lake Tanganyika, one single genome was identified for this lineage from metagenomes reconstructed from a deep subsurface aquifer, Crystal Geyser near Green River, Utah[27]. This second CP named Ziwabacteria contains facultative organisms with heterotrophic metabolism supplemented by hydrogen metabolism (FeFe hydrogenase, Ni-Fe Hydrogenase Group 1, Group 3c), oxygen metabolism, and sulfide oxidation (*sqr).*

Nitrogen cycling

Nitrogen is a major nutrient shaping the ecology of Lake Tanganyika[9]. While 477 T of N/year flow into LT, dissolved inorganic nitrogen at the surface waters remains low suggesting rapid nitrogen depletion with important implications for primary productivity that limits yields of critical fisheries[29]. Operating in both oxic and anoxic conditions present in Lake Tanganyika, the microbially-driven nitrogen cycle produces biologically-relevant compounds like ammonia, amino acids, proteins, urea, nitrate, and nitrite. The spatial dynamics of nitrogen forms in Lake Tanganyika have been studied from biogeochemical data and show that upwelling from the chemocline (~50-100m) can deliver fixed N to surface[8,29–31]. Such vertical transfers are critical to supporting productive food webs and the ecosystem[32]. Here, we identify microorganisms that catalyze these processes while emphasizing specific physiologies that may have implications for ecosystem-level nitrogen budgets.

We identified microbes involved in ammonia oxidation, nitrite oxidation, complete ammonia oxidation (comammox), denitrification, urea utilization, anaerobic ammonia oxidation (anammox), dissimilatory nitrate reduction to ammonia (DNRA), and nitrogen fixation (Figure S7). In total, nitrogen cycling organisms accounted for 23% of the microbial community abundance across all the samples. Nitrification (ammonia oxidation to nitrite, nitrite oxidation to nitrate) is catalyzed by oxygen-dependent enzymes (*amoABC, nxrAB*), and expected to be prevalent in the surface. Ammonia oxidation was found in bacteria (*Alpha*-, *Gamma*- and *Betaproteobacteria*), and also in Thaumarchaeota. The potential for nitrite oxidation was found in *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Chloroflexi*, *Deltaproteobacteria, Fibrobacteres*, *Nitrospira*, CP Ziwabacteria*, Planctomycetes,* and *Verrucomicrobia*. Organisms from many CP were found to be capable of nitrite oxidation including Poribacteria,

237 Handelsmanbacteria, Eisenbacteria, Rokubacteria, Hydrogenedentes, Tanganyikabacteria and
238 Ziwabacteria.
239 We recovered two Nitrospira genomes with *amoABC* and *nxrAB* suggesting comammox
240 ability for the first time in a freshwater lake water column (to our knowledge), although previous
241 studies have hinted at their existence[33]. These genomes belong to the previously described
242 lineage II Clade A, consistent with known *Nitrospira* comammox lineages (Figure 5A).
243 Additionally, we performed a full genome-wide metabolic comparison with *Nitrospira*
244 (comammox and not) from other environments (wastewater treatment plants, iron pipes,
245 biofilm)[34](Figure 5B). When comparing comammox, nitrite oxidizing bacteria (NOB) and
246 anammox bacteria, which are competing among themselves for either ammonia or nitrite, we
247 observed that they were present also in the anoxic layer (Figure 5C). Ammonia oxidizers have a
248 versatile metabolism and can be heterotrophs[35]. In anoxic environment, comammox Nitrospira
249 can potentially use H_2, which provides an advantage in hypoxic and anoxic habitats[36].
250 Despite a prominently anoxic water column, our data does not suggest that full
251 denitrification and DNRA are prevalent genomic features of the microbial community.
252 Denitrification and DNRA can co-occur under oxygen-limited or anoxic conditions, such as in
253 the anoxic hypolimnion of lakes. Unlike denitrification which removes reactive nitrogen from
254 the system by converting fully to N_2 gas, DNRA conserves nitrogen within the system by
255 reducing nitrate to ammonium. DNRA therefore does not produce N_2 nor N_2O that is released
256 into the atmosphere. Nitrate reduction to nitrite (*napAB, narGH*) is the first step of denitrification
257 and DNRA. Six MAGs (*Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*,
258 *Planctomycetes* and CP Tanganyikabacteria) had potential for nitrate reduction (Table S4). Only
259 3 organisms from *Betaproteobacteria* and CP Tanganyikabacteria possessed the capacity for
260 complete DNRA. Meanwhile, a single organism from *Betaproteobacteria* was capable of
261 complete denitrification to N_2, while one from CP Tanganyikabacteria could undertake partial
262 denitrification to nitrous oxide (N_2O). Individual steps in denitrification such as nitrite reduction
263 to nitric oxide (NO), NO reduction to N_2O, and N_2O reduction to N_2 were inconsistently
264 distributed in organisms suggesting the presence of widespread metabolic handoffs in the
265 reductive cycle of nitrogen cycle[37] (Table S4).
266 Lake water column anammox was first discovered in Lake Tanganyika[38]. Anammox
267 bypasses the steps of nitrite oxidation and nitrite reduction to directly produce N_2. We identified
268 one MAG closely related to Candidatus *Brocadiacea* in the phylum *Planctomycetes* that
269 contained the hydrazine oxidase (*hzoA*) and hydrazine synthase (*hzsA*) genes necessary for
270 anammox. Like observed previously, anammox organisms peaked in abundance around 100-
271 150m (Figure 5C)[38]. In LT, nitrite oxidizing bacteria (NOB) and anammox, which both compete
272 for nitrite, peak in abundance at 100-150m, but NOB are relatively more abundant than
273 anammox bacteria (Figure 5C). Other important components of the nitrogen cycle in LT include
274 nitrogen fixation and transformations of organic nitrogen including urea, amino acids, and
275 proteins (Table S4).
276

Sulfur Cycling

The sulfur cycle spans the entire water column of LT and is characterized by intermediates that transcend and connect other biogeochemical cycles (Figure S7). Sulfide oxidation to elemental sulfur was identified by the presence of sulfide quinone oxidoreductases (*sqr)* and was found in 94 MAGs, including Cyanobacteria. We compared the distribution of sulfide oxidizing *Cyanobacteria* with non-Cyanobacteria (Figure S4). We noted that non-Cyanobacterial sulfide oxidizers were more prominent in the dark anoxic zone, whereas anoxygenic photosynthetic *Cyanobacteria* that were able to oxidize sulfide were more abundant at the oxic-photic zone. We propose that competition for sulfide results in higher abundance of anoxygenic photosynthetic cyanobacteria at the surface (where light is present), whereas below the euphotic zone, other sulfide oxidizers such as *Chlorobi*, might have an advantage. The process of anoxygenic photosynthesis coupled with sulfide oxidation has been studied in microbial mats[39], but rarely in pelagic freshwater environment. Interestingly, we did not find any purple sulfur bacteria in Lake Tanganyika. Finally, thiosulfate disproportionation (using the sox complex without *soxCD*) and thiosulfate oxidation (using the sox complex) were abundant in LT MAGs representing organisms from *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*, *Chloroflexi*, *Planctomycetes*, and CP Rokubacteria.

Sulfate reduction was found to be highly prominent in the anoxic water column of LT. In total, we identified 28 MAGs representing organisms from six different phyla, namely *Deltaproteobacteria*, *Acidobacteria*, *Nitrospirae*, *Calditrochaeota*, *Aminicenantes*, *Chloroflexi* that possessed the dissimilatory sulfite reductase complex. No organisms capable of only sulfite reduction or containing the anaerobic sulfite reductase complex were identified. Organisms involved in Sulfur metabolism were the third most diverse after carbon and oxygen metabolism (Figure S8).

Carbon transformations as a direct link to the food chain

Autotrophy involving carbon fixation can take many forms. In photosynthesis, CO_2 is fixed into organic carbon and involves Rubisco enzymes. Rubisco Form I and II are found in algae, plants and bacteria. Form IV are found in both Bacteria and Archaea and perform functions distinct from carbon fixation, such as methionine salvage, sulfur metabolism and D-apiode catabolism. In LT, Form I Rubisco were prevalent in MAGs from *Betaproteobacteria*, *Chloroflexi*, *Alphaproteobacteria*, *Cyanobacteria* while Form II was only found in *Betaproteobacteria* (Figure S9*)*. A well supported clade of Rubisco sequences from LT could not be classified to any known forms. This newly discovered Form IV-like clade was composed of sequences from *Planctomycetes*, *Verrucomicrobia*, *Chloroflexi*, *Poribacteria,* CP Handelsmanbacteria, and CP Hydrogenedentes. In addition to the CBB cycle, the Wood-Ljungdahl pathway was also found to be prevalent in LT. This pathway is dominant in anoxic environments and was found in 48 MAGs, representing organisms from *Actinobacteria, Armatimonadetes, Chloroflexi, Deltaproteobacteria, Euryarchaeota, Firmicutes, Lentisphaerae, Nitrospirae, Planctomycetes,* and *Omnitrophica*. Most organisms with the Wood-Ljungdahl

317 pathway decreased in abundance with increased depth. While is it counterintuitive that
318 fermentative organisms are present in the oxic waters of Lake Tanganyika, this could provide an
319 advantage for them under microoxic or anoxic conditions.
320 Aerobic and anaerobic methanotrophs consume methane. While previously thought of as
321 limited to the sediments, it is now known that both aerobic and anaerobic methane oxidation
322 occur in the water column[40,41]. Methane oxidation is prevalent in the well-oxygenated part of the
323 water column in many lakes including Lake Kivu, also part of the African Rift Valley lakes[42,43].
324 Methanogenesis (production of methane) is generally expected to be restricted to the
325 hypolimnion, but can also be found in the oxygenated water column[44]. Understanding the
326 potential for microbially-driven methanotrophy and methanogenesis in the water column is
327 important since it modulates methane fluxes, as they travels from the sediment, through the water
328 column, to the atmosphere. We identified three MAGs (*Verstraetaerchaeota*, *Euryarchaeaota*)
329 that possessed the capacity for methanogenesis and were most abundant in the hypolimnion
330 (Figure S7). The capacity for methane oxidation was identified in organisms from
331 *Betaproteobacteria* whose abundance was highest in the oxic-anoxic interface, suggesting an
332 aerobic methanotrophic lifestyles. Similar to observations in marine systems where capacities for
333 methane oxidation coexist with sulfur oxidation, we propose that these *Betaproteobacteria*
334 adapted to thrive in low-oxygen, transient environments characteristic of the sub-oxic zone, with
335 its ability to utilize many reduced compounds such as sulfide, sulfur, and hydrogen.
336 To investigate the heterotrophic potential of microbes to breakdown polysaccharides in
337 Lake Tanganyika, we annotated the carbohydrate-degrading enzymes (CAZYmes) in all MAGs
338 (Figures S10). Overall, *Planctomycetes* seemed to be a major contributor to carbon degradation
339 since this group had the largest number of protein families associated with auxiliary activities
340 (AAs), carbohydrate-binding modules (CBMs), carbohydrates esterases (CEs), glycoside
341 hydrolases (GHs), and Polysaccharide Lyases (PLs). *Deltaproteobacteria* had the most glycosyl
342 transferases (GTs). Despite their small genome sizes, CPR especially CP Shapirobacteria have
343 relatively many CAZYme hits per genome size[45]. We observed the total number of identified
344 GTs to be more abundant that GHs which begs the question as to why GT are more abundant
345 than GH in this system, and if the same pattern is found in similar or contrasting environments.
346
347 ***Metabolic connections across vertical gradients***
348 Overall, we identified and assessed the distribution of organisms involved in nitrogen,
349 carbon and sulfur cycles through their genomic content. Along a vertical oxygen depth gradient,
350 we observed a prominence of both phototrophic and heterotrophic metabolism in the photic zone
351 (Figure 6, Figure S11). At the sub-oxic zone, we noticed the start of the increase in nitrogen-
352 based metabolism, and complex carbon degradation, and hydrogen metabolism (Figure 6, Figure
353 S11). Methane metabolism was found throughout the water column but increased below the sub-
354 oxic zone suggesting the prominent use of non-oxygen electron acceptors for this process. Our
355 overall findings about microbial metabolisms in LT in relation to light, oxygen and temperature
356 are reminiscent of classical understanding of partitioning between aerobic and anaerobic

357 metabolism. Yet, while it is classically understood that certain metabolic processes such as nitrite
358 oxidation are extremely limited in their distribution (e.g. in *Nitrospirae, Nitrospinae,*
359 *Proteobacteria, Nitrospira*), we found that diverse co-occurring groups are actually capable of
360 performing these transformations. Similarly, identification and characterization of common
361 freshwater lineages such as *Thaumarchaeota* and *Actinobacteria* below the sub-oxic zone in LT
362 presents new insights in the ecology of these organisms. These novel findings demonstrate that
363 genomic-based investigations, when combined with environmental chemistry, add significant
364 value in comparison to taxonomy-based assumptions of microbial metabolism in the
365 environment and highlight potential inter-species metabolic dependencies.
366 Here, we identify connections between organisms in the water column via the
367 intermediate substrates, or end compounds that they produce. In other words, each cycle can be
368 seen not in isolation, but in terms of intersections among cycles. This serves as a conceptual map
369 for studying microbial metabolism in LT and can be used as a framework in other meromictic
370 lakes with strong oxygen and nutrient gradients. The connections between microbial groups and
371 metabolisms vary within and across the different lake layers. Finally, we perform a community
372 wide profiling of microbes with intersecting "metabolic guilds". For example, we found
373 methanotrophs that can denitrify (*Alphaproteobacteria*), chemolithoautotrophs like anammox
374 bacteria (*Planctomycetes)* and nitrifiers (*Nitrospira*), and anoxygenic phototrophs
375 (*Cyanobacteria*) that use hydrogen sulfide which is known to also happen in stable mats
376 environments[39], but unknown of in meromictic lakes. These findings point to the complexity and
377 diverse microbial metabolisms in natural ecosystems and for our understanding of pelagic
378 freshwater microbial paradigms.
379
380 **Discussion**
381 Our study of bacterial and archaeal depth-discrete community offers a whole water
382 column perspective of this incredibly deep, voluminous, and ancient freshwater lake. Our study
383 encompassed the LT's continuous vertical redox gradient. Microbial communities were
384 dominated by from core freshwater taxa at the surface to Archaea and uncultured Candidate
385 Phyla in the latter. We provide genomic evidence for the capabilities of Bacteria and Archaea in
386 biogeochemical cycling and describe links between spatial distribution of organisms and
387 biogeochemical processes. These processes are essential in replenishing the water surface in
388 nutrients, and possibly impact critical food webs that are renowned for their high biodiversity
389 and that serve as important protein sources for local human populations in four surrounding
390 African countries[46].
391 Finally, this study serves as a baseline of microbial community ecology and
392 biogeochemistry of tropical lakes and highlights the importance of characterizing and linking
393 microbial processes in deep anoxic waters. Future follow-up studies including genome informed
394 cultivation efforts, stable isotope probing and biogeochemical modeling, can be employed to
395 validate our conceptual model of transformations in tropical lake waters, and quantitatively

396 model the flow of nutrients from microorganisms to the food web, especially the impacts of the
397 little-studied globally distributed sub-oxic and anoxic freshwaters.
398
399 **Methods**
400 ***Sample collection***
401 Water samples were collected from two stations, Kigoma and Mahale in 2015 (Table 1).
402 Two casts were employed at each station (Kigoma Deep Cast, Kigoma Surface, Mahale Deep
403 Cast, Mahale Offshore). Geographic coordinates of the sample sites are listed in Table S1.
404 Water samples were collected with a vertically oriented Van Dorn bottle in 2015.
405 Environmental data was collected in 2013, on a previous research cruise, using a YSI 6600 sonde
406 with optical DO and chl-a sensors. All data including temperature, dissolved oxygen (DO),
407 conductivity, chlorophyll a) and was collected down to approximately 150m.
408
409 ***DNA extraction and sequencing***
410 DNA extractions were performed using the MP Biomedicals FastDNA Spin Kit with
411 minor protocol modifications as described previously[47]. Metagenome sequencing was conducted
412 at the DOE Joint Genome Institute (JGI). DNA was sequenced on the Illumina HiSeq 2500
413 platform (Illumina, San Diego, CA, U.S.A.), which produces 2 x 150 base pairs (bp) reads with a
414 targeted insert size of ~240 bp. Cells collected at 1200m at the Kigoma station were sent to the
415 DOE JGI for single-cell sequencing.
416 Single amplified genomes (SAGs) were generated following the Department of Energy
417 Joint Genome Institute's (JGI) standard protocol[48]. Briefly, individual cells were sorted using an
418 Influx flow cytometer (BD Biosciences) and treated with Ready-Lyse lysozyme (Epicenter; 5
419 U μL−1 final concentration) for 15 min at room temperature. Next, cell lysis and whole-genome
420 amplification were performed with the REPLI-g Single Cell Kit (Qiagen) in 2 μL reactions.
421 Lysis and stop reagents from the REPLI-g kit received UV treatment to remove potential DNA
422 contamination[49]. An Illumina shotgun library was constructed from each single cell and
423 sequenced on the Illumina NextSeq platform (Illumina, San Diego, CA, U.S.A.). Sequencing
424 reads were filtered using BBTools[50] and assembled into SAGs using SPAdes[51].
425
426 ***Metagenome Assembly and binning***
427 Each of the 24 individual samples were assembled de novo to obtain 24 metagenomes
428 assemblies. Metagenomes were quality filtered, then assembled using MetaSPADEs[52].
429 Sequencing coverage for each assembled scaffold was performed using Bowtie2[53]. Each 24
430 metagenomes were binned individually using three binning program: MetaBat1[54], MetaBat2[55]
431 and MaxBin2[56] test each program's ability to recover bins, and to consolidate bins generated
432 using different binners using DASTool[57]. DAStool was run with default parameters except for --
433 score_threshold 0.4. Using this approach, we reconstructed 3948 MAGs, and calculated genomes
434 statistics using CheckM[58]. We then used dRep[59] to dereplicate these 3948 MAGs to generate a
435 non-redundant set of genomes for downstream analyses using default parameters. dRep identifies

436 the highest quality genomes from a set of similar genomes, and reduces overall redundancy for
437 the downstream analyses. Finally, we used CheckM[58] lineage_wf and ssu_finder functions to
438 generate basic genome statistics and find 16S rRNA sequences for all 802 dereplicated genomes.
439
440 ***Identification of phylogenetic markers***
441 We used a curated set of Hidden Markov Models (HMM) for 16 single-copy ribosomal proteins
442 (rpL2, rpL3, rpL4, rpL5, rpL6, rpL14, rpL14, rpL15, rpL16, rpL18, rpL22, rpL24, rpS3, rpS3,
443 rpS8, rpS10, rpS17, rpS19). Initially, the nucleotides of all genomes were translated into amino
444 acid sequences using Prodigal[60] (V2.6.3). Then the HMM models were searched against the
445 MAGs and SAGs amino acid sequences using hmmsearch (HMMER 3.1b2)[61] with the setting --
446 cut_tc and by saving the alignment as a multiple alignment for all hits (-A option). The cut_tc
447 option was manually determined for each HMM by the threshold resulting in the sharpest decline
448 in score during the initial run (no --cut_tc option); and are hardcoded in the HMM profiles. The
449 esl-reformat.sh script available with hmmsearch was used to extract alignment hits.
450
451 ***Phylogenetic tree***
452 To create the concatenated gene phylogeny, we used publicly available metagenome-
453 assembled genomes which represented a wide range of environments including marine, soil,
454 hydrothermal environments, coastal and estuarine environments. We used 16 genes for the
455 bacterial tree, and 14 genes for the archaeal tree as described previously[22]. The amino acid
456 sequences corresponding to each gene were imported to Geneious Prime V.2019.0.04
457 (https://www.geneious.com) separately for each Bacterial or Archaeal tree. For each gene, we
458 aligned the sequences using MAFFT (v7.388, with parameters: Automatic algorithm,
459 BLOSUM62 scoring matrix, 1.53 gap open penalty, 0.123 offset value) [62]. In the case that more
460 than one copy of the ribosomal protein was found, we performed a sequence alignment of that
461 gene using MAFFT (same settings) and compared the alignments for those copies. For example,
462 if they correspond exactly to a split gene, we concatenated them to obtain a full-length gene. If
463 they were the same section (overlap) of the gene, but one was shorter than the other, the longer
464 fragment was retained.
465 We used Geneious Prime (https://www.geneious.com) to apply a 50% gap masking and
466 concatenated the 16 (or 14) proteins. The concatenated alignment was exported into the fasta
467 format and used as an input for RAxML-HPC, which was run on the CIPRES server[63] with the
468 following settings: datatype = protein, maximum-likelihood search = TRUE, no bfgs = FALSE,
469 print br length = false, protein matrix spec: JTT, runtime=168 hours, use bootstrapping = TRUE.
470 The resulting Newick format tree was visualized with FigTree
471 (http://tree.bio.ed.ac.uk/software/figtree/). The Newick format trees for the Archaea and
472 Bacteria are available as Supplementary Data 1 and Supplementary Data 2. The concatenated
473 ribosomal protein alignments can be found in Supplementary Data 3 and 4.
474 The same procedure was followed to create the taxon-specific tree, for example for the
475 CP Tanganyikabacteria, CP Ziwabacteria, Actinobacteria ac groups, LD12, and Nitrospira, using

476 phylum-specific references, except with no gap masking since sequences were highly similar and
477 had few gaps. The *amoA* gene phylogeny was performed using UniProt *amoA* sequences with the
478 two comammox genomes and with 90% applied masking.
479
480 ***Taxonomic assignment and comparison of manual versus automated methods***
481 Taxonomic classification of MAGs and SAGs was manually performed by careful
482 inspection of the concatenated-RP gene phylogeny, bootstrap values of each group, and closest
483 named representatives. For comparative purposes, we also assigned taxonomy using GTDB-tk[10],
484 which uses ANI comparisons to reference genomes and 120 marker genes. While most
485 taxonomic assignment was consistent between methods, we noted that automatic assignment was
486 inconsistent for phyla with few representatives in the databases, such Archaea or CP. In total, 12
487 of the 48 SAGs had enough marker ribosomal sequences to be included in the bacterial tree, and
488 taxonomic classification was manually assigned, confirming classification by JGI-IMG. In most
489 cases, we were able to resolve a finer taxonomic identity for the SAGs manually. Additionally,
490 16S rRNA genes were found in 444 out of 802 MAGs using CheckM[58]. We used the curated
491 freshwater bacteria database "FreshTrain", implemented in TaxAss[64], to assigned taxonomy to
492 the 16S rRNA sequences. This freshwater-specific database, albeit focused on epilimnions of
493 temperate lakes is useful for comparable terminology between the Lake Tanganyika genomes
494 and the "typical" freshwater bacterial clades, lineages and tribes terminology defined
495 previously[11]. Taken together, the concatenated gene phylogeny, manual curation, automated
496 taxonomic classification and 16S rRNA sequences were used to provide support for phylogenetic
497 classification.
498
499 ***Confirmation and support for novel candidate phyla***
500 To ensure that candidate phyla (CP Tanganyikabacteria and CP Ziwabacteria) described
501 in this study are indeed novel, we combined evidence from genome-wide phylogenetic analysis,
502 16S rRNA sequences and phylogeny of the MAGs, and average nucleotide identity (ANI) with
503 closely related genomes in the literature. Taken together, this providence evidence of the
504 monophyly and deep-branching of these lineages suggesting that they represent novel bacterial
505 phyla.
506 To visualize the distribution of CP Tanganyikabacteria globally, we annotated the 16S
507 rRNA sequence using ssu_finder from CheckM. Then we searched the LT's 16S rRNA sequence
508 against all the 16S rRNA sequence database from IMG/M[65] (accessed June 2019). We extracted
509 the latitude and longitude of each hit and used R to visualize the results. The points shown are
510 results with 81-100% percent ID using BLASTn[66] and have a e-value between 0 and 0.0000653.
511
512 ***Relative abundance across water column depths***
513 To obtain a matrix of relative coverage as a proxy for abundance across the samples, we
514 first mapped each metagenomic paired-read sets to its respective metagenome assemblies using
515 BBMAP[50], using default settings. BBMap uses pileup.sh, which normalizes the coverage per

516 scaffold and genome size. We combined the mapping table from the 24 metagenomes, and
517 summed the total coverage based on the associated MAG ID for each scaffold. To normalize the
518 coverage by the metagenome size, we obtained the number of raw reads per metagenome. We
519 also classified the metagenomes by depth as oxic (0-50m, 100% to 80% DO), sub-oxic (50-
520 100m, 80% to 0% DO) and anoxic (100-1200m, 0% DO). We divided each coverage value by
521 the sum of reads in that layer. To determine whether a taxonomic group was "abundant" we
522 applied the criterion that the lineage must be among the 20 most abundant taxa in each of the 24
523 metagenomes
524
525 ***Metabolic potential analysis***
526 Metabolic potential of the Tanganyika MAGs was assessed using 143 custom HMM
527 profiles[67] , made using *hmm-build* and searched with *hmmsearch* (HMMER 3.1b2)[61]. Threshold
528 cutoffs were manually created for each 143 HMMs by identifying the score at which the sharpest
529 drop happened and manually writing the TC value in the HMM database file. Hmmsearch --
530 use_tc and esl-reformat (part of the "easel" tool suite that comes with hmmsearch) were used to
531 export the alignments for the HMM hits. We classified the number of genes involved in sulfur
532 metabolism, hydrogen metabolism, methane, nitrogen, oxygen, C1-metabolism, C monoxide
533 (CO) metabolism, carbon fixation, organic nitrogen (urea), halogenated compounds, arsenic,
534 selenium, nitriles, and metals metabolism/utilization. To determine if a MAG could perform a
535 metabolic function, one copy of each representative gene of the pathway must have been present,
536 which a value of 1 (presence) was written, as opposed to 0 (absence) (Supplementary Table 3).
537 For novel candidate phyla, we also used kofamKOALA[68] to further annotate the genomes
538 (Supplementary Table 5). To investigate heterotrophy, carbohydrates degrading enzymes were
539 annotated using dbCan2[69] (Supplementary Table 6). The dbCAN-HMMdb-V7 databases was
540 downloaded (June 2019), and hmmscan and hmmscan-parser.sh were used to run dbCAN2.
541
542 ***Annotation and differentiation of amoA and pmoA genes***
543 We annotated *pmoA* using custom HMMs as described above. However, we did not have
544 a custom HMM for *amoA*. *amoA* and *pmoA* can be difficult to distinguish using gene calling
545 methods. Therefore, we selected sequences annotated as "*pmoA*" (amoA-like sequences) and
546 built a single-gene phylogeny using references in Supplementary File 3 from [70]. We first called
547 the protein coding sequences using Prodigal[60] on the File 1 named
548 "AamoA.db_an96.aln_tax.annotated.fasta". Then we aligned the sequences using MAFFT[62] in
549 Geneious Prime, and applied 90% gap masking. We used RAxML[71] to build a comprehensive
550 tree. Based on the phylogenetic position of the sequences, we delineated *amoA* and *pmoA*
551 sequences (Figure S12). We noted that the pmoA custom HMM did not pick up any Archaeal
552 sequences. To identify archaeal ammonia oxidizers in LT, we searched the amoA, amoB and
553 amoC subunits of *Candidatus Nitrosophaera gargensis* Ga9.2 (Phylum Thaumarchaeota,
554 Kingdom Crenarchaeota, Class Nitrososphaeria) using BLASTp 2.2.31[66] . We identified 5

555 *Thaumarchaeota* MAGs with either subunits, and 2 MAGs (with all 3 subunits). We added the
556 amoA sequences from these *Thaumarchaeota* to the tree in Figure S12.
557
558 ***Annotation and classification of Rubisco sequences***
559 Following the methodology described in reference [72], we further classified the Rubisco
560 enzymes in Lake Tanganyika MAGs. We used manually curated HMMs (described in our
561 Methods) to retrieve Form I, II, II/III, III and IV. We only retained sequences over 200 basepairs.
562 To generate the concatenated tree, we retrieved the unmasked sequences from Jaffe et al., 2019,
563 and used MAFFT to include our sequences from Lake Tanganyika. A 95% sequence alignment
564 masking was described as mentioned in the methods in Jaffe et al., 2019. We used RAxML[71] on
565 the CIPRES server[63] to generate the concatenated tree. Rubisco forms were classified based on
566 their position in the phylogenetic tree.
567
568 **Data availability**
569 The 802 MAGs can be accessed under the NCBI BioProject ID: PRJNA523022. The 24
570 metagenomes can be accessed on the Integrated Microbial Genomes & Microbiomes IMG/M
571 portal using the following IMG Genome ID's: 3300020220, 3300020083, 3300020183,
572 3300020200, 3300021376, 3300021093, 3300021091, 3300020109, 3300020074, 3300021092,
573 3300021424, 3300020179, 3300020193, 3300020204, 3300020221, 3300020196, 3300020190,
574 3300020197, 3300020222, 3300020214, 3300020084, 3300020198, 3300020603, 3300020578.
575 An interactive version of the Archaeal and Bacteria trees (Supplementary Data 1 and 2) can be
576 accessed at iTOL at the url: https://itol.embl.de/shared/patriciatran . Code to generate the figures
577 can be accessed at https://github.com/patriciatran/LakeTanganyika/.
578
579 **Acknowledgments**
580 We thank the University of Wisconsin - Office of the Vice Chancellor for Research and
581 Graduate Education, University of Wisconsin – Department of Bacteriology, and University of
582 Wisconsin – College of Agriculture and Life Sciences for their support. The Lake Tanganyika
583 project was supposed by the United States National Science Foundation (DEB-1030242 to
584 P.B.M and DEB-0842253 to Y.V.). We are thankful to the Tanzania Commission for Science
585 and Technology (COSTECH) for providing the research permits to collect the samples. This
586 research was also supported by the U.S. Department of Energy Joint Genome Institute (JGI)
587 through a JGI-Community Science Program Award to K.D.M (Proposal ID: CSP 2796). The
588 work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of
589 Science User Facility, is supported by the Office of Science of the U.S. Department of Energy
590 under Contract No. DE-AC02-05CH11231. K.D.M. received funding from the United States
591 National Science Foundation via an INSPIRE award (DEB-1344254) and the Wisconsin Alumni
592 Research Foundation at UW-Madison. B.M.K was supported by funding from the Leibniz
593 Institute for Freshwater Ecology and Inland Fisheries' International Postdoctoral Research
594 Fellowship and from the German Research Foundation through the LimnoScenES project (AD

91/22-1). P.Q.T is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Alexander Graham Bell Fellowship Canadian Graduate Scholarship – Doctoral (CGS-D). We would like to thank our colleagues for valuable feedback throughout the project.

Contributions

P.Q.T, K.D.M, and K.A designed the study. P.B.M and B.K. conducted the sampling and provided environmental data. P.Q.T. and K.A. performed genome-resolved metagenomics analyses. P.Q.T. performed phylogenetic, metabolic and community analyses. P.Q.T, K.D.M, and K.A. wrote the manuscript. All authors provided feedback and revised the manuscript.

Corresponding author
Correspondence to Karthik Anantharaman

Competing interests
The authors declare no competing financial interests.

Supplementary Information
Supplementary Methods, Supplementary Figures 1–12, Supplementary Tables 1-7 and Supplementary Data 1-4

615 **List of Figures**
616
617 **Figure 1. Sampling site and environmental data**
618 *A.* Sampling sites in Lake Tanganyika, collected in July and October 2015
619 *B.* Weekly temperature profiles from January to July 2013.
620 *C.* Weekly dissolved oxygen profiles from January to July 2013.
621 Contours in *Figure 1A* and *1B* were generated *interp* in the package Akima (Akima and
622 Gebhardt 2016), and plots were done with the *fields* package (Nychka et al., 2017) in R.
623 Thermocline depths are calculated from the rLakeAnalyzer (Winslow et al., 2019) package in R,
624 and are shown on the plots as points. Each point represents a calculated thermocline depth.
625 *Figure 1D.* Nitrate profiles (ug/L) in Mahale and Kigoma collected in 2015.
626
627 **Figure 2. Phylogeny of Archaea and Bacteria**
628 Concatenated gene phylogeny showing the distribution of (*Figure 2A*) Archaeal and (*Figure 2B*)
629 Bacterial metagenome-assembled genomes (MAGs) and single-cell amplified genomes (SAGs)
630 in Lake Tanganyika, using 14 and 16 concatenated ribosomal proteins respectively, and
631 visualized using FigTree. The scale bar shows branch length corresponding to 0.3 and 0.4
632 substitutions per site.
633 *A*. Colored names are a subset of groups from Lake Tanganyika, with the number dereplicated of
634 MAGs listed in parenthesis.
635 *B.* Colored groups are the most abundant groups in Lake Tanganyika. Additionally, the
636 Candidate Phyla Radiation (CPR) and candidate phyla (CP) are listed. The two novel phyla from
637 this study are italicized. In both panels, the color scale of the circles represents the normalized
638 abundance of that taxa across depths, summarized by the values in the oxic (0-50m), sub-oxic
639 (50-100m) and anoxic (100-12000m) zones.
640
641 **Figure 3. Rank abundance curve of major taxonomic groups.** The total sum of coverage was
642 calculated for each taxonomic group, then the top 20 most abundant groups overall were selected
643 to be plotted here. The y axis represents the relative abundance (sum of abundance of specific
644 taxonomic group in a given layer divided by total sum of abundance across all taxa in that given
645 layer). Alphaproteobacteria and Actinobacteria are shown as classes, lineages, or tribes for
646 freshwater microbiologists/ecologists to relate to their ecosystem.
647
648 **Figure 4. Novel Candidate Phyla Tanganyikabacteria**
649 *A.* Genome-wide phylogeny of the 3 CP Tanganyikabacteria MAGs from Lake Tanganyika in
650 the context of Cyanobacteria and the sister lineages Melainabacteria, Saganbacteria,
651 Margulisbacteria and Blackallbacteria.
652 *B.* Global distribution of the 16S rRNA gene from CP Tanganyikabacteria.
653 *C.* Cellular map overview of genes common to all CP Tanganyikabacteria, with some common
654 features of CP Tanganyikabacteria.

655
656 **Figure 5. Comammox *Nitrospira* Genome Comparison and Phylogeny.**
657 *A.* Concatenated gene phylogeny of 16 ribosomal proteins of *Nitrospira* genomes from Lake
658 Tanganyika (renamed LTC-1 and LTC-2) and other environments, including comammox.
659 *B.* Presence (filled circles) and absence (empty circles) of selected genes in ammonia and nitrite
660 oxidation to nitrate, comammox, ammonium transporters and urea utilization. *amoABC* were
661 annotated using single-gene phylogenies, ammonium transporters and ureases were annotated
662 with Interpro, whereas *nxrAB* were annotated with custom HMM.
663 *C.* Distribution of MAGs identified as nitrite oxidizing bacteria (NOB), comammox and
664 anammox bacteria in Lake Tanganyika. The grey dashed lines at 50 and 100m represent the
665 oxygen layer boundaries. The red line (70m) represents the depth of the photic zone.
666
667 **Figure 6. Overview of vertical biogeochemical cycling**
668 Samples were collected along a vertical profile to capture changes in environmental variables.
669 The lake is separated into three operational zones: oxic, sub-oxic and anoxic. A selected group of
670 cycles (e.g. methane, fermentation, nitrogen, sulfur) are shown with the approximate vertical
671 distribution are depicted on the right.
672
673 **List of Supplementary Tables**
674
675 **Supplementary Table 1. List of metagenomes**
676 List of the 24 metagenome samples collected in Lake Tanganyika with accession ID's.
677
678 **Supplementary Table 2. Metagenome-assembled genomes information**
679 Information for the 802 Metagenome-assembled genomes (MAGs) from Lake Tanganyika,
680 including taxonomic assignment, genome statistics (completeness, size, number of coding DNA
681 sequences), and more.
682
683 **Supplementary Table 3. Relative abundances of MAGs**
684 Relative abundances of each 802 MAGs through depths and strata. The relative abundances
685 (Columns D to AA) were obtained by first dividing each coverage value obtained by BBMap
686 pileup.sh by the number of total reads in that given metagenome. Then the value was multiplied
687 by the average number of reads of the 24 metagenomes. Column "AB","AC", and "AB", are the
688 values obtained divided by the maximum number of reads of all given metagenomes per layer
689 (oxic, sub-oxic and anoxic).
690
691 **Supplementary Table 4. HMM metabolic markers in MAGs**
692 The number of hits for each curated HMM of selected metabolic marker genes, in all 802 MAGs.
693

694 **Supplementary Table 5. CP Tanganyikabacteria and CP Ziwabacteria KOfam KEGG**
695 **annotations**
696 Six genomes belonging to CP Tanganyikabacteria and CP Ziwabacteria were further annotated
697 using KOFAM KEGG via the online website.
698
699 **Supplementary Table 6. The 802 genomes were annotated using dbCAN.**
700 The genome sizes and percent coding DNA sequences (CDS) were obtained by CheckM. The
701 number of CAZYme hits were normalized by the genome size of each MAG (Column "E"). The
702 sum, and the detailed hits for Auxiliary Activities (AA), Carbohydrate-binding domains (CBD),
703 cohesin, glycoside hydrolases (GH), glycosyl transferases (GT) and Polysaccharide lyases (PL)
704 are reported.
705
706 **Supplementary Table 7.** All >3000 MAGs generated, and dereplication results. Each MAG was
707 assigned to a cluster, then one representative MAG was selected for each cluster.
708
709 **List of Supplementary Figures**
710 **Figure S1.** Environmental data collected in 2010, 2011, 2012 and 2013 for dissolved oxygen,
711 temperature, chlorophyll a and conductivity. Data in was collected during 6 days in 2010 [Julian
712 days 189 to 230], 8 days in 2011 [Julian day 182 to 237], 26 days in 2012 [Julian day 194 to 362]
713 and 33 days in 2013 [Julian day 3 to 227]. Data in 2013 covers the largest day span annually.
714 Data was plotted with ggplot2 (Wickham 2016) in R, dates were converted using *lubridate*
715 package (Grolemund and Wickham 2011).
716
717 **Figure S2. Light information**
718 *A*. Box plots showing the Secchi depths distribution by month, for data collected in 2012 and
719 2013. Secchi depths is a visual measure of lake transparency. The coefficient of light extinction
720 (kd) is estimated from the empirical relationship 1.7/SecchiDepth(m) = kd. The water is clearest
721 in April. The 24 metagenomes from our study were collected in July and October. These months
722 are shown in red on the x-axis.
723 *B*. Boxplot showing the range of light extinction coefficient. The coefficient of light extinction
724 (*kd*) is estimated from the empirically derived relationship 1.7/SecchiDepth(meters) = *kd*. The
725 values are derived from the Secchi depths values *A.* The lower the kd value, the clearer the lake,
726 and the higher the kd value, the less transparent the lake. The 24 metagenomes were collected in
727 July and October, as shown in red on the x-axis.
728
729 **Figure S3.** Index of replication (iRep) values for three high-quality Cyanobacterial MAGs from
730 Lake Tanganyika, along with their abundance (coverage) values in the 24 metagenomes (color
731 coded in pink and blue for Kigoma and Mahale). Dashed grey lines at 50 and 100m represent
732 oxygen layer boundaries, and red 70m dashed line represents approximate photic zone depth
733 based in previously published environmental profiles. Note that the most abundant

Cyanobacteria MAG did not meet the thresholds to calculate iRep values, therefore we were unable to assess its replication value.

Figure S4. A. Distribution of all Cyanobacterial MAGs in Lakes Tanganyika. In blue are Cyanobacterial MAGs which are sulfide oxidizers, based on *sqr* presence in their genomes. In Pannel B, the zoomed-in distribution on Cyanobacterial sulfide oxidizers, and all other sulfide oxidizers in Lake Tanganyika. The x-axis the now the sum of coverage group by the taxonomic group in the legend. The dominating sulfide oxidizers in the oxic layer (grey dashed lines) are Cyanobacteria, whereas a wider diversity of organisms, including Chlorobi, reach higher abundance in the anoxic layer. Pannel C. Conceptual diagram hypothesing that competition for sulfide results in higher abundance of anoxygenic photosynthetic cyanobacteria at the surface (where light is present), whereas below the euphotic zone, other sulfide oxidizers might have an advantage. Note that purple sulfur bacteria (Gammaproteobacteria, Order Chromatiales) were not identified in Lake Tanganyika.

Figure S5. Detailed Actinobacteria genome phylogeny with a focus on common freshwater lineages, using MAGs from Lake Tanganyika, published datasets (see text).

Figure S6. Zoom into the CP Ziwabacteria. See Newick tree (Supplementary Data 2) for detailed phylogeny.

Figure S7. Overview of the biogeochemical cycles using the data show in Table S3. Overall, the taxa involved in nitrogen, sulfur, hydrogen, oxygen, carbon and other (e.g. metals) metabolism are shown.

Figure S8. As a supplement to Figure S7, a bar plot shows the number of distinct taxa involved in each biogeochemical category, versus the number of distinct MAGs corresponding to these categories.

Figure S9. Phylogeny of Rubisco genes in Lake Tanganyika, using the sequences in[72] as the backbone reference tree.

Figure S10. Distribution of carbohydrate-degrading enzymes (CAZymes) among the Lake Tanganyika MAGs, organized alphabetically. The CAZyme density is the total number of Cazyme hits divided by the respective MAG's genome size.

Figure 11. Distribution of organisms involved in the various metabolic cycles from Figure S7 and S8.

773 **Figure S12.** Single-gene phylogeny of bacterial and archaeal *amoA* genes (See Methods). Lake
774 Tanganyika genomes are bolded.
775
776 **List of Supplementary Data**
777
778 **Supplementary Data 1.** Archaeal tree in Newick format.
779
780 **Supplementary Data 2.** Bacterial tree in Newick format.
781
782 **Supplementary Data 3.** Concatenated alignment of 14 ribosomal proteins for Archaea.
783
784 **Supplementary Data 4.** Concatenated alignment of 16 ribosomal proteins for Bacteria
785
786

References

1. Verburga, P. & Hecky, R. E. The physics of the warming of Lake Tanganyika by climate change. *Limnol. Oceanogr.* **54**, 2418–2430 (2009).
2. Alin, S. R. & Johnson, T. C. Carbon cycling in large lakes of the world: A synthesis of production, burial, and lake-atmosphere exchange estimates: CARBON CYCLING IN LARGE LAKES. *Glob. Biogeochem. Cycles* **21**, n/a-n/a (2007).
3. Durisch-Kaiser, E. *et al.* What prevents outgassing of methane to the atmosphere in Lake Tanganyika? *J. Geophys. Res.* **116**, (2011).
4. Cohen, A. S. *et al.* Climate warming reduces fish production and benthic habitat in Lake Tanganyika, one of the most biodiverse freshwater ecosystems. *Proc. Natl. Acad. Sci.* **113**, 9563–9568 (2016).
5. Salzburger, W., Van Bocxlaer, B. & Cohen, A. S. Ecology and Evolution of the African Great Lakes and Their Faunas. *Annu. Rev. Ecol. Evol. Syst.* **45**, 519–545 (2014).
6. Pirlot, S., Unrein, F., Descy, J.-P. & Servais, P. Fate of heterotrophic bacteria in Lake Tanganyika (East Africa): Fate of bacteria in Lake Tanganyika. *FEMS Microbiol. Ecol.* **62**, 354–364 (2007).
7. De Wever, A. *et al.* Bacterial Community Composition in Lake Tanganyika : Vertical and Horizontal Heterogeneity. *Society* **71**, 5029–5037 (2005).
8. Edmond, J. M. *et al.* Nutrient chemistry of the water column of Lake Tanganyika. *Limnol. Oceanogr.* **38**, 725–738 (1993).
9. Järvinen, M., Salonen, K., Sarvala, J., Vuorio, K. & Virtanen, A. The stoichiometry of particulate nutrients in Lake Tanganyika – implications for nutrient limitation of phytoplankton. *Hydrobiologia* **407**, 81–88 (1999).
10. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
11. Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D. & Bertilsson, S. *A guide to the natural history of freshwater lake bacteria.* vol. 75 (2011).
12. Stal, L. J. & Moezelaar, R. Fermentation in cyanobacteria. *FEMS Microbiol. Rev.* **21**, 179–211 (1997).
13. Brown, C. T., Olm, M. R., Thomas, B. C. & Banfield, J. F. Measurement of bacterial replication rates in microbial communities. *Nat. Biotechnol.* **34**, 1256–1263 (2016).
14. Tonolla, M., Peduzzi, S., Demarta, A., Peduzzi, R. & Hahn, D. Phototropic sulfur and sulfate-reducing bacteria in the chemocline of meromictic Lake Cadagno, Switzerland. *J. Limnol.* **63**, 161 (2004).
15. Bendall, M. L. *et al.* Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *ISME J.* **10**, 1589–1601 (2016).
16. Cabello-Yeves, P. J. *et al.* Genomes of novel microbial lineages assembled from the sub-ice waters of Lake Baikal. *Appl. Environ. Microbiol.* AEM.02132-17 (2017) doi:10.1128/AEM.02132-17.
17. Tran, P. *et al.* Microbial life under ice: Metagenome diversity and *in situ* activity of Verrucomicrobia in seasonally ice-covered lakes. *Environ. Microbiol.* **20**, 2568–2584 (2018).
18. Linz, A. M., He, S., Stevens, S. L. R., Anantharaman, K. & Robin, R. Connections between freshwater carbon and nutrient cycles revealed through. **221**, (2018).
19. Haas, S., Desai, D. K., LaRoche, J., Pawlowicz, R. & Wallace, D. Geomicrobiology of the Carbon, Nitrogen and Sulfur Cycles in Powell Lake: A Permanently Stratified Water

832 Column Containing Ancient Seawater. *Environ. Microbiol.* (2019) doi:10.1111/1462-
833 2920.14743.
834 20. Neuenschwander, S. M., Ghai, R., Pernthaler, J. & Salcher, M. M. Microdiversification in
835 genome-streamlined ubiquitous freshwater Actinobacteria. *ISME J.* 1–14 (2017)
836 doi:10.1038/ismej.2017.156.
837 21. Kang, I., Kim, S., Islam, Md. R. & Cho, J.-C. The first complete genome sequences of the
838 acI lineage, the most abundant freshwater Actinobacteria, obtained by whole-genome-
839 amplification of dilution-to-extinction cultures. *Sci. Rep.* **7**, (2017).
840 22. Hug, L. A. *et al.* A new view of the tree of life. *Nat. Microbiol.* **1**, 1–6 (2016).
841 23. Peura, S. *et al.* Distinct and diverse anaerobic bacterial communities in boreal lakes
842 dominated by candidate division OD1. *ISME J.* **6**, 1640–1652 (2012).
843 24. Soo, R. M. *et al.* An Expanded Genomic Representation of the Phylum Cyanobacteria.
844 *Genome Biol. Evol.* **6**, 1031–1045 (2014).
845 25. Di Rienzi, S. C. *et al.* The human gut and groundwater harbor non-photosynthetic bacteria
846 belonging to a new candidate phylum sibling to Cyanobacteria. *eLife* **2**, (2013).
847 26. Matheus Carnevali, P. B. *et al.* Hydrogen-based metabolism as an ancestral trait in lineages
848 sibling to the Cyanobacteria. *Nat. Commun.* **10**, (2019).
849 27. Probst, A. J. *et al.* Differential depth distribution of microbial function and putative
850 symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat.*
851 *Microbiol.* **3**, 328–336 (2018).
852 28. Yarza, P. *et al.* Uniting the classification of cultured and uncultured bacteria and archaea
853 using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* **12**, 635–645 (2014).
854 29. Brion, N. *et al.* Inorganic Nitrogen Uptake and River Inputs in Northern Lake Tanganyika. *J.*
855 *Gt. Lakes Res.* **32**, 553–564 (2006).
856 30. Kimbadi, S., Vandelannoote, A., Deelstra, H., Mbemba, M. & Ollevier, F. Chemical
857 composition of the small rivers of the north-western part of Lake Tanganyika. in *From*
858 *Limnology to Fisheries: Lake Tanganyika and Other Large Lakes* (eds. Lindqvist, O. V.,
859 Mölsä, H., Salonen, K. & Sarvala, J.) 75–80 (Springer Netherlands, 1999). doi:10.1007/978-
860 94-017-1622-2_7.
861 31. Verschuren, D. The heat on Lake Tanganyika. *Nature* **424**, 731–732 (2003).
862 32. Kilham, P. & Kilham, S. S. OPINION Endless summer: internal loading processes dominate
863 nutrient cycling in tropical lakes. *Freshw. Biol.* **23**, 379–389 (1990).
864 33. Alfreider, A. *et al.* Autotrophic carbon fixation strategies used by nitrifying prokaryotes in
865 freshwater lakes. *FEMS Microbiol. Ecol.* **94**, (2018).
866 34. Poghosyan, L. *et al.* Metagenomic recovery of two distinct comammox *Nitrospira* from the
867 terrestrial subsurface. *Environ. Microbiol.* (2019) doi:10.1111/1462-2920.14691.
868 35. Daims, H., Lücker, S. & Wagner, M. A New Perspective on Microbes Formerly Known as
869 Nitrite-Oxidizing Bacteria. *Trends Microbiol.* **24**, 699–712 (2016).
870 36. Koch, H., van Kessel, M. A. H. J. & Lücker, S. Complete nitrification: insights into the
871 ecophysiology of comammox Nitrospira. *Appl. Microbiol. Biotechnol.* **103**, 177–189 (2019).
872 37. Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected
873 biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 13219 (2016).
874 38. Schubert, C. J. *et al.* Anaerobic ammonium oxidation in a tropical freshwater system (Lake
875 Tanganyika). *Environ. Microbiol.* **8**, 1857–1863 (2006).

39. Cohen, Y., Jørgensen, B. B., Revsbech, N. P. & Poplawski, R. Adaptation to Hydrogen Sulfide of Oxygenic and Anoxygenic Photosynthesis among Cyanobacteria. *Appl. Environ. Microbiol.* **51**, 398–407 (1986).
40. Eller, G., Kanel, L. & Kruger, M. Cooccurrence of Aerobic and Anaerobic Methane Oxidation in the Water Column of Lake Plu see. *Appl. Environ. Microbiol.* **71**, 8925–8928 (2005).
41. Thottathil, S. D., Reis, P. C. J. & Prairie, Y. T. Methane oxidation kinetics in northern freshwater lakes. *Biogeochemistry* **143**, 105–116 (2019).
42. Zigah, P. K. *et al.* Methane oxidation pathways and associated methanotrophic communities in the water column of a tropical lake: Lake Kivu methane oxidation pathways. *Limnol. Oceanogr.* **60**, 553–572 (2015).
43. Roland, F. A. E. *et al.* Denitrification, anaerobic ammonium oxidation, and dissimilatory nitrate reduction to ammonium in an East African Great Lake (Lake Kivu): *Denitrification* , *anammox* , *and DNRA in Lake Kivu*. *Limnol. Oceanogr.* **63**, 687–701 (2018).
44. Grossart, H.-P., Frindte, K., Dziallas, C., Eckert, W. & Tang, K. W. Microbial methane production in oxygenated water column of an oligotrophic lake. *Proc. Natl. Acad. Sci.* **108**, 19657–19661 (2011).
45. Danczak, R. E. *et al.* Members of the Candidate Phyla Radiation are functionally differentiated by carbon- and nitrogen-cycling capabilities. *Microbiome* **5**, (2017).
46. Knaap, M. V. der, Katonda, K. I. & Graaf, G. J. D. Lake Tanganyika fisheries frame survey analysis: Assessment of the options for management of the fisheries of Lake Tanganyika. *Aquat. Ecosyst. Health Manag.* **17**, 4–13 (2014).
47. Shade, A. *et al.* Interannual dynamics and phenology of bacterial communities in a eutrophic lake. *Limnol. Oceanogr.* **52**, 487–494 (2007).
48. Obtaining genomes from uncultivated environmental microorganisms using FACS–based single-cell genomics | Nature Protocols. https://www.nature.com/articles/nprot.2014.067.
49. Woyke, T. *et al.* Decontamination of MDA Reagents for Single Cell Whole Genome Amplification. *PLoS ONE* **6**, e26161 (2011).
50. Bushnell, B. BBMAP. https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbmap-guide/.
51. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
52. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
53. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
54. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
55. Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
56. Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinforma. Oxf. Engl.* **32**, 605–607 (2016).
57. Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* **3**, 836–843 (2018).

58. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–55 (2015).
59. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* 1–5 (2017) doi:10.1038/ismej.2017.126.
60. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, (2010).
61. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, (2011).
62. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
63. Miller, M. A., Pfeiffer, W. & Schwartz, Terri. Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. in *Proceedings of the Gateway Computing Environments Workshop* 1–8 (2010).
64. Rohwer, R. R., Hamilton, J. J., Newton, R. J. & McMahon, K. D. TaxAss: Leveraging a Custom Freshwater Database Achieves Fine-Scale Taxonomic Resolution. *mSphere* **3**, (2018).
65. Chen, I.-M. A. *et al.* IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* **47**, D666–D677 (2019).
66. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
67. Zhou, Z., Tran, P., Liu, Y., Kieft, K. & Anantharaman, K. METABOLIC: A scalable high-throughput metabolic and biogeochemical functional trait profiler based on microbial genomes. *bioRxiv* 761643 (2019) doi:10.1101/761643.
68. Aramaki, T. *et al.* KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *bioRxiv* (2019) doi:10.1101/602110.
69. Zhang, H. *et al.* dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**, W95–W101 (2018).
70. Alves, R. J. E., Minh, B. Q., Urich, T., von Haeseler, A. & Schleper, C. Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on amoA genes. *Nat. Commun.* **9**, 1517 (2018).
71. Liu, K., Linder, C. R. & Warnow, T. RAxML and FastTree: Comparing Two Methods for Large-Scale Maximum Likelihood Phylogeny Estimation. *PLoS ONE* **6**, e27731 (2011).
72. Jaffe, A. L., Castelle, C. J., Dupont, C. L. & Banfield, J. F. Lateral Gene Transfer Shapes the Distribution of RuBisCO among Candidate Phyla Radiation Bacteria and DPANN Archaea. *Mol. Biol. Evol.* **36**, 435–446 (2019).

Figure 1

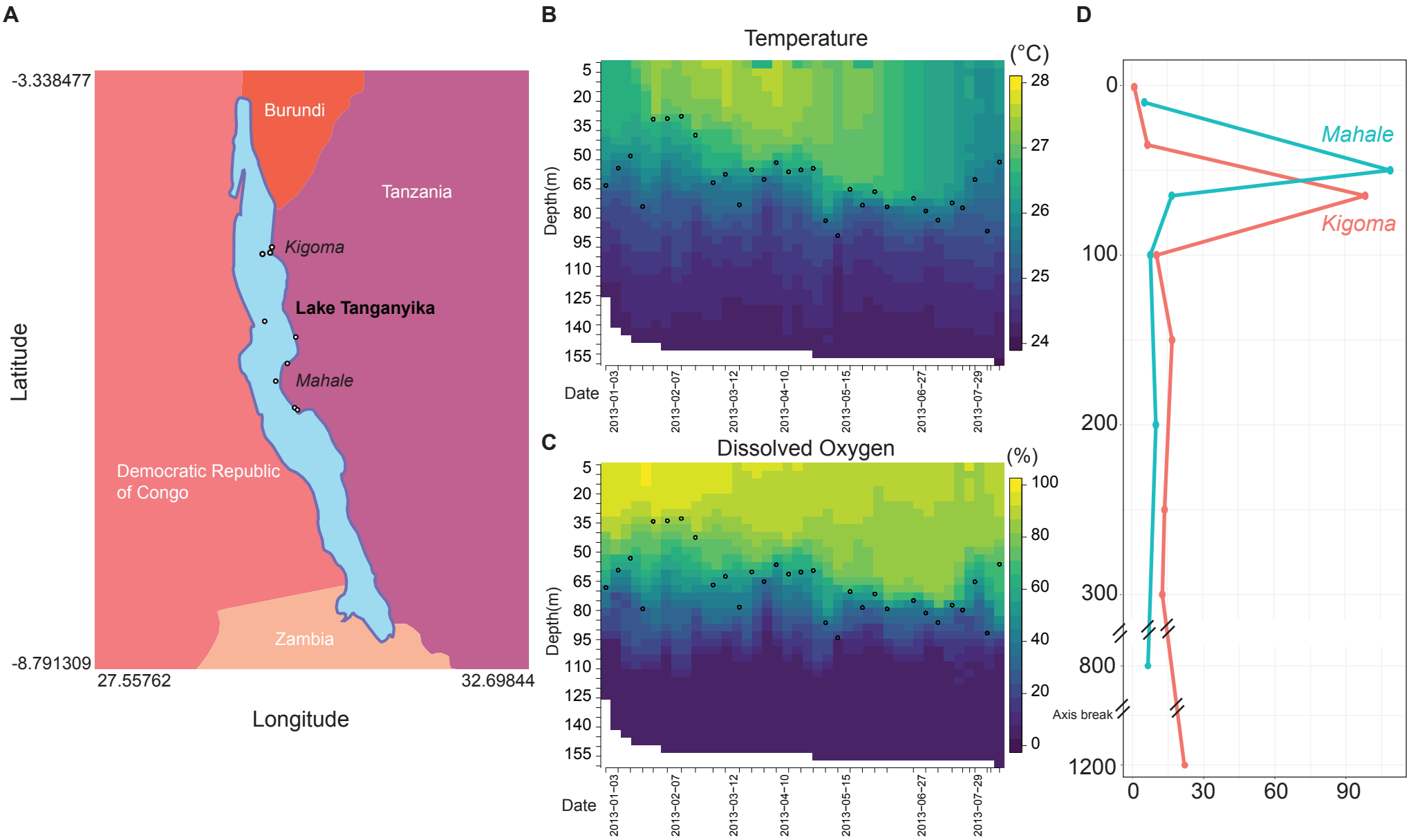
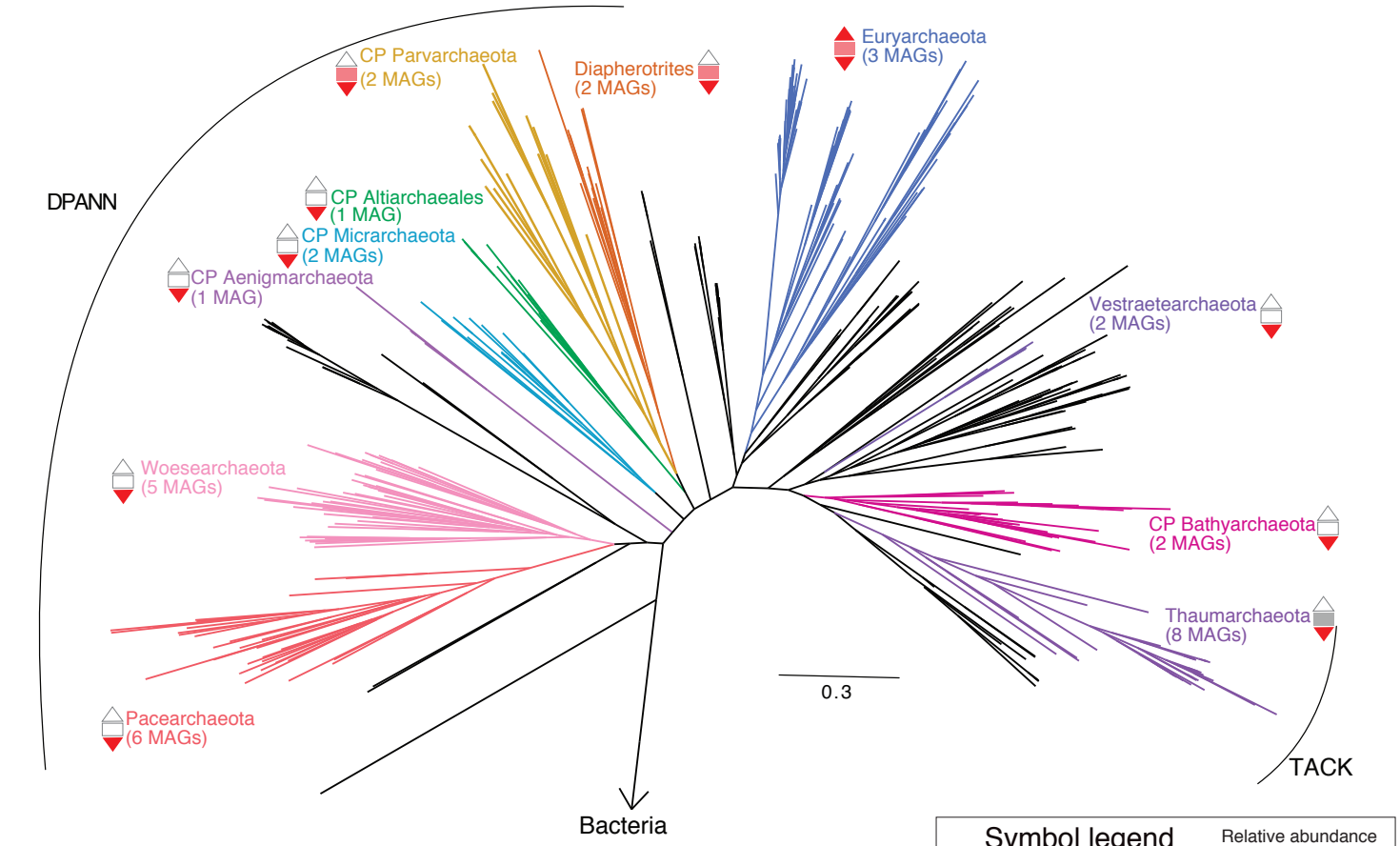

A
-3.338477
-8.791309
27.55762
32.69844
Latitude
Longitude
Burundi
Tanzania
Kigoma
Lake Tanganyika
Mahale
Democratic Republic of Congo
Zambia
B
Temperature
(°C)
Depth(m)
Date
C
Dissolved Oxygen
(%)
Depth(m)
Date
D
Mahale
Kigoma
Axis break

Figure 2

A.

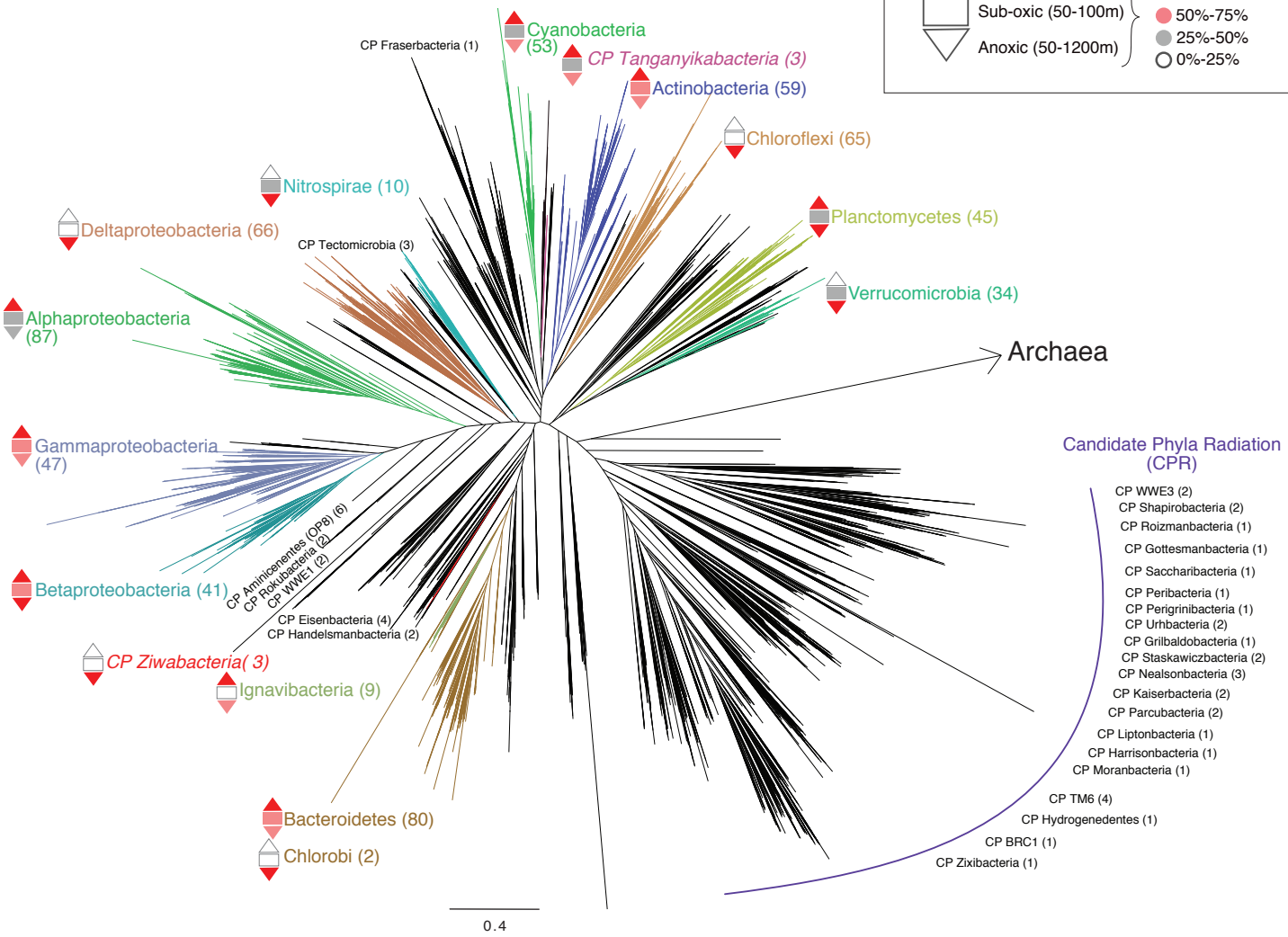


Symbol legend

Oxic (0-50 m)
Sub-oxic (50-100m)
Anoxic (50-1200m)

Relative abundance of each lineage in each layer:
- 75%-100%
- 50%-75%
- 25%-50%
- 0%-25%

B.



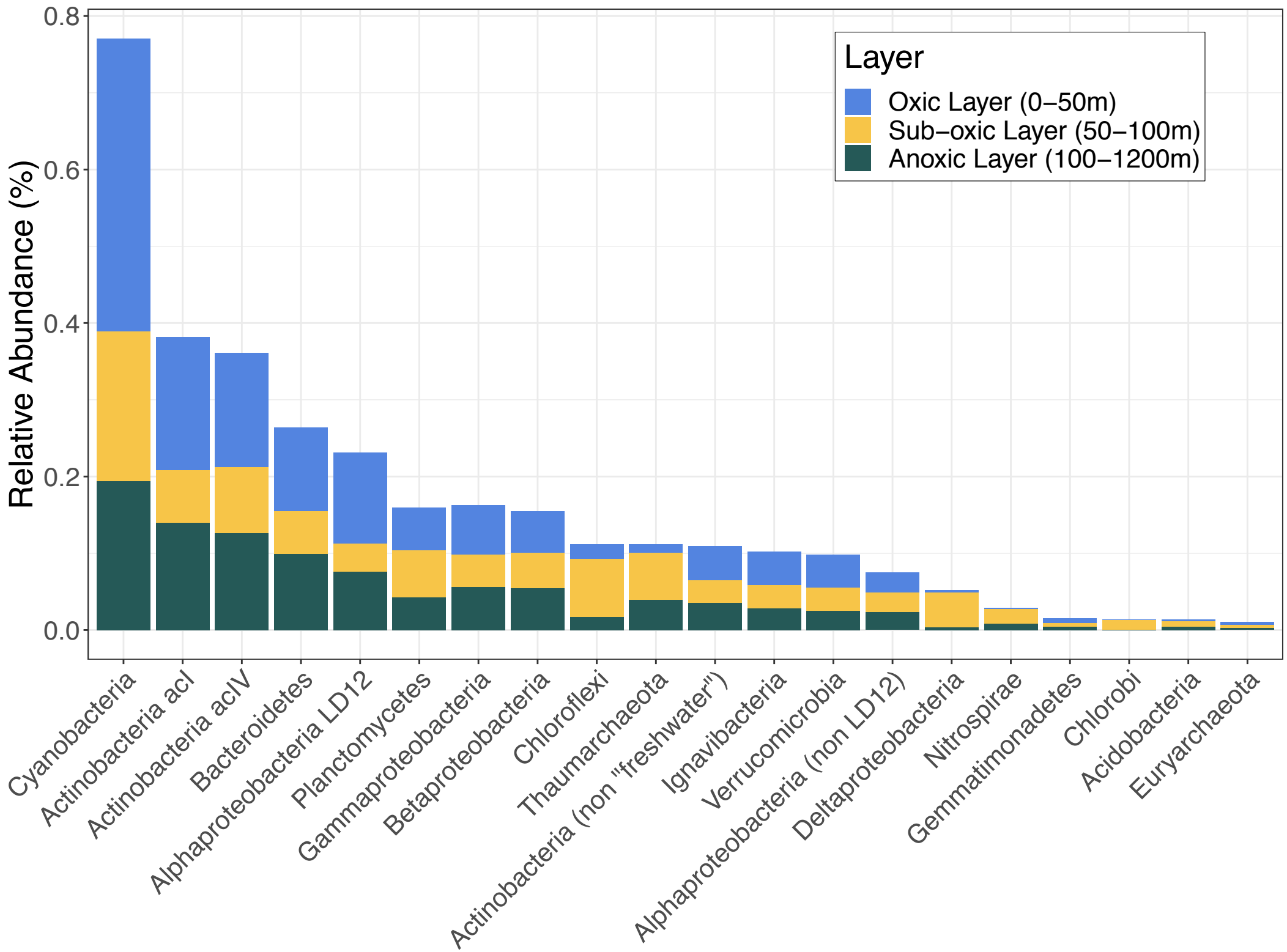

Layer
Oxic Layer (0–50m)
Sub–oxic Layer (50–100m)
Anoxic Layer (100–1200m)
Relative Abundance (%)
0.8
0.6
0.4
0.2
0.0
Cyanobacteria
Actinobacteria acI
Actinobacteria acIV
Bacteroidetes
Alphaproteobacteria LD12
Planctomycetes
Gammaproteobacteria
Betaproteobacteria
Chloroflexi
Thaumarchaeota
Actinobacteria (non "freshwater")
Ignavibacteria
Verrucomicrobia
Alphaproteobacteria (non LD12)
Deltaproteobacteria
Nitrospirae
Gemmatimonadetes
Chlorobi
Acidobacteria
Euryarchaeota
Taxonomic Group, ranked by total abundance across all layers

Figure 3

Figure 4

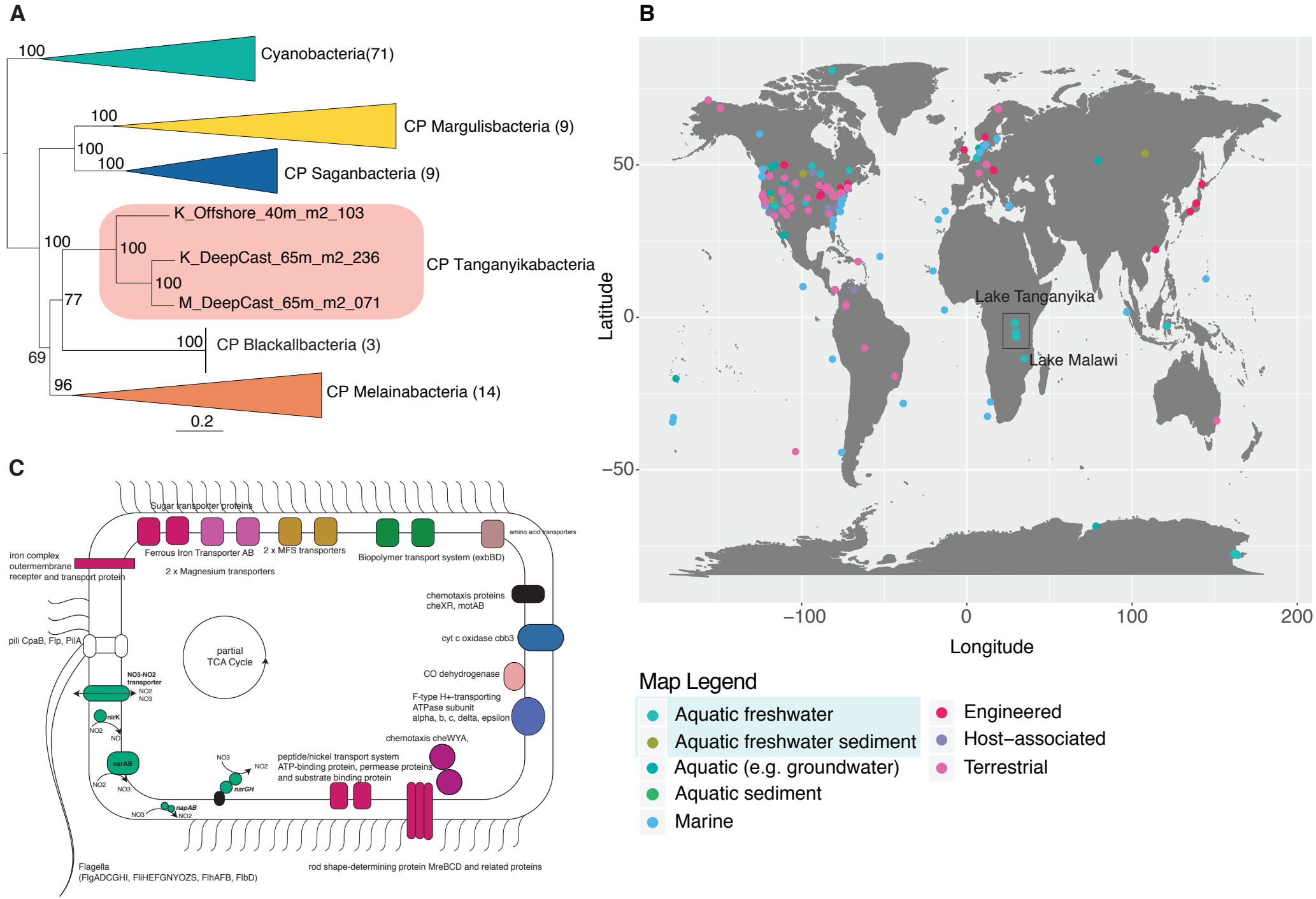

A
100
Cyanobacteria(71)
100
CP Margulisbacteria (9)
100
100
CP Saganbacteria (9)
K_Offshore_40m_m2_103
100
100
K_DeepCast_65m_m2_236
CP Tanganyikabacteria
100
M_DeepCast_65m_m2_071
77
100
CP Blackallbacteria (3)
69
96
CP Melainabacteria (14)
0.2
B
50
0
−50
Latitude
Lake Tanganyika
Lake Malawi
−100
0
100
200
Longitude
Map Legend
Aquatic freshwater
Aquatic freshwater sediment
Aquatic (e.g. groundwater)
Aquatic sediment
Marine
Engineered
Host−associated
Terrestrial
C
Sugar transporter proteins
amino acid transporters
Ferrous Iron Transporter AB
2 x MFS transporters
Biopolymer transport system (exbBD)
iron complex outermembrane recepter and transport protein
2 x Magnesium transporters
chemotaxis proteins cheXR, motAB
pili CpaB, Flp, PilA
partial TCA Cycle
cyt c oxidase cbb3
NO3-NO2 transporter
NO2
NO3
CO dehydrogenase
F-type H+-transporting ATPase subunit alpha, b, c, delta, epsilon
nirK
NO2
NO
chemotaxis cheWYA,
nxrAB
NO3
NO2
peptide/nickel transport system ATP-binding protein, permease proteins and substrate binding protein
NO2
NO3
narGH
napAB
NO3
NO2
Flagella (FlgADCGHI, FliHEFGNYOZS, FlhAFB, FlbD)
rod shape-determining protein MreBCD and related proteins

Figure 5

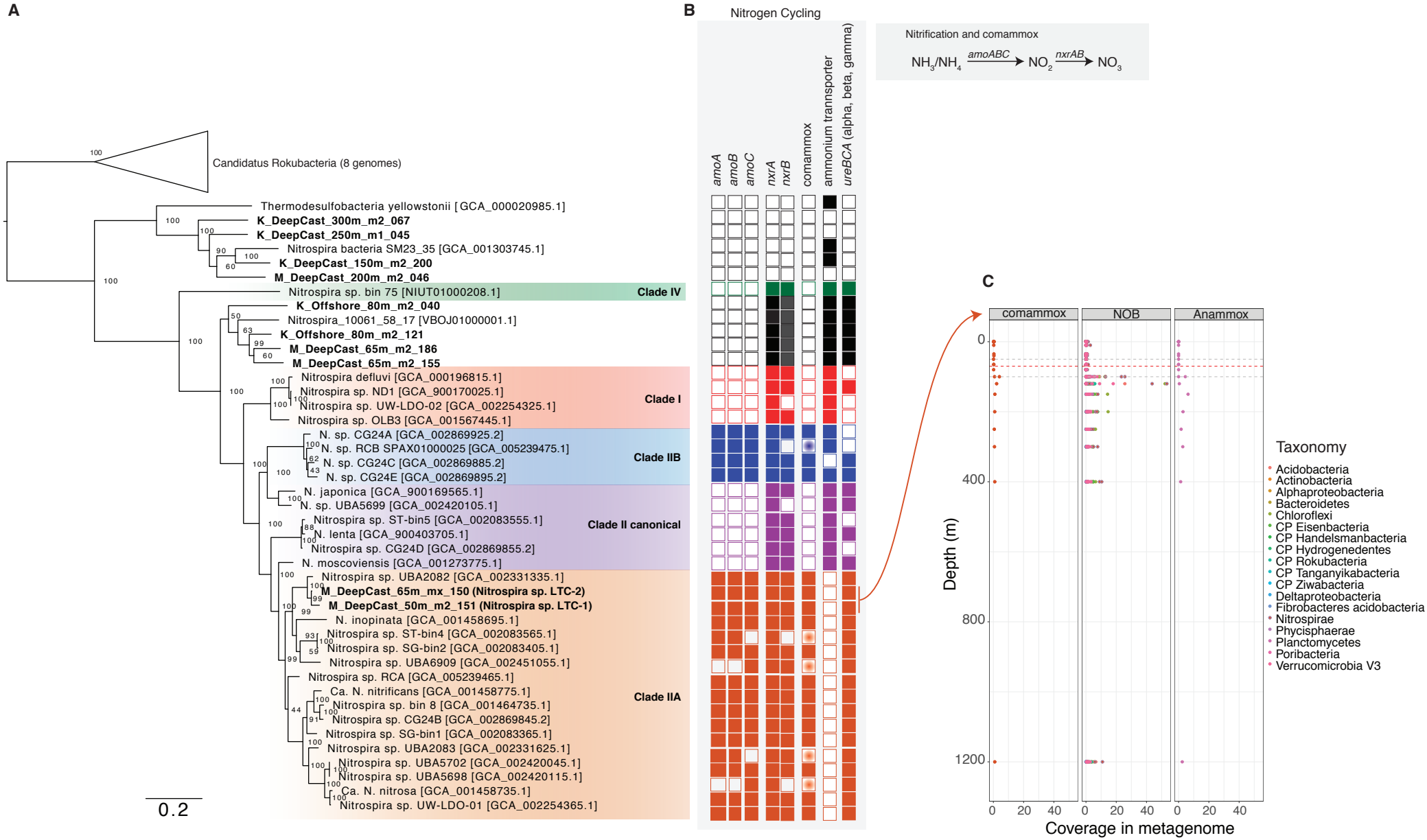

A
Candidatus Rokubacteria (8 genomes)
Thermodesulfobacteria yellowstonii [GCA_000020985.1]
K_DeepCast_300m_m2_067
K_DeepCast_250m_m1_045
Nitrospira bacteria SM23_35 [GCA_001303745.1]
K_DeepCast_150m_m2_200
M_DeepCast_200m_m2_046
Nitrospira sp. bin 75 [NIUT01000208.1]
Clade IV
K_Offshore_80m_m2_040
Nitrospira_10061_58_17 [VBOJ01000001.1]
K_Offshore_80m_m2_121
M_DeepCast_65m_m2_186
M_DeepCast_65m_m2_155
Nitrospira defluvi [GCA_000196815.1]
Nitrospira sp. ND1 [GCA_900170025.1]
Nitrospira sp. UW-LDO-02 [GCA_002254325.1]
Nitrospira sp. OLB3 [GCA_001567445.1]
Clade I
N. sp. CG24A [GCA_002869925.2]
N. sp. RCB SPAX01000025 [GCA_005239475.1]
N. sp. CG24C [GCA_002869885.2]
N. sp. CG24E [GCA_002869895.2]
Clade IIB
N. japonica [GCA_900169565.1]
N. sp. UBA5699 [GCA_002420105.1]
Nitrospira sp. ST-bin5 [GCA_002083555.1]
N. lenta [GCA_900403705.1]
Nitrospira sp. CG24D [GCA_002869855.2]
N. moscoviensis [GCA_001273775.1]
Clade II canonical
Nitrospira sp. UBA2082 [GCA_002331335.1]
M_DeepCast_65m_mx_150 (Nitrospira sp. LTC-2)
M_DeepCast_50m_m2_151 (Nitrospira sp. LTC-1)
N. inopinata [GCA_001458695.1]
Nitrospira sp. ST-bin4 [GCA_002083565.1]
Nitrospira sp. SG-bin2 [GCA_002083405.1]
Nitrospira sp. UBA6909 [GCA_002451055.1]
Nitrospira sp. RCA [GCA_005239465.1]
Ca. N. nitrificans [GCA_001458775.1]
Nitrospira sp. bin 8 [GCA_001464735.1]
Nitrospira sp. CG24B [GCA_002869845.2]
Nitrospira sp. SG-bin1 [GCA_002083365.1]
Nitrospira sp. UBA2083 [GCA_002331625.1]
Nitrospira sp. UBA5702 [GCA_002420045.1]
Nitrospira sp. UBA5698 [GCA_002420115.1]
Ca. N. nitrosa [GCA_001458735.1]
Nitrospira sp. UW-LDO-01 [GCA_002254365.1]
Clade IIA
0.2
B
Nitrogen Cycling
amoA
amoB
amoC
nxrA
nxrB
comammox
ammonium trannsporter
ureBCA (alpha, beta, gamma)
Nitrification and comammox
NH3/NH4 amoABC NO2 nxrAB NO3
C
comammox
NOB
Anammox
Depth (m)
0
400
800
1200
0 20 40
Coverage in metagenome
Taxonomy
Acidobacteria
Actinobacteria
Alphaproteobacteria
Bacteroidetes
Chloroflexi
CP Eisenbacteria
CP Handelsmanbacteria
CP Hydrogenedentes
CP Rokubacteria
CP Tanganyikabacteria
CP Ziwabacteria
Deltaproteobacteria
Fibrobacteres acidobacteria
Nitrospirae
Phycisphaerae
Planctomycetes
Poribacteria
Verrucomicrobia V3

Figure 6

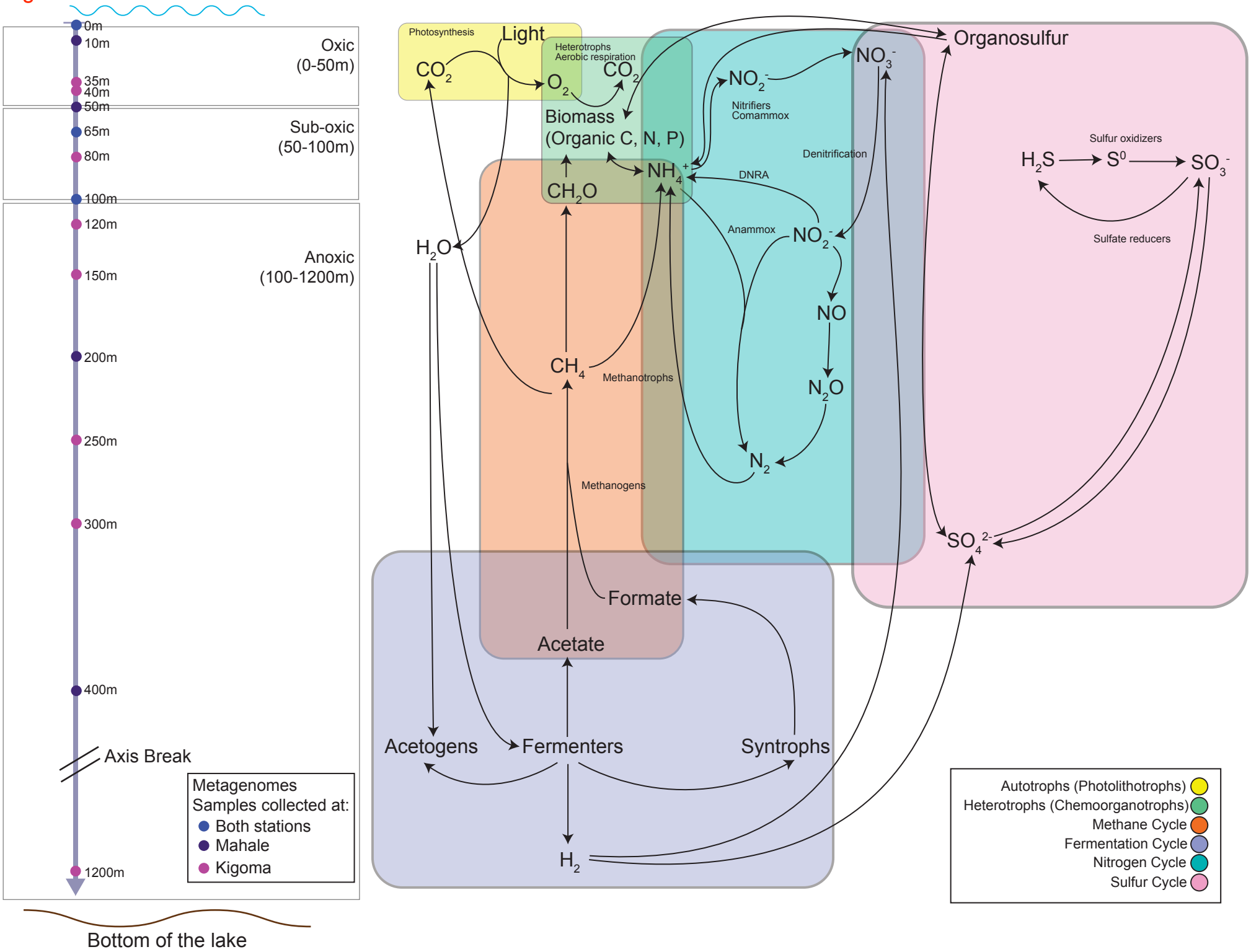

Lake Surface
0m
10m
35m
40m
50m
65m
80m
100m
120m
150m
200m
250m
300m
400m
1200m
Oxic
(0-50m)
Sub-oxic
(50-100m)
Anoxic
(100-1200m)
Axis Break
Metagenomes
Samples collected at:
Both stations
Mahale
Kigoma
Bottom of the lake
Photosynthesis
Light
CO_2
O_2
Heterotrophs
Aerobic respiration
CO_2
Biomass
(Organic C, N, P)
CH_2O
H_2O
CH_4
Methanotrophs
Methanogens
Formate
Acetate
Acetogens
Fermenters
Syntrophs
H_2
NH_4^+
NO_2^-
Nitrifiers
Comammox
NO_3^-
Denitrification
DNRA
Anammox
NO_2^-
NO
N_2O
N_2
Organosulfur
Sulfur oxidizers
H_2S
S^0
SO_3^-
Sulfate reducers
SO_4^{2-}
Autotrophs (Photolithotrophs)
Heterotrophs (Chemoorganotrophs)
Methane Cycle
Fermentation Cycle
Nitrogen Cycle
Sulfur Cycle

Figure S1

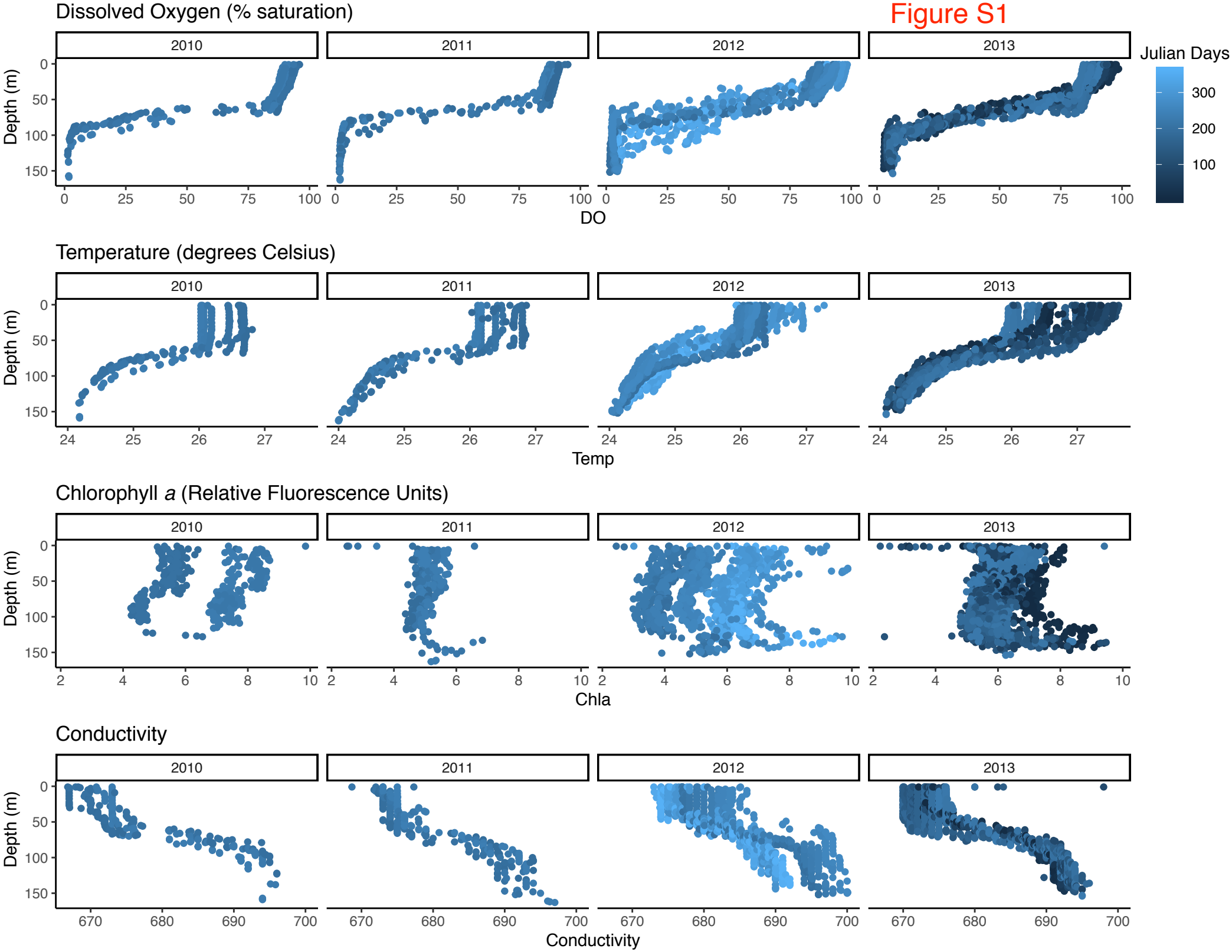
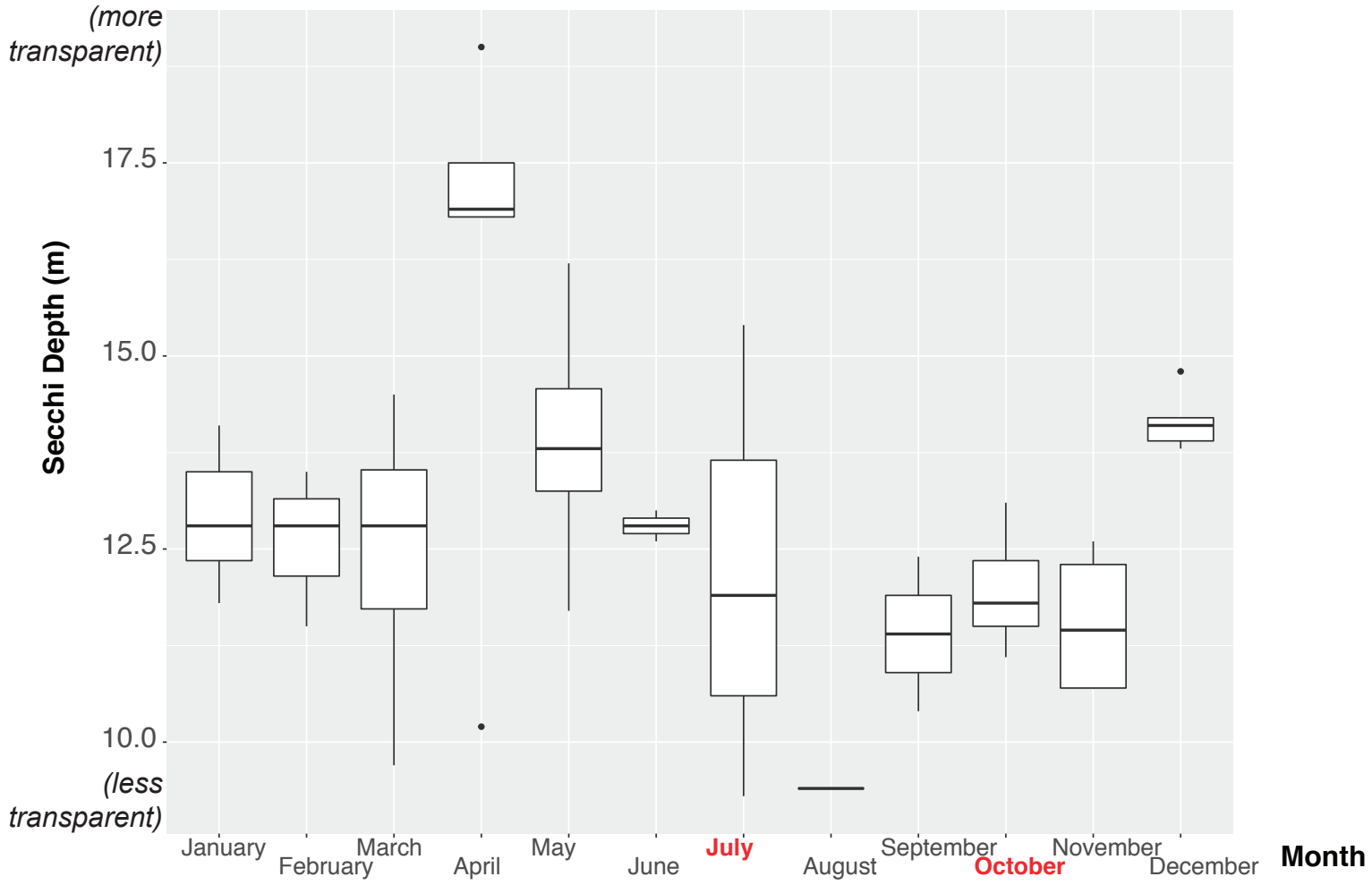

Dissolved Oxygen (% saturation)
2010
2011
2012
2013
Depth (m)
DO
Julian Days
300
200
100
Temperature (degrees Celsius)
Temp
Chlorophyll a (Relative Fluorescence Units)
Chla
Conductivity

Figure S2

A.


(more transparent)
17.5
15.0
12.5
10.0
(less transparent)
Secchi Depth (m)
January
February
March
April
May
June
July
August
September
October
November
December
Month

B.

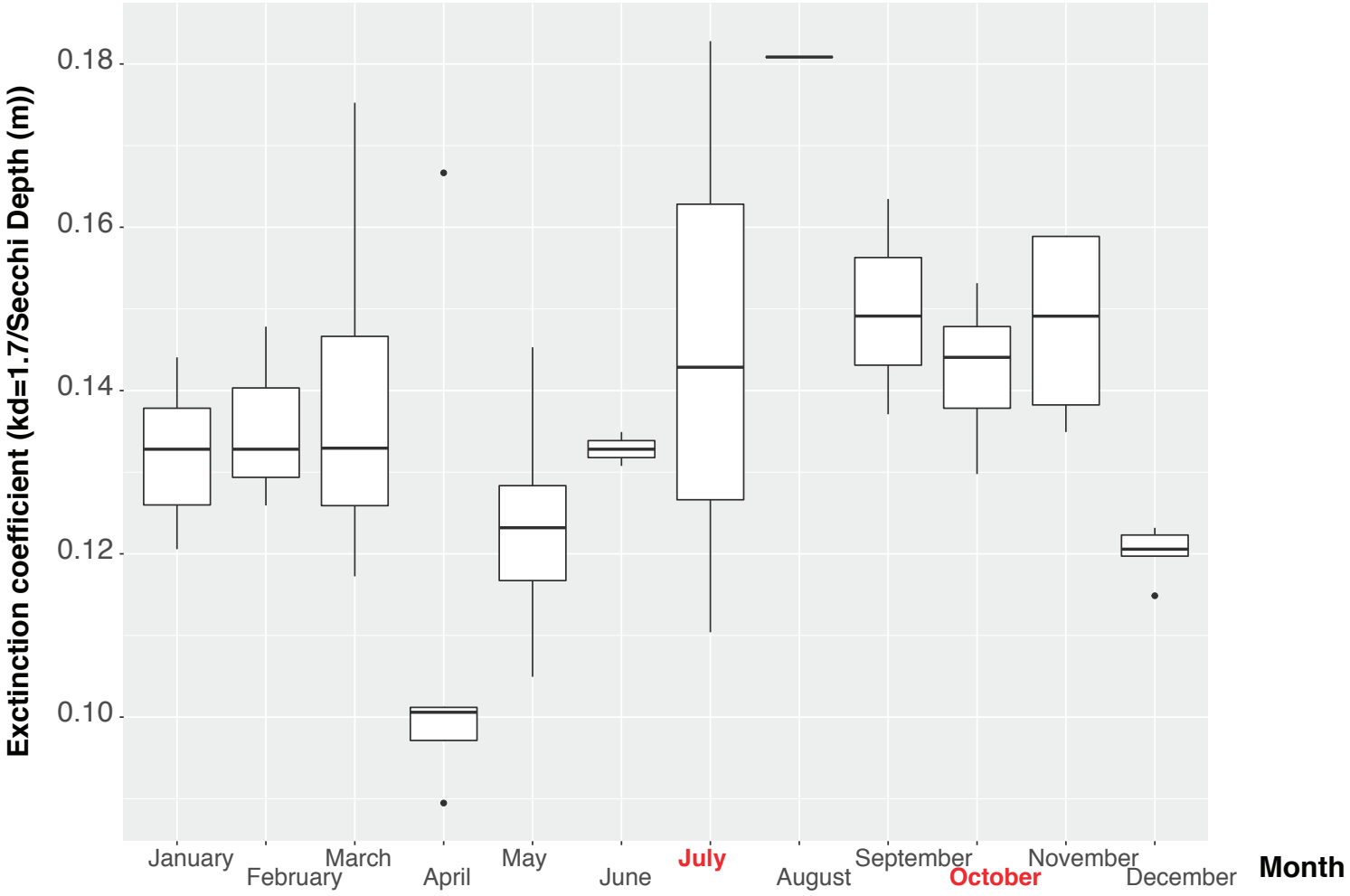

0.18
0.16
0.14
0.12
0.10
Exctinction coefficient (kd=1.7/Secchi Depth (m))
January
February
March
April
May
June
July
August
September
October
November
December
Month

Figure S3

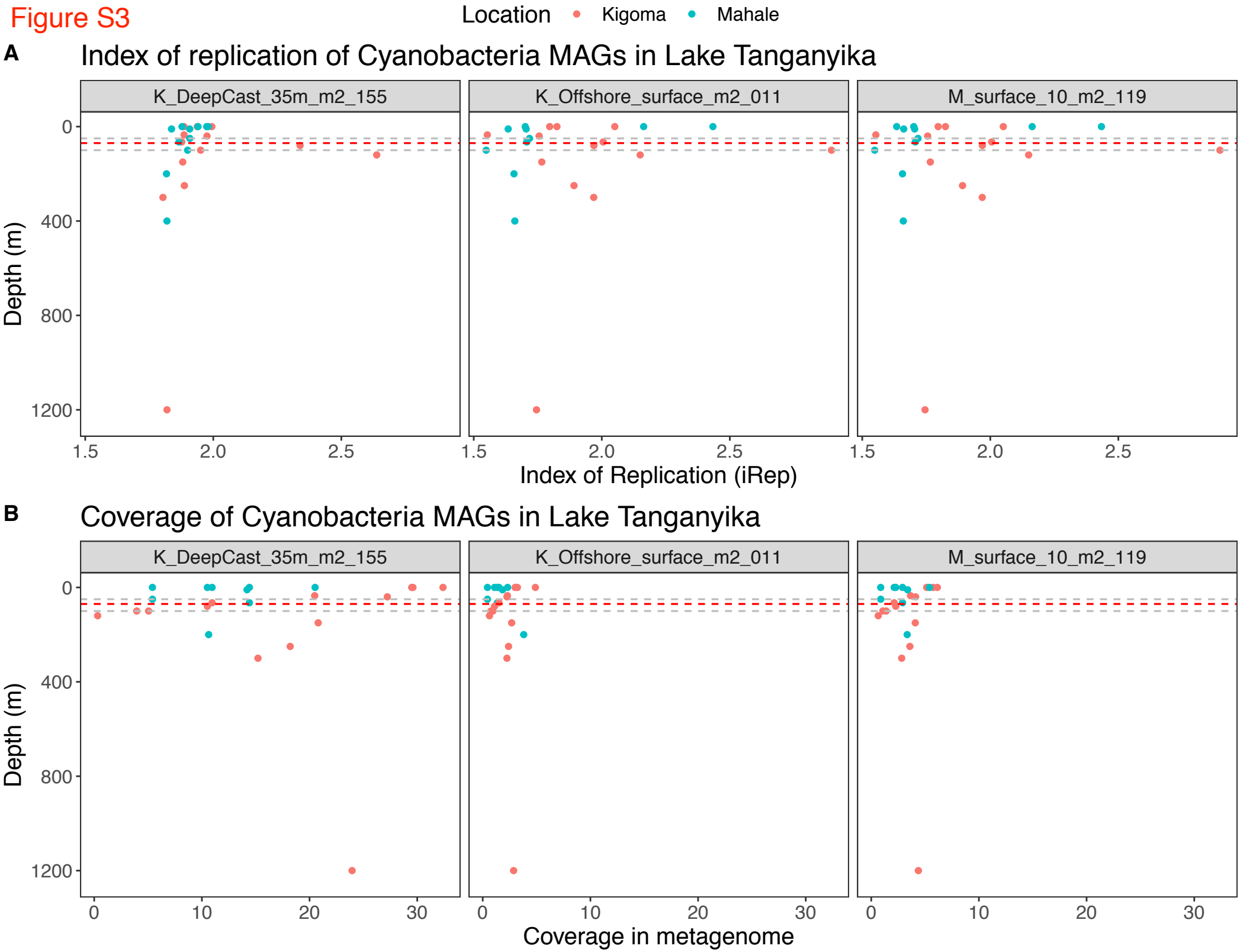

Location
Kigoma
Mahale
A
Index of replication of Cyanobacteria MAGs in Lake Tanganyika
K_DeepCast_35m_m2_155
K_Offshore_surface_m2_011
M_surface_10_m2_119
Depth (m)
0
400
800
1200
1.5
2.0
2.5
Index of Replication (iRep)
B
Coverage of Cyanobacteria MAGs in Lake Tanganyika
K_DeepCast_35m_m2_155
K_Offshore_surface_m2_011
M_surface_10_m2_119
Depth (m)
0
400
800
1200
0
10
20
30
Coverage in metagenome

Figure S4

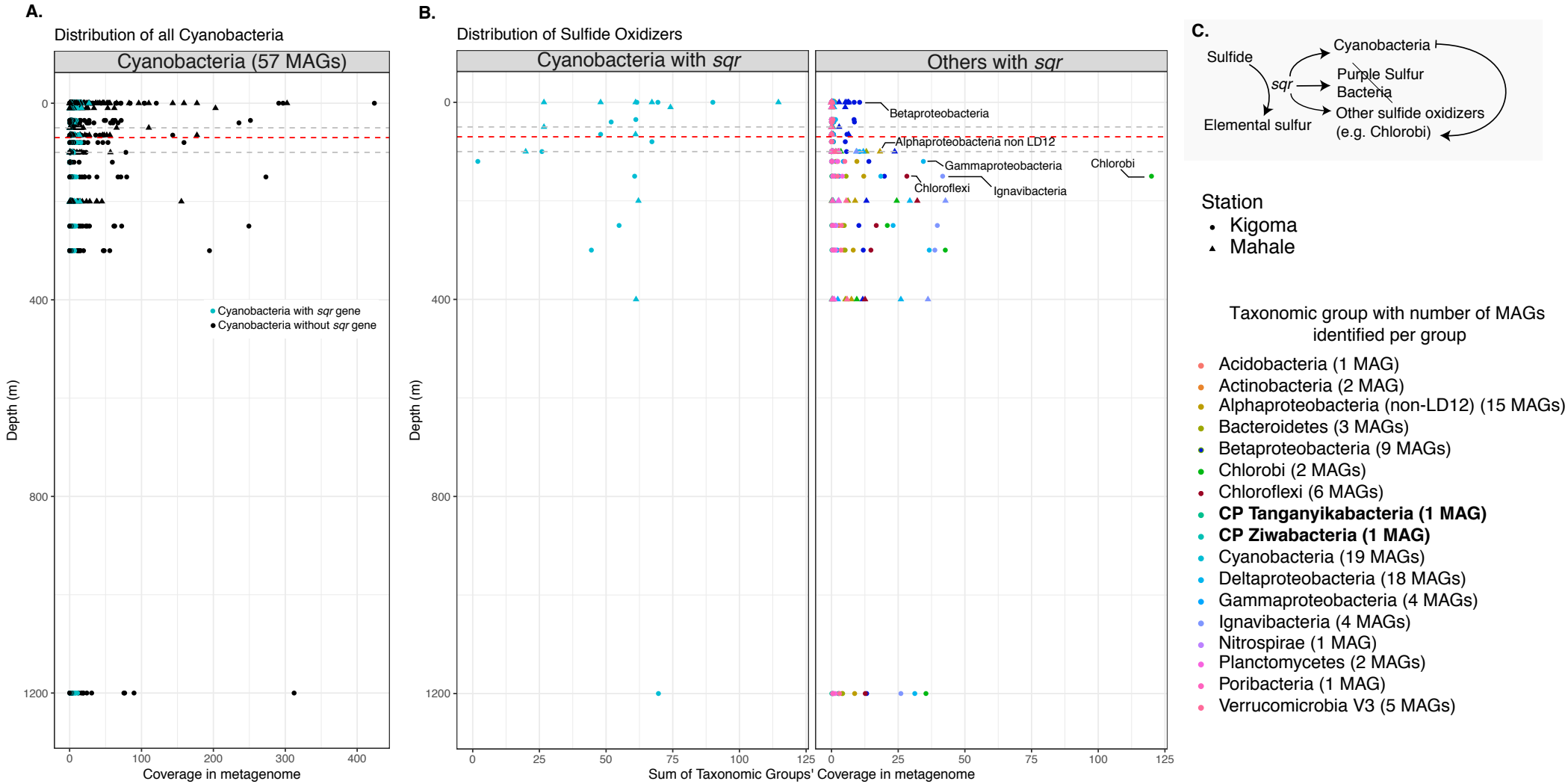

A.
Distribution of all Cyanobacteria
Cyanobacteria (57 MAGs)
Cyanobacteria with sqr gene
Cyanobacteria without sqr gene
Depth (m)
Coverage in metagenome
B.
Distribution of Sulfide Oxidizers
Cyanobacteria with sqr
Others with sqr
Betaproteobacteria
Alphaproteobacteria non LD12
Gammaproteobacteria
Chlorobi
Chloroflexi
Ignavibacteria
Sum of Taxonomic Groups' Coverage in metagenome
C.
Sulfide
sqr
Elemental sulfur
Cyanobacteria
Purple Sulfur Bacteria
Other sulfide oxidizers (e.g. Chlorobi)
Station
Kigoma
Mahale
Taxonomic group with number of MAGs identified per group
Acidobacteria (1 MAG)
Actinobacteria (2 MAG)
Alphaproteobacteria (non-LD12) (15 MAGs)
Bacteroidetes (3 MAGs)
Betaproteobacteria (9 MAGs)
Chlorobi (2 MAGs)
Chloroflexi (6 MAGs)
CP Tanganyikabacteria (1 MAG)
CP Ziwabacteria (1 MAG)
Cyanobacteria (19 MAGs)
Deltaproteobacteria (18 MAGs)
Gammaproteobacteria (4 MAGs)
Ignavibacteria (4 MAGs)
Nitrospirae (1 MAG)
Planctomycetes (2 MAGs)
Poribacteria (1 MAG)
Verrucomicrobia V3 (5 MAGs)

Figure S5

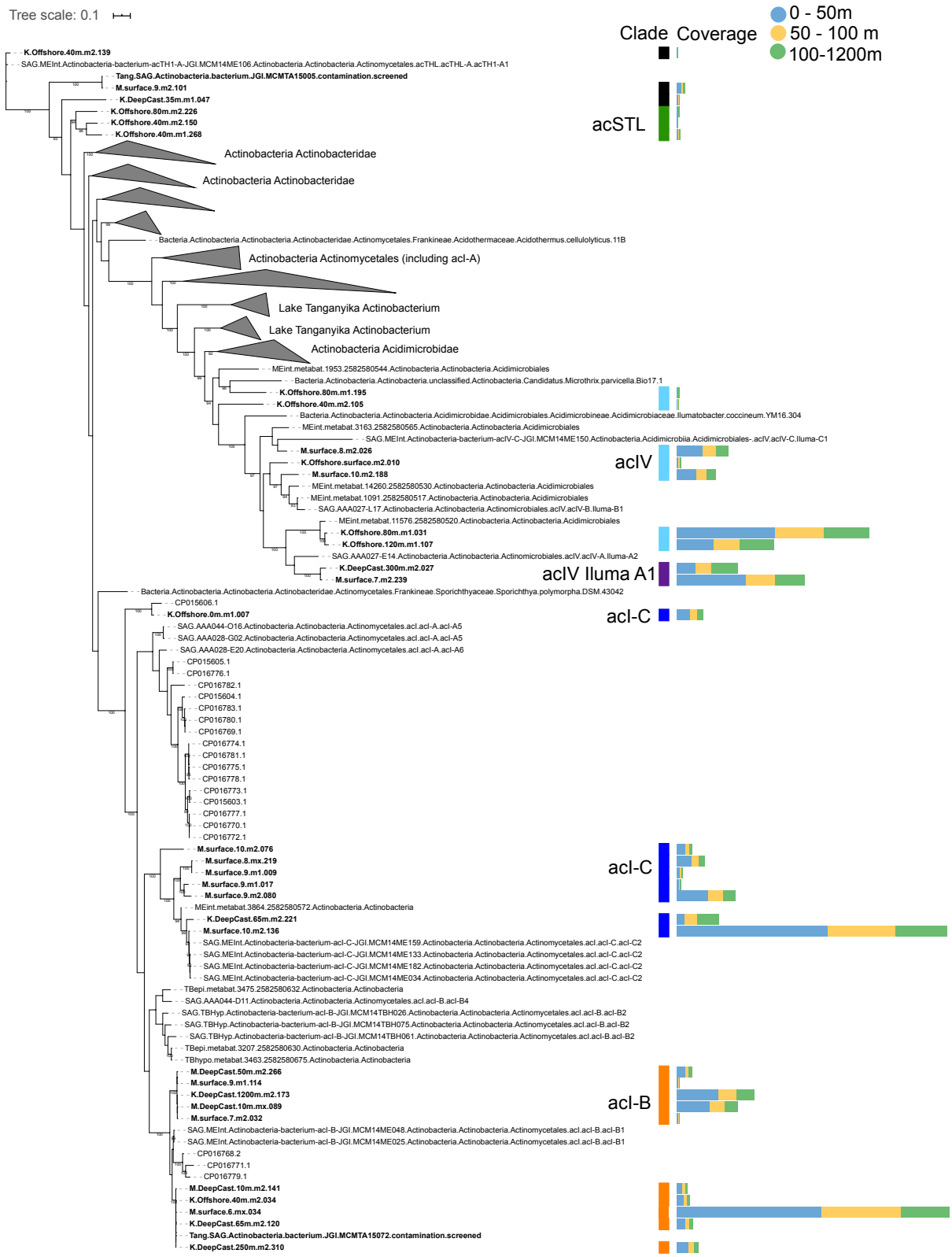

Tree scale: 0.1
Clade
Coverage
0 - 50m
50 - 100 m
100-1200m
K.Offshore.40m.m2.139
SAG.MEInt.Actinobacteria-bacterium-acTH1-A-JGI.MCM14ME106.Actinobacteria.Actinobacteria.Actinomycetales.acTHL.acTHL-A.acTH1-A1
Tang.SAG.Actinobacteria.bacterium.JGI.MCMTA15005.contamination.screened
M.surface.9.m2.101
K.DeepCast.35m.m1.047
K.Offshore.80m.m2.226
K.Offshore.40m.m2.150
K.Offshore.40m.m1.268
Actinobacteria Actinobacteridae
Actinobacteria Actinobacteridae
Bacteria.Actinobacteria.Actinobacteria.Actinobacteridae.Actinomycetales.Frankineae.Acidothermaceae.Acidothermus.cellulolyticus.11B
Actinobacteria Actinomycetales (including acl-A)
Lake Tanganyika Actinobacterium
Lake Tanganyika Actinobacterium
Actinobacteria Acidimicrobidae
MEint.metabat.1953.2582580544.Actinobacteria.Actinobacteria.Acidimicrobiales
Bacteria.Actinobacteria.Actinobacteria.unclassified.Actinobacteria.Candidatus.Microthrix.parvicella.Bio17.1
K.Offshore.80m.m1.195
K.Offshore.40m.m2.105
Bacteria.Actinobacteria.Actinobacteria.Acidimicrobidae.Acidimicrobiales.Acidimicrobineae.Acidimicrobiaceae.Ilumatobacter.coccineum.YM16.304
MEint.metabat.3163.2582580565.Actinobacteria.Actinobacteria.Acidimicrobiales
SAG.MEInt.Actinobacteria-bacterium-acIV-C-JGI.MCM14ME150.Actinobacteria.Acidimicrobiia.Acidimicrobiales-.acIV.acIV-C.Iluma-C1
M.surface.8.m2.026
K.Offshore.surface.m2.010
M.surface.10.m2.188
MEint.metabat.14260.2582580530.Actinobacteria.Actinobacteria.Acidimicrobiales
MEint.metabat.1091.2582580517.Actinobacteria.Actinobacteria.Acidimicrobiales
SAG.AAA027-L17.Actinobacteria.Actinobacteria.Actinomicrobiales.acIV.acIV-B.Iluma-B1
MEint.metabat.11576.2582580520.Actinobacteria.Actinobacteria.Acidimicrobiales
K.Offshore.80m.m1.031
K.Offshore.120m.m1.107
SAG.AAA027-E14.Actinobacteria.Actinobacteria.Actinomicrobiales.acIV.acIV-A.Iluma-A2
K.DeepCast.300m.m2.027
M.surface.7.m2.239
Bacteria.Actinobacteria.Actinobacteria.Actinobacteridae.Actinomycetales.Frankineae.Sporichthyaceae.Sporichthya.polymorpha.DSM.43042
CP015606.1
K.Offshore.0m.m1.007
SAG.AAA044-O16.Actinobacteria.Actinobacteria.Actinomycetales.acl.acl-A.acl-A5
SAG.AAA028-G02.Actinobacteria.Actinobacteria.Actinomycetales.acl.acl-A.acl-A5
SAG.AAA028-E20.Actinobacteria.Actinobacteria.Actinomycetales.acl.acl-A.acl-A6
CP015605.1
CP016776.1
CP016782.1
CP015604.1
CP016783.1
CP016780.1
CP016769.1
CP016774.1
CP016781.1
CP016775.1
CP016778.1
CP016773.1
CP015603.1
CP016777.1
CP016770.1
CP016772.1
M.surface.10.m2.076
M.surface.8.mx.219
M.surface.9.m1.009
M.surface.9.m1.017
M.surface.9.m2.080
MEint.metabat.3864.2582580572.Actinobacteria.Actinobacteria
K.DeepCast.65m.m2.221
M.surface.10.m2.136
SAG.MEInt.Actinobacteria-bacterium-acl-C-JGI.MCM14ME159.Actinobacteria.Actinobacteria.Actinomycetales.acl.acl-C.acl-C2
SAG.MEInt.Actinobacteria-bacterium-acl-C-JGI.MCM14ME133.Actinobacteria.Actinobacteria.Actinomycetales.acl.acl-C.acl-C2
SAG.MEInt.Actinobacteria-bacterium-acl-C-JGI.MCM14ME182.Actinobacteria.Actinobacteria.Actinomycetales.acl.acl-C.acl-C2
SAG.MEInt.Actinobacteria-bacterium-acl-C-JGI.MCM14ME034.Actinobacteria.Actinobacteria.Actinomycetales.acl.acl-C.acl-C2
TBepi.metabat.3475.2582580632.Actinobacteria.Actinobacteria
SAG.AAA044-D11.Actinobacteria.Actinobacteria.Actinomycetales.acl.acl-B.acl-B4
SAG.TBHyp.Actinobacteria-bacterium-acl-B-JGI.MCM14TBH026.Actinobacteria.Actinobacteria.Actinomycetales.acl.acl-B.acl-B2
SAG.TBHyp.Actinobacteria-bacterium-acl-B-JGI.MCM14TBH075.Actinobacteria.Actinobacteria.Actinomycetales.acl.acl-B.acl-B2
SAG.TBHyp.Actinobacteria-bacterium-acl-B-JGI.MCM14TBH061.Actinobacteria.Actinobacteria.Actinomycetales.acl.acl-B.acl-B2
TBepi.metabat.3207.2582580630.Actinobacteria.Actinobacteria
TBhypo.metabat.3463.2582580675.Actinobacteria.Actinobacteria
M.DeepCast.50m.m2.266
M.surface.9.m1.114
K.DeepCast.1200m.m2.173
M.DeepCast.10m.mx.089
M.surface.7.m2.032
SAG.MEInt.Actinobacteria-bacterium-acl-B-JGI.MCM14ME048.Actinobacteria.Actinobacteria.Actinomycetales.acl.acl-B.acl-B1
SAG.MEInt.Actinobacteria-bacterium-acl-B-JGI.MCM14ME025.Actinobacteria.Actinobacteria.Actinomycetales.acl.acl-B.acl-B1
CP016768.2
CP016771.1
CP016779.1
M.DeepCast.10m.m2.141
K.Offshore.40m.m2.034
M.surface.6.mx.034
K.DeepCast.65m.m2.120
Tang.SAG.Actinobacteria.bacterium.JGI.MCMTA15072.contamination.screened
K.DeepCast.250m.m2.310
acSTL
acIV
acIV Iluma A1
acl-C
acl-C
acl-B

Figure S6

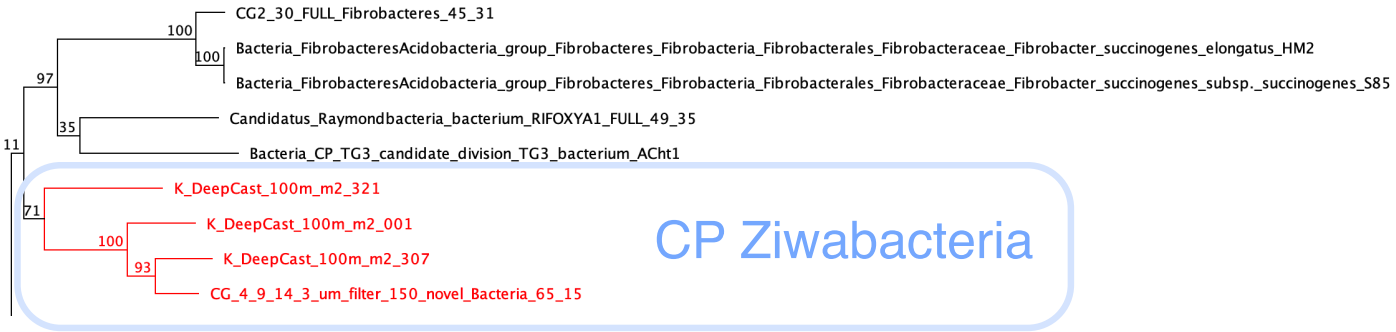

CG2_30_FULL_Fibrobacteres_45_31
Bacteria_FibrobacteresAcidobacteria_group_Fibrobacteres_Fibrobacteria_Fibrobacterales_Fibrobacteraceae_Fibrobacter_succinogenes_elongatus_HM2
Bacteria_FibrobacteresAcidobacteria_group_Fibrobacteres_Fibrobacteria_Fibrobacterales_Fibrobacteraceae_Fibrobacter_succinogenes_subsp._succinogenes_S85
Candidatus_Raymondbacteria_bacterium_RIFOXYA1_FULL_49_35
Bacteria_CP_TG3_candidate_division_TG3_bacterium_ACht1
K_DeepCast_100m_m2_321
K_DeepCast_100m_m2_001
K_DeepCast_100m_m2_307
CG_4_9_14_3_um_filter_150_novel_Bacteria_65_15
CP Ziwabacteria
100
100
97
35
11
71
100
93

Figure S7 A

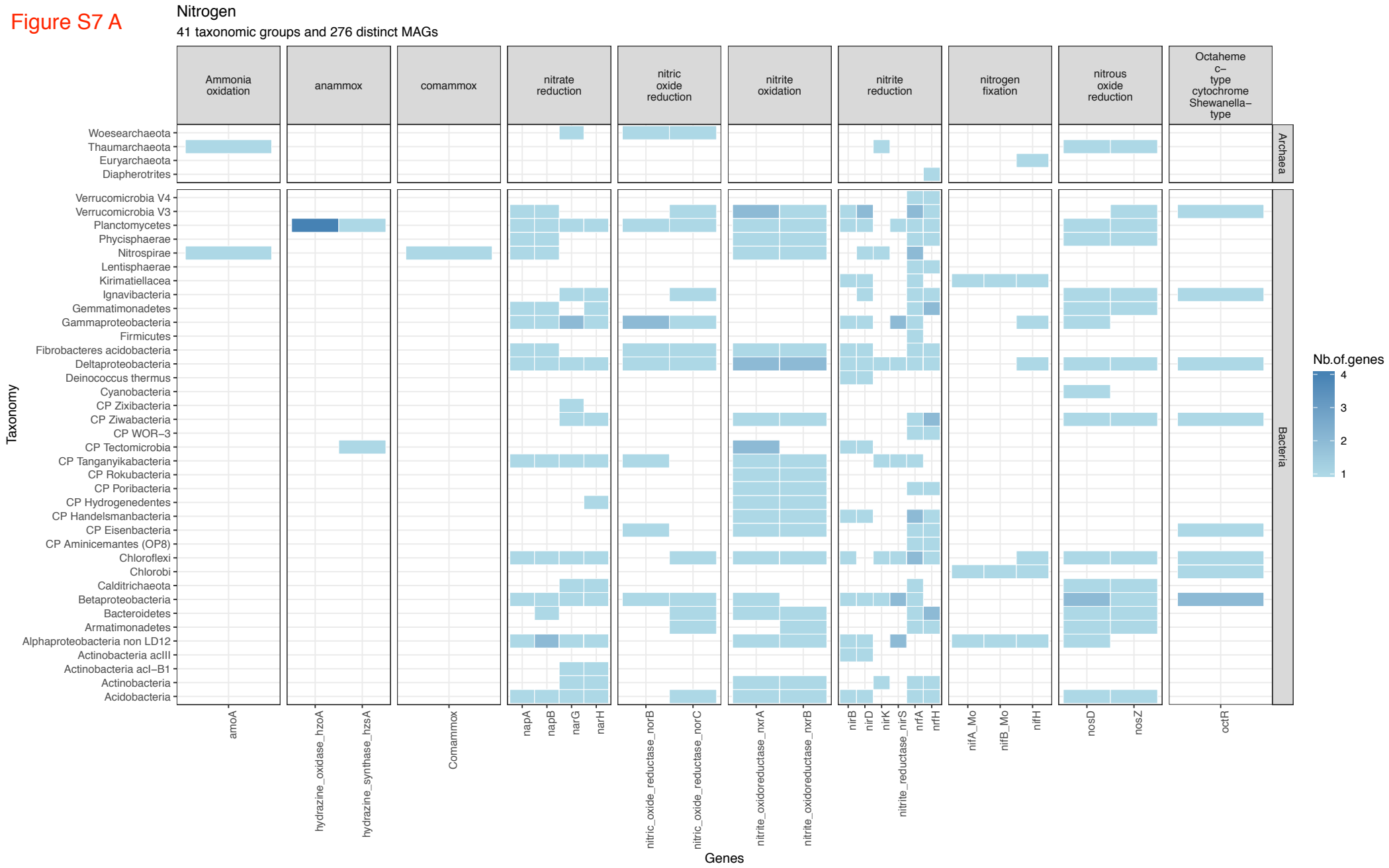

Nitrogen
41 taxonomic groups and 276 distinct MAGs
Ammonia oxidation
anammox
comammox
nitrate reduction
nitric oxide reduction
nitrite oxidation
nitrite reduction
nitrogen fixation
nitrous oxide reduction
Octaheme c–type cytochrome Shewanella–type
Archaea
Bacteria
Woesearchaeota
Thaumarchaeota
Euryarchaeota
Diapherotrites
Verrucomicrobia V4
Verrucomicrobia V3
Planctomycetes
Phycisphaerae
Nitrospirae
Lentisphaerae
Kirimatiellacea
Ignavibacteria
Gemmatimonadetes
Gammaproteobacteria
Firmicutes
Fibrobacteres acidobacteria
Deltaproteobacteria
Deinococcus thermus
Cyanobacteria
CP Zixibacteria
CP Ziwabacteria
CP WOR–3
CP Tectomicrobia
CP Tanganyikabacteria
CP Rokubacteria
CP Poribacteria
CP Hydrogenedentes
CP Handelsmanbacteria
CP Eisenbacteria
CP Aminicemantes (OP8)
Chloroflexi
Chlorobi
Calditrichaeota
Betaproteobacteria
Bacteroidetes
Armatimonadetes
Alphaproteobacteria non LD12
Actinobacteria acIII
Actinobacteria acI–B1
Actinobacteria
Acidobacteria
Taxonomy
amoA
hydrazine_oxidase_hzoA
hydrazine_synthase_hzsA
Comammox
napA
napB
narG
narH
nitric_oxide_reductase_norB
nitric_oxide_reductase_norC
nitrite_oxidoreductase_nxrA
nitrite_oxidoreductase_nxrB
nirB
nirD
nirK
nitrite_reductase_nirS
nrfA
nrfH
nifA_Mo
nifB_Mo
nifH
nosD
nosZ
octR
Genes
Nb.of.genes
4
3
2
1

Figure S7 B

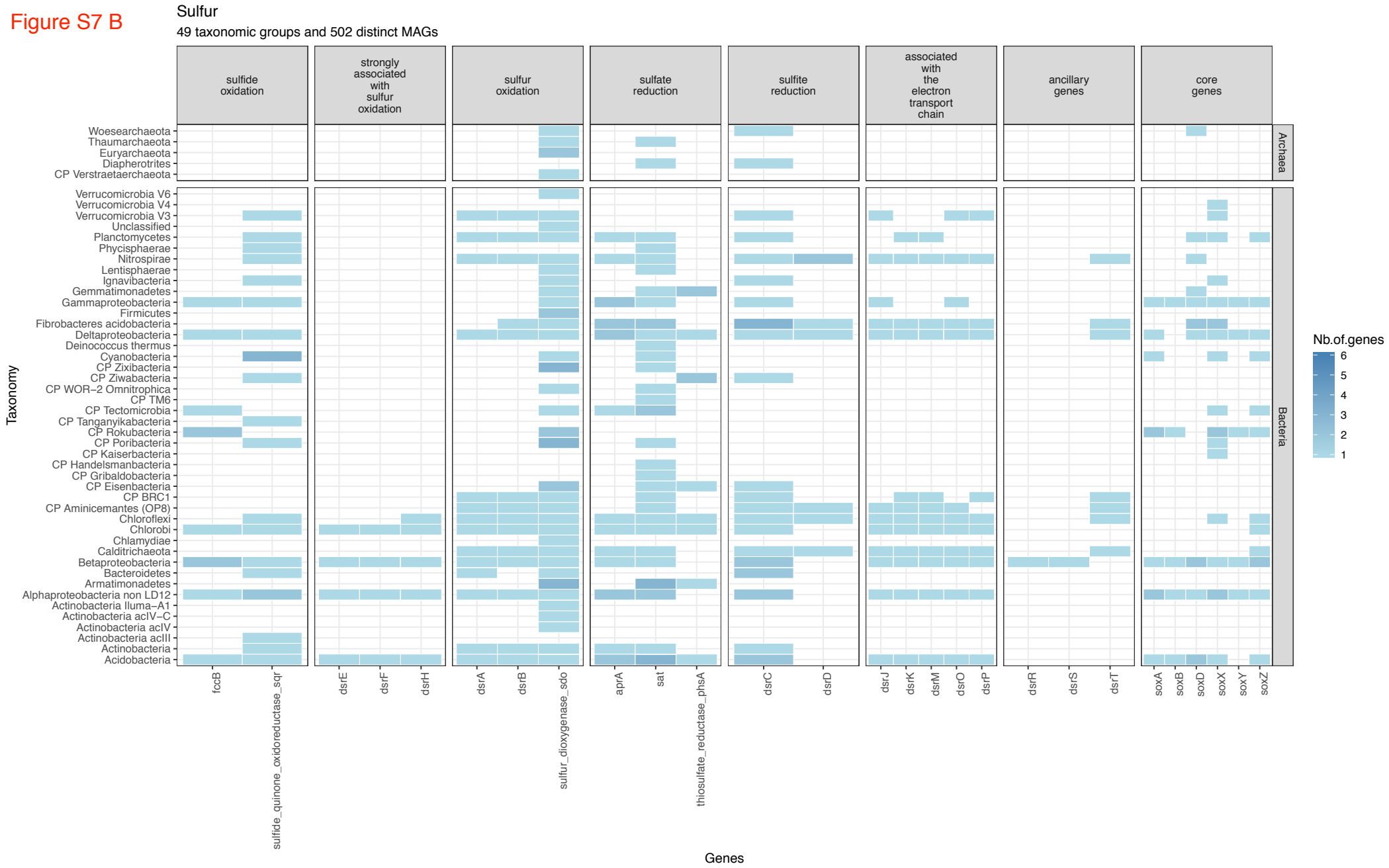

Sulfur
49 taxonomic groups and 502 distinct MAGs
sulfide oxidation
strongly associated with sulfur oxidation
sulfur oxidation
sulfate reduction
sulfite reduction
associated with the electron transport chain
ancillary genes
core genes
Archaea
Bacteria
Taxonomy
Woesearchaeota
Thaumarchaeota
Euryarchaeota
Diapherotrites
CP Verstraetaerchaeota
Verrucomicrobia V6
Verrucomicrobia V4
Verrucomicrobia V3
Unclassified
Planctomycetes
Phycisphaerae
Nitrospirae
Lentisphaerae
Ignavibacteria
Gemmatimonadetes
Gammaproteobacteria
Firmicutes
Fibrobacteres acidobacteria
Deltaproteobacteria
Deinococcus thermus
Cyanobacteria
CP Zixibacteria
CP Ziwabacteria
CP WOR−2 Omnitrophica
CP TM6
CP Tectomicrobia
CP Tanganyikabacteria
CP Rokubacteria
CP Poribacteria
CP Kaiserbacteria
CP Handelsmanbacteria
CP Gribaldobacteria
CP Eisenbacteria
CP BRC1
CP Aminicemantes (OP8)
Chloroflexi
Chlorobi
Chlamydiae
Calditrichaeota
Betaproteobacteria
Bacteroidetes
Armatimonadetes
Alphaproteobacteria non LD12
Actinobacteria Iluma−A1
Actinobacteria acIV−C
Actinobacteria acIV
Actinobacteria acIII
Actinobacteria
Acidobacteria
fccB
sulfide_quinone_oxidoreductase_sqr
dsrE
dsrF
dsrH
dsrA
dsrB
sulfur_dioxygenase_sdo
aprA
sat
thiosulfate_reductase_phsA
dsrC
dsrD
dsrJ
dsrK
dsrM
dsrO
dsrP
dsrR
dsrS
dsrT
soxA
soxB
soxD
soxX
soxY
soxZ
Genes
Nb.of.genes
6
5
4
3
2
1

Figure S7 C

Hydrogen
46 taxonomic groups and 220 distinct MAGs
FeFeHydrogenase
NiFe Hydrogenase
Archaea
Bacteria
Taxonomy
Woesearchaeota
Euryarchaeota
Diapherotrites
CP Verstraetaerchaeota
CP Parvarchaeota
CP Pacearchaeota
CP Bathyarchaeota
CP Altiarchaeota
Verrucomicrobia V4
Verrucomicrobia V3
Planctomycetes
Phycisphaerae
Nitrospirae
Lentisphaerae
Ignavibacteria
Gemmatimonadetes
Gammaproteobacteria
Fibrobacteres acidobacteria
Deltaproteobacteria
Cyanobacteria
CP Ziwabacteria
CP WWE1
CP WOR−3
CP WOR−2 Omnitrophica
CP Shapirobacteria
CP Roizmanbacteria
CP Poribacteria
CP Nealsonbacteria
CP Moranbacteria
CP Handelsmanbacteria
CP Gottesmanbacteria
CP Fraserbacteria
CP Eisenbacteria
CP BRC1
CP Aminicemantes (OP8)
Chloroflexi
Chlorobi
Chlamydiae
Calditrichaeota
Betaproteobacteria
Bacteroidetes
Armatimonadetes
Alphaproteobacteria non LD12
Actinobacteria acSTL
Actinobacteria
Acidobacteria
FeFeHydrogenase
FeFeHydrogenase_2
Hydrogenase_Group_1
Hydrogenase_Group_2a
Hydrogenase_Group_2b
Hydrogenase_Group_3b
Hydrogenase_Group_3c
Hydrogenase_Group_3d
Hydrogenase_Group_4
Genes
Nb.of.genes
4
3
2
1

Figure S7 D

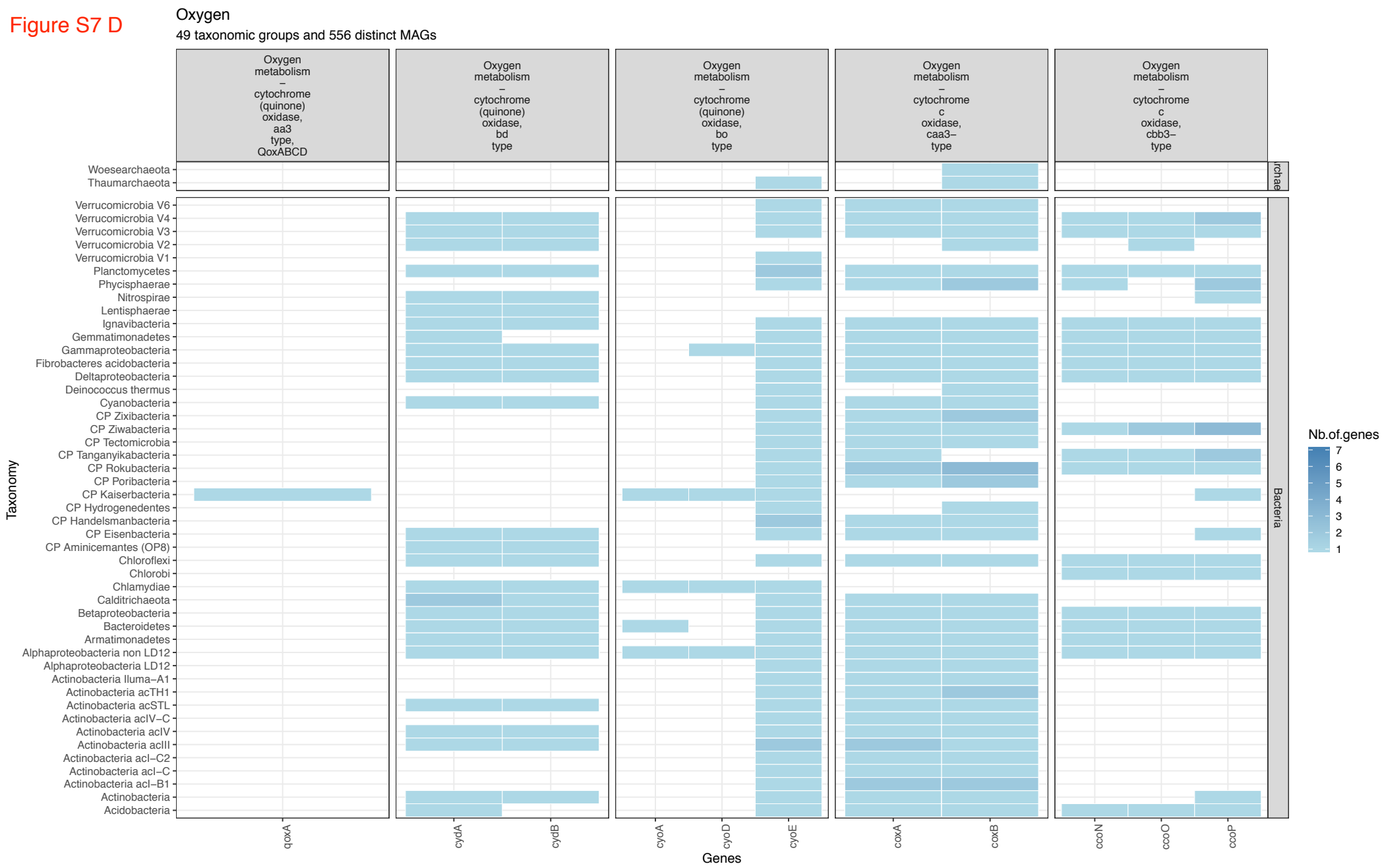

Oxygen
49 taxonomic groups and 556 distinct MAGs
Oxygen metabolism – cytochrome (quinone) oxidase, aa3 type, QoxABCD
Oxygen metabolism – cytochrome (quinone) oxidase, bd type
Oxygen metabolism – cytochrome (quinone) oxidase, bo type
Oxygen metabolism – cytochrome c oxidase, caa3– type
Oxygen metabolism – cytochrome c oxidase, cbb3– type
Woesearchaeota
Thaumarchaeota
Verrucomicrobia V6
Verrucomicrobia V4
Verrucomicrobia V3
Verrucomicrobia V2
Verrucomicrobia V1
Planctomycetes
Phycisphaerae
Nitrospirae
Lentisphaerae
Ignavibacteria
Gemmatimonadetes
Gammaproteobacteria
Fibrobacteres acidobacteria
Deltaproteobacteria
Deinococcus thermus
Cyanobacteria
CP Zixibacteria
CP Ziwabacteria
CP Tectomicrobia
CP Tanganyikabacteria
CP Rokubacteria
CP Poribacteria
CP Kaiserbacteria
CP Hydrogenedentes
CP Handelsmanbacteria
CP Eisenbacteria
CP Aminicemantes (OP8)
Chloroflexi
Chlorobi
Chlamydiae
Calditrichaeota
Betaproteobacteria
Bacteroidetes
Armatimonadetes
Alphaproteobacteria non LD12
Alphaproteobacteria LD12
Actinobacteria Iluma–A1
Actinobacteria acTH1
Actinobacteria acSTL
Actinobacteria acIV–C
Actinobacteria acIV
Actinobacteria acIII
Actinobacteria acI–C2
Actinobacteria acI–C
Actinobacteria acI–B1
Actinobacteria
Acidobacteria
Taxonomy
Bacteria
qoxA
cydA
cydB
cyoA
cyoD
cyoE
coxA
coxB
ccoN
ccoO
ccoP
Genes
Nb.of.genes
7
6
5
4
3
2
1

Figure S7 E

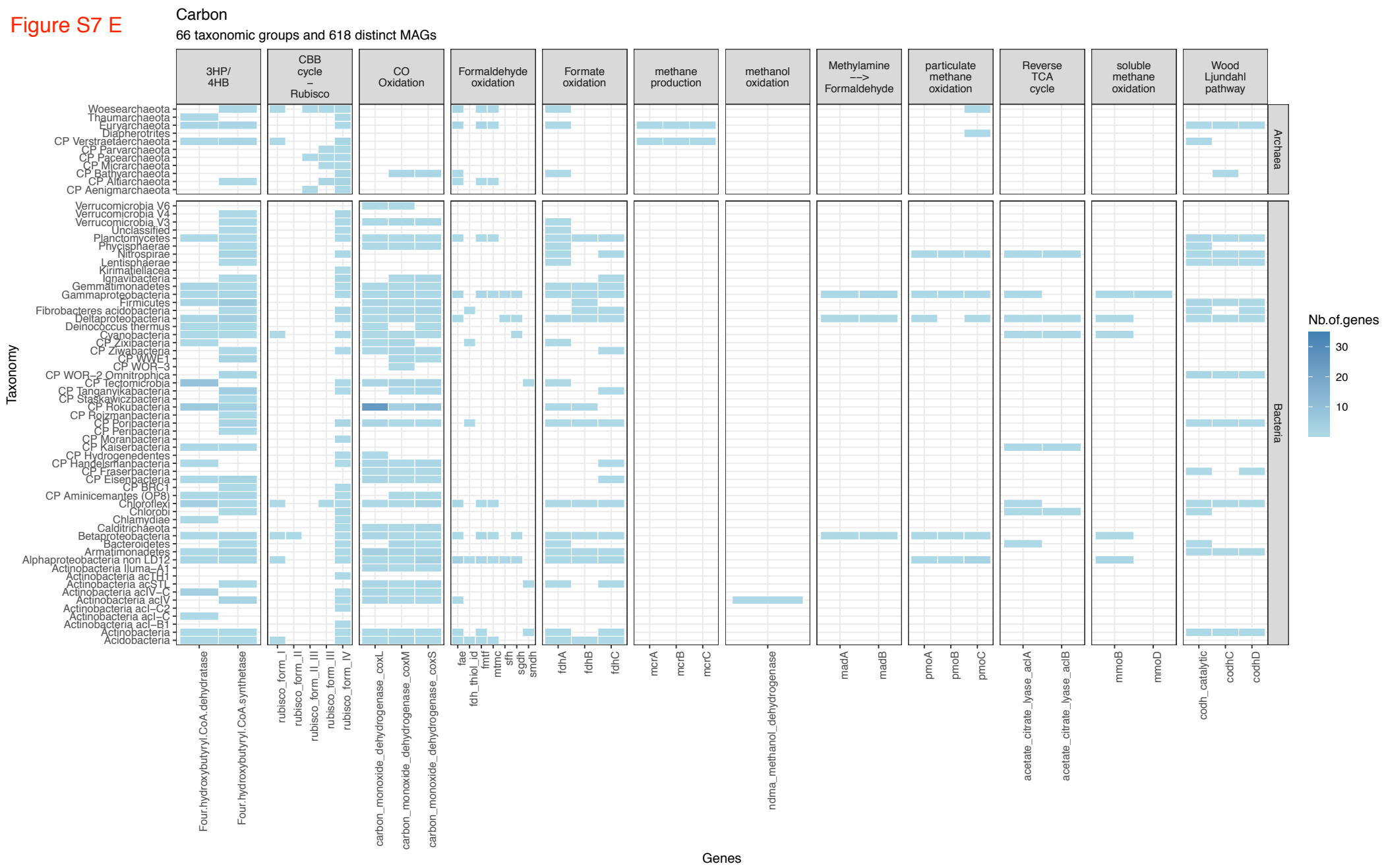
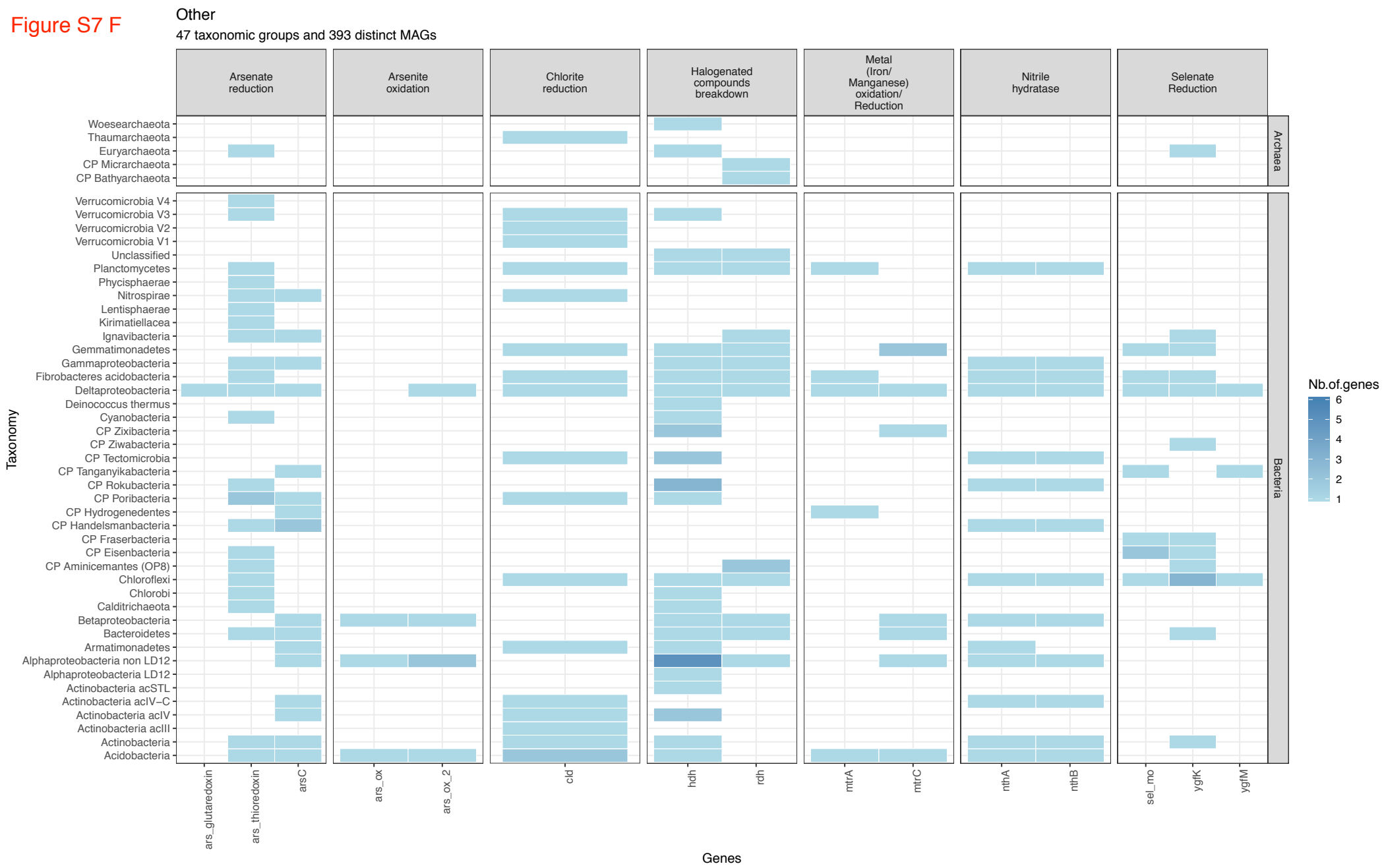

Carbon
66 taxonomic groups and 618 distinct MAGs
3HP/ 4HB
CBB cycle – Rubisco
CO Oxidation
Formaldehyde oxidation
Formate oxidation
methane production
methanol oxidation
Methylamine --> Formaldehyde
particulate methane oxidation
Reverse TCA cycle
soluble methane oxidation
Wood Ljundahl pathway
Archaea
Bacteria
Taxonomy
Woesearchaeota
Thaumarchaeota
Euryarchaeota
Diapherotrites
CP Verstraetaerchaeota
CP Parvarchaeota
CP Pacearchaeota
CP Micrarchaeota
CP Bathyarchaeota
CP Altiarchaeota
CP Aenigmarchaeota
Verrucomicrobia V6
Verrucomicrobia V4
Verrucomicrobia V3
Unclassified
Planctomycetes
Phycisphaerae
Nitrospirae
Lentisphaerae
Kirimatiellacea
Ignavibacteria
Gemmatimonadetes
Gammaproteobacteria
Firmicutes
Fibrobacteres acidobacteria
Deltaproteobacteria
Deinococcus thermus
Cyanobacteria
CP Zixibacteria
CP Ziwabacteria
CP WWE1
CP WOR-3
CP WOR-2 Omnitrophica
CP Tectomicrobia
CP Tanganyikabacteria
CP Staskawiczbacteria
CP Rokubacteria
CP Roizmanbacteria
CP Poribacteria
CP Peribacteria
CP Moranbacteria
CP Kaiserbacteria
CP Hydrogenedentes
CP Handelsmanbacteria
CP Fraserbacteria
CP Eisenbacteria
CP BHC1
CP Aminicemantes (OP8)
Chloroflexi
Chlorobi
Chlamydiae
Calditrichaeota
Betaproteobacteria
Bacteroidetes
Armatimonadetes
Alphaproteobacteria non LD12
Actinobacteria Iluma-A1
Actinobacteria acTH1
Actinobacteria acST1
Actinobacteria acIV-C
Actinobacteria acIV
Actinobacteria acI-C2
Actinobacteria acI-C
Actinobacteria acI-B1
Actinobacteria
Acidobacteria
Four.hydroxybutyryl.CoA.dehydratase
Four.hydroxybutyryl.CoA.synthetase
rubisco_form_I
rubisco_form_II
rubisco_form_II_III
rubisco_form_III
rubisco_form_IV
carbon_monoxide_dehydrogenase_coxL
carbon_monoxide_dehydrogenase_coxM
carbon_monoxide_dehydrogenase_coxS
fae
fdh_thiol_id
fmtf
mtmc
sfh
sgdh
smdh
fdhA
fdhB
fdhC
mcrA
mcrB
mcrC
ndma_methanol_dehydrogenase
madA
madB
pmoA
pmoB
pmoC
acetate_citrate_lyase_aclA
acetate_citrate_lyase_aclB
mmoB
mmoD
codh_catalytic
codhC
codhD
Genes
Nb.of.genes
30
20
10

Figure S7 F


Other
47 taxonomic groups and 393 distinct MAGs
Arsenate reduction
Arsenite oxidation
Chlorite reduction
Halogenated compounds breakdown
Metal (Iron/ Manganese) oxidation/ Reduction
Nitrile hydratase
Selenate Reduction
Archaea
Bacteria
Woesearchaeota
Thaumarchaeota
Euryarchaeota
CP Micrarchaeota
CP Bathyarchaeota
Verrucomicrobia V4
Verrucomicrobia V3
Verrucomicrobia V2
Verrucomicrobia V1
Unclassified
Planctomycetes
Phycisphaerae
Nitrospirae
Lentisphaerae
Kirimatiellacea
Ignavibacteria
Gemmatimonadetes
Gammaproteobacteria
Fibrobacteres acidobacteria
Deltaproteobacteria
Deinococcus thermus
Cyanobacteria
CP Zixibacteria
CP Ziwabacteria
CP Tectomicrobia
CP Tanganyikabacteria
CP Rokubacteria
CP Poribacteria
CP Hydrogenedentes
CP Handelsmanbacteria
CP Fraserbacteria
CP Eisenbacteria
CP Aminicemantes (OP8)
Chloroflexi
Chlorobi
Calditrichaeota
Betaproteobacteria
Bacteroidetes
Armatimonadetes
Alphaproteobacteria non LD12
Alphaproteobacteria LD12
Actinobacteria acSTL
Actinobacteria acIV−C
Actinobacteria acIV
Actinobacteria acIII
Actinobacteria
Acidobacteria
Taxonomy
ars_glutaredoxin
ars_thioredoxin
arsC
ars_ox
ars_ox_2
cld
hdh
rdh
mtrA
mtrC
nthA
nthB
sel_mo
ygfK
ygfM
Genes
Nb.of.genes
6
5
4
3
2
1

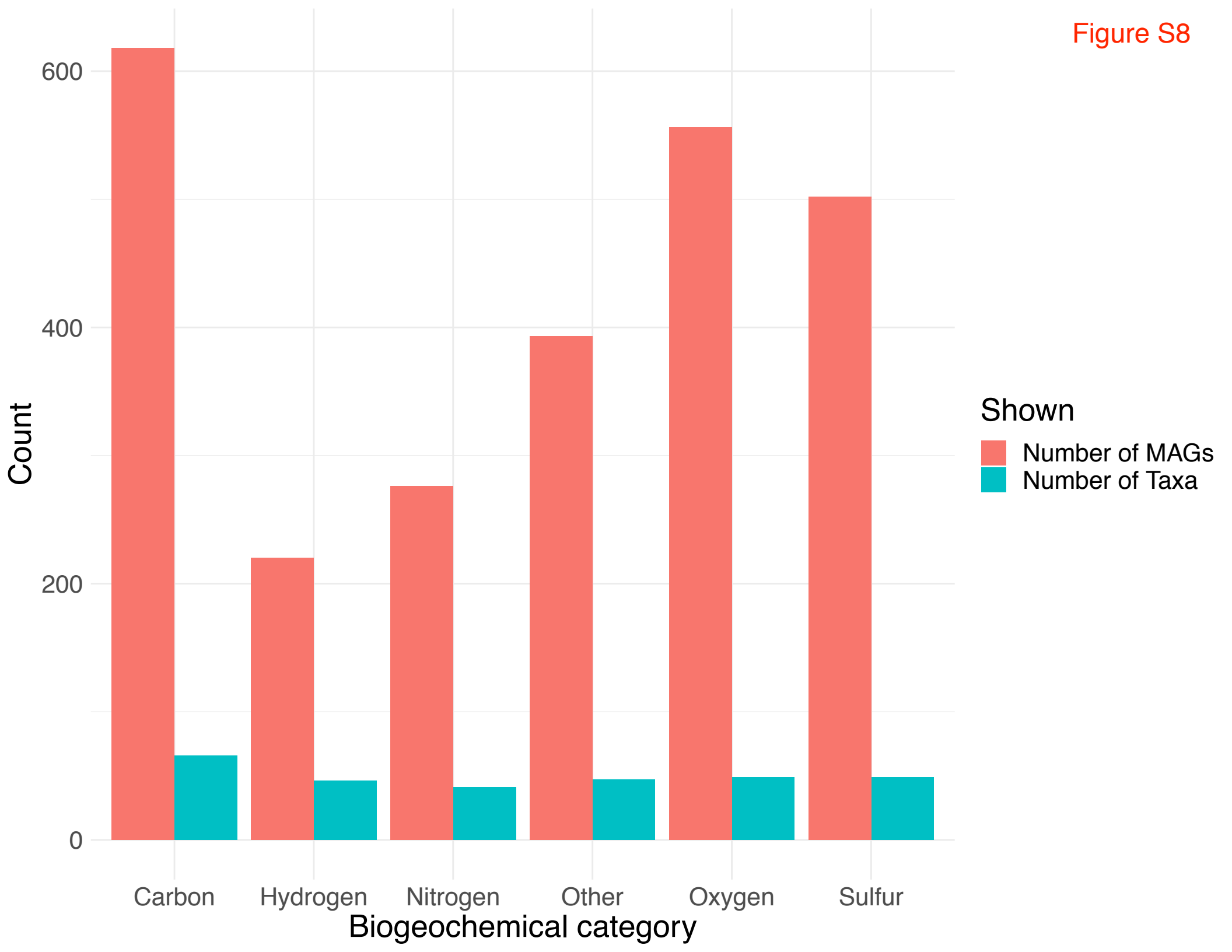


Figure S8
600
400
200
0
Count
Carbon
Hydrogen
Nitrogen
Other
Oxygen
Sulfur
Biogeochemical category
Shown
Number of MAGs
Number of Taxa

Figure S9

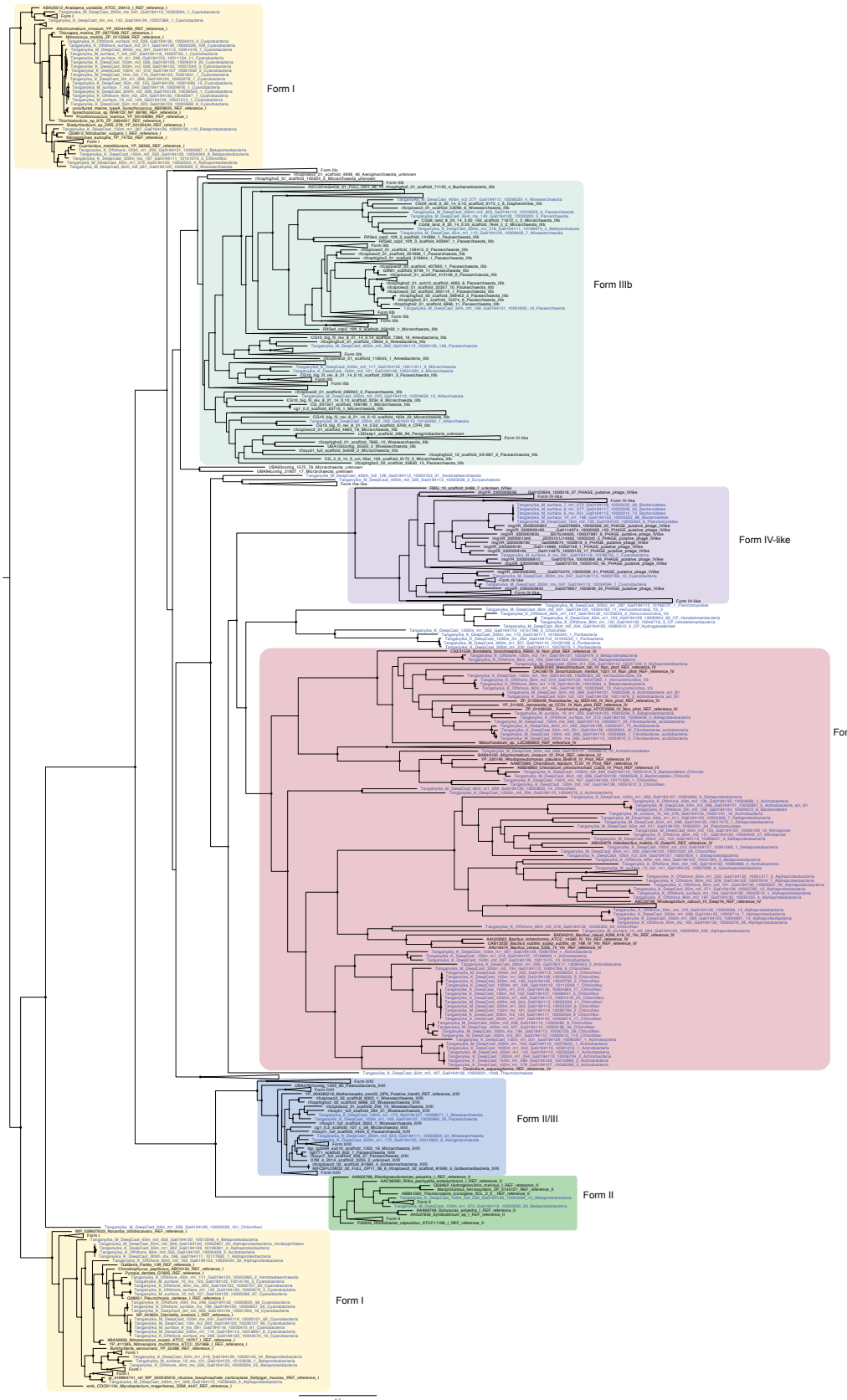


Figure S10

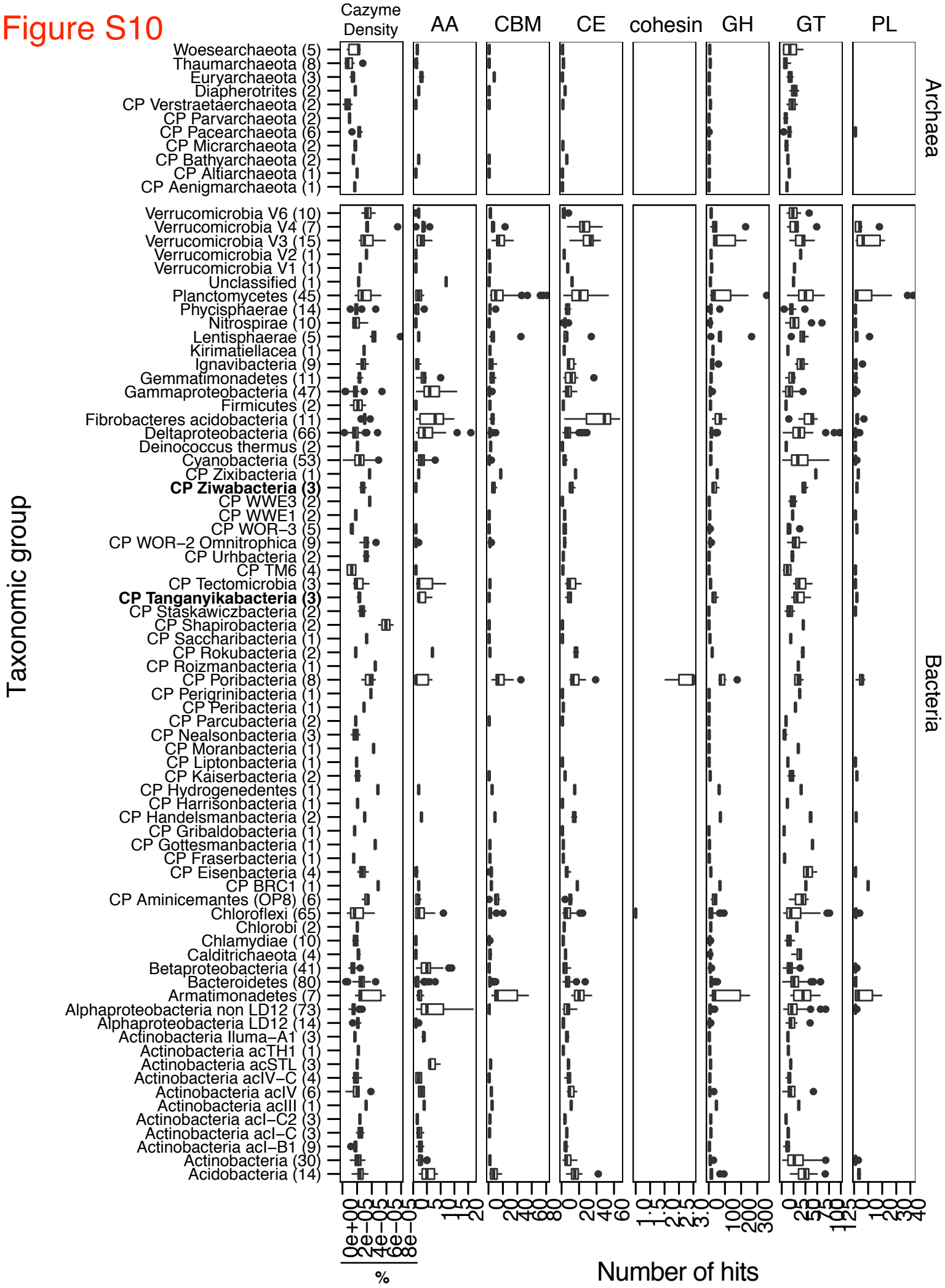

Cazyme Density
AA
CBM
CE
cohesin
GH
GT
PL
Archaea
Bacteria
Taxonomic group
Woesearchaeota (5)
Thaumarchaeota (8)
Euryarchaeota (3)
Diapherotrites (2)
CP Verstraetaerchaeota (2)
CP Parvarchaeota (2)
CP Pacearchaeota (6)
CP Micrarchaeota (2)
CP Bathyarchaeota (2)
CP Altiarchaeota (1)
CP Aenigmarchaeota (1)
Verrucomicrobia V6 (10)
Verrucomicrobia V4 (7)
Verrucomicrobia V3 (15)
Verrucomicrobia V2 (1)
Verrucomicrobia V1 (1)
Unclassified (1)
Planctomycetes (45)
Phycisphaerae (14)
Nitrospirae (10)
Lentisphaerae (5)
Kirimatiellacea (1)
Ignavibacteria (9)
Gemmatimonadetes (11)
Gammaproteobacteria (47)
Firmicutes (2)
Fibrobacteres acidobacteria (11)
Deltaproteobacteria (66)
Deinococcus thermus (2)
Cyanobacteria (53)
CP Zixibacteria (1)
CP Ziwabacteria (3)
CP WWE3 (2)
CP WWE1 (2)
CP WOR−3 (5)
CP WOR−2 Omnitrophica (9)
CP Urhbacteria (2)
CP TM6 (4)
CP Tectomicrobia (3)
CP Tanganyikabacteria (3)
CP Staskawiczbacteria (2)
CP Shapirobacteria (2)
CP Saccharibacteria (1)
CP Rokubacteria (2)
CP Roizmanbacteria (1)
CP Poribacteria (8)
CP Perigrinibacteria (1)
CP Peribacteria (1)
CP Parcubacteria (2)
CP Nealsonbacteria (3)
CP Moranbacteria (1)
CP Liptonbacteria (1)
CP Kaiserbacteria (2)
CP Hydrogenedentes (1)
CP Harrisonbacteria (1)
CP Handelsmanbacteria (2)
CP Gribaldobacteria (1)
CP Gottesmanbacteria (1)
CP Fraserbacteria (1)
CP Eisenbacteria (4)
CP BRC1 (1)
CP Aminicemantes (OP8) (6)
Chloroflexi (65)
Chlorobi (2)
Chlamydiae (10)
Calditrichaeota (4)
Betaproteobacteria (41)
Bacteroidetes (80)
Armatimonadetes (7)
Alphaproteobacteria non LD12 (73)
Alphaproteobacteria LD12 (14)
Actinobacteria Iluma−A1 (3)
Actinobacteria acTH1 (1)
Actinobacteria acSTL (3)
Actinobacteria acIV−C (4)
Actinobacteria acIV (6)
Actinobacteria acIII (1)
Actinobacteria acI−C2 (3)
Actinobacteria acI−C (3)
Actinobacteria acI−B1 (9)
Actinobacteria (30)
Acidobacteria (14)
%
Number of hits

Figure S11

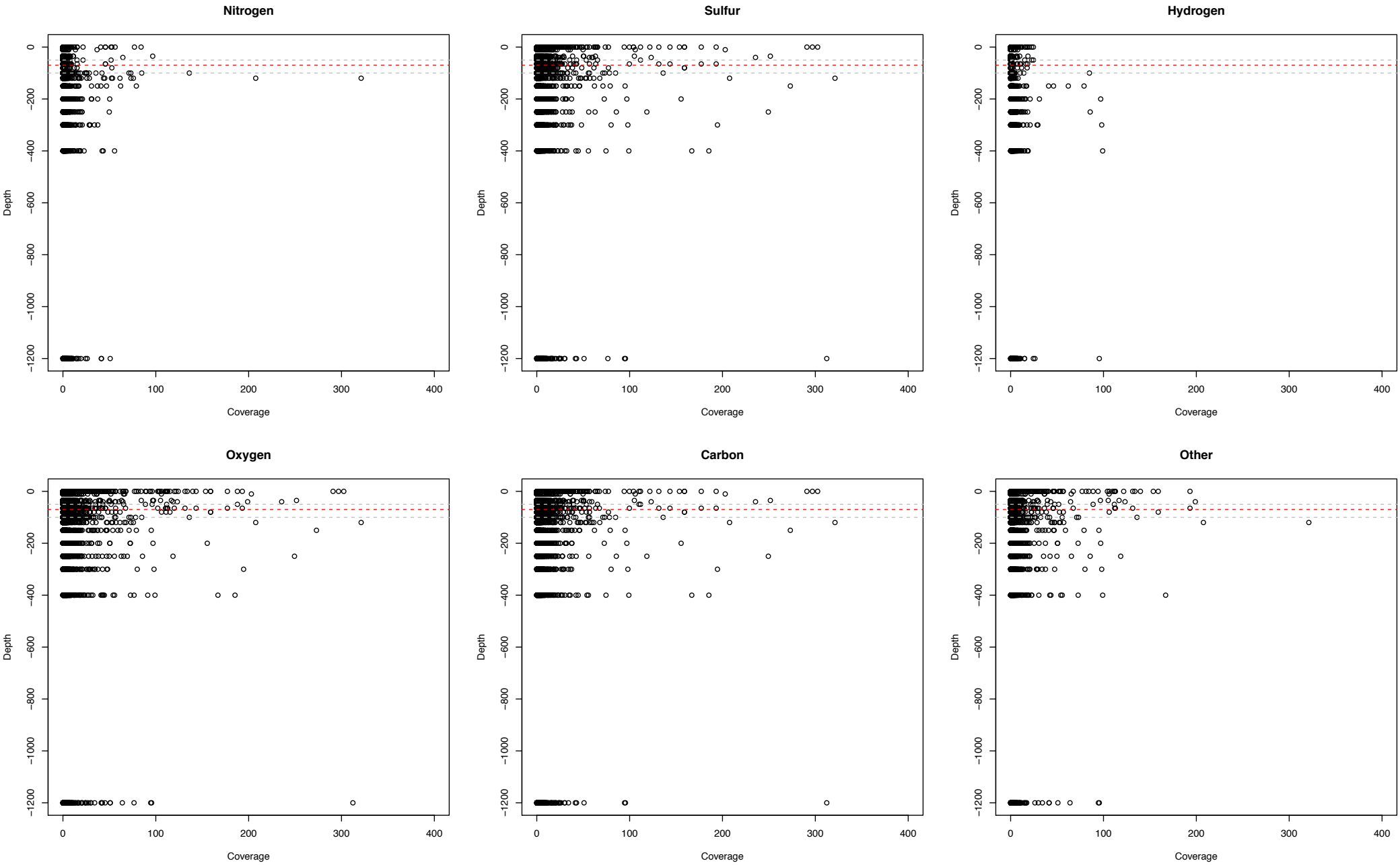
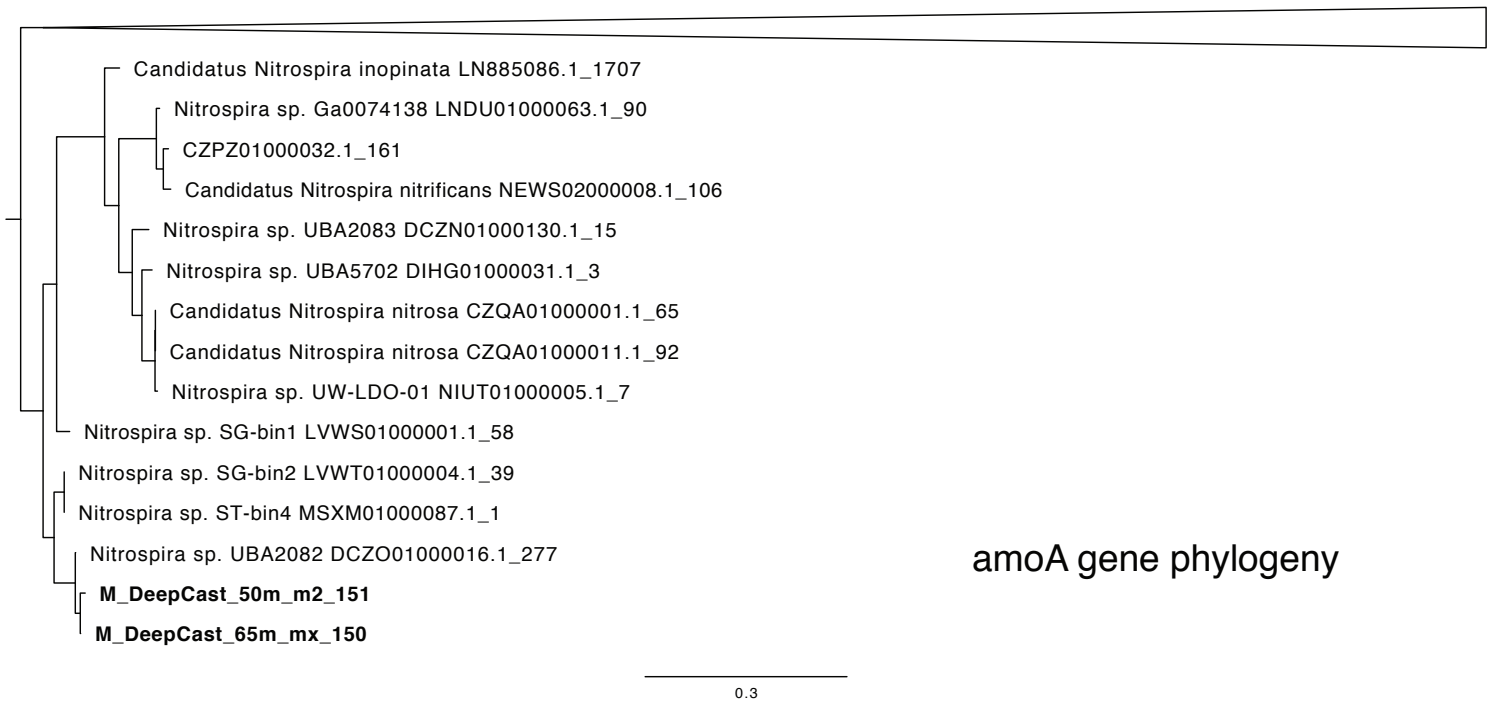

Nitrogen
Sulfur
Hydrogen
Oxygen
Carbon
Other
Depth
Coverage
0
−200
−400
−600
−800
−1000
−1200
0
100
200
300
400

Figure S12



amoA gene phylogeny