Cultivation and genomics of the first freshwater SAR11 (LD12) isolate

Michael W. Henson[1], V. Celeste Lanclos[1], Brant C. Faircloth[1,2], and J. Cameron Thrash[1,3]

1. Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, U.S.A.

2. Museum of Natural History, Louisiana State University, Baton Rouge, LA 70803, U.S.A.

3. Materials and Correspondence:
J. Cameron Thrash
Department of Biological Sciences
202 Life Sciences Bldg.
Louisiana State University
Baton Rouge, LA 70803
thrashc@lsu.edu
225-578-8210

Key Words: LD12, SAR11, freshwater microbiology, genomics, physiology

Abstract

Evolutionary transitions between fresh and salt water happen infrequently among bacterioplankton. Within the ubiquitous and highly abundant heterotrophic Alphaproteobacteria order *Pelagibacterales* (SAR11), most members live in marine habitats but the LD12 subclade has evolved as a unique freshwater lineage. LD12 cells occur as some of the most dominant freshwater bacterioplankton, yet this group has remained elusive to cultivation, hampering a more thorough understanding of its biology. Here we report the first successful isolation of an LD12 representative, strain LSUCC0530, using high throughput dilution to extinction cultivation methods, and its complete genome sequence. Growth experiments corroborate ecological data indicating active populations of LD12 in brackish water up to salinities of ~5. LSUCC0530 has the smallest closed genome thus far reported for a SAR11 strain (1.16 Mbp). The genome affirms many previous metabolic predictions from cultivation-independent analyses, like a complete Embden-Meyerhof-Parnas glycolysis pathway, but also provides novel insights, such as the first isocitrate dehydrogenase in LD12, a likely homologous recombination of malate synthase from outside of the SAR11 clade, and analogous substitutions of ion transporters with others that occur throughout the rest of the SAR11 clade. Growth data support metagenomic recruitment results suggesting temperature-based ecotype diversification within LD12, and key gene losses for osmolyte uptake provide a succinct hypothesis for the evolutionary transition of LD12 from salt to fresh water. For strain LSUCC0530 we propose the provisional nomenclature *Candidatus* Fonsibacter ubiquis.

Introduction

Bacterioplankton in the SAR11 clade of *Alphaproteobacteria* are dominant heterotrophs in both marine and freshwater systems. In the oceans, SAR11 can represent 25-50% of total planktonic cells[1,2]. Numerous subclades with unique spatio-temporal distributions comprise SAR11[1,3-5]. At least nine subclades defined via 16S rRNA gene sequences occupy marine niches[4], and more likely exist[6]. However, in spite of its global distribution[1], massive predicted population size of 10^{28} cells[1], and an estimated divergence time from the last common ancestor of 1.1 billion years ago[7], the bulk of existing evidence suggests that SAR11 has only successfully colonized freshwater environments once in its natural history[8-10]. All known freshwater SAR11 belong to subclade IIIb, a.k.a. LD12, although a very recent report challenges this assertion[11]. Regardless, the limited evolutionary diversification into less saline habitats has not prevented LD12 from achieving ecosystem prominence where it lives. In many lotic and lentic environments, LD12 occupies similar relative abundances as its marine cousins[9,12-15]. Study of LD12 is important to understand SAR11 evolution specifically, and how successful transitions between marine and freshwater environments occur in bacterioplankton[16] more generally.

Ecological, functional, and sequence-based inference from single amplified genomes (SAGs) and metagenomes support the hypothesis that LD12 bacterioplankton evolved from a genome-streamlined marine ancestor[7,9,17,18]. As such they share many of the same characteristics as marine SAR11, such as small cell volumes; adaptation to oligotrophic habitats; small, streamlined genomes; an obligate aerobic chemoorganoheterotrophic lifestyle with limited metabolic flexibility; preference for small molecular weight compounds like carboxylic and amino acids as carbon/energy sources; and auxotrophies for some amino acids and vitamins[5,9,10,15,17-24]. Previous research suggests that where LD12 differ from their marine counterparts lies in specific elements of metabolic potential that suggest a greater emphasis on production, rather than uptake, of osmolytes, and important metabolic changes related to energy production[15,17]. For example, metagenomic population data showed a correlation between decreased salinity and greater proportion of the Embden-Meyerhof-Parnass (EMP) vs. Entner-Doudoroff (ED) glycolysis pathways[15]. Comparative genomic analyses of SAGs from different SAR11 strains concurred: LD12 genomes contained the EMP pathway which is not found in most marine SAR11[17,19]. SAGs also suggested that LD12 lacks the glyoxylate shunt and some single carbon (C1) metabolism[17].

In spite of what has been learned from cultivation-independent methods, the lack of cultured LD12 representatives has hampered a more detailed exploration of the group. Potential ecotypes within LD12 have been identified[18], but we cannot delineate what distinguishes them without better physiological and genomic data. Interpreting data on the ecological distribution of LD12 remains challenging without information on growth tolerances and optima for salinity and temperature. We also don't understand whether a connection exists between more efficient energy production through glycolysis and the

84 freshwater lifestyle, or what other adaptations explain LD12 evolution away from salt
85 water.
86 The next steps in translating 'omics-based predictions into measured data for
87 integration with ecosystem models require living experimental subjects. For example,
88 cultures of marine SAR11, such as HTCC1062 and HTCC7211, have facilitated testing of
89 metabolism and growth[21-23,25-29], structural organization[30], and virus-host interactions[31].
90 We need cultivated representatives to provide this kind of understanding of other
91 important bacterioplankton like LD12. In service of this goal, we pursued a systematic high
92 throughput cultivation effort from coastal regions in the northern Gulf of Mexico. Here we
93 report the first isolation of a member of the LD12 clade, strain LSUCC0530, its complete
94 genome sequence, preliminary physiological data, and an examination of the ecological
95 distribution of this organism compared with other LD12 clade members. Our results
96 provide evidence for temperature-dependent ecotype diversification within LD12, and a
97 hypothesis about evolutionary process that led to the unique SAR11 colonization of fresh
98 water.
99
100 **Results**
101 *Isolation and growth of LSUCC0530*
102 We isolated strain LSUCC0530 as part of a series of high throughput dilution-to-extinction
103 cultivation experiments (HTC) conducted with inoculation water obtained across the
104 southern Louisiana coast. The particular experiment that yielded LSUCC0530 utilized
105 surface water from the coastal lagoon of Lake Borgne, 39 km southeast of New Orleans (Fig.
106 S1). At the time of collection, the surface water had a salinity of 2.39 and was 30.5°C. Later
107 community analysis demonstrated that LD12-type organisms represented 8.7% of the
108 bacterioplankton population in the inoculum (see below). We diluted Lake Borgne surface
109 water for inoculation to an estimated 2 cells well^{-1} into our JW5 medium (salinity 1.45,
110 Table S1), and incubated cultures for 30 days at 23°C (room temperature in the Thrash
111 Lab). LSUCC0530 grew slowly, reaching 8.6 x 10^5 cells mL^{-1} after 24 days. It had a low
112 fluorescence and low side scatter flow cytometry signature characteristic of other SAR11
113 cells (Fig. S1).
114
115 *Taxonomy and morphology*
116 The LSUCC0530 16S rRNA gene sequence had > 99% identity with eight sequences from
117 members of LD12, and only 91% and 92% sequence identity with sequences from the
118 sister subclade IIIa representatives IMCC9063 and HIMB114, respectively. LSUCC0530 also
119 had average amino acid identities (AAI) of 88.7% and 89.8% with HIMB114 and IMCC9063,
120 respectively, corroborating LSUCC0530 as belonging to separate species from subclade
121 IIIa[32]. Phylogenetic inference using 16S rRNA genes placed LSUCC0530 within the Order
122 *Pelagibacteriales* (a.k.a. SAR11[5]), in a monophyletic group with the original LD12 clone
123 library sequence (Fig. S2). Phylogenomic inference using 83 single copy marker genes from

124 49 SAR11 genomes also strongly supported placement of LSUCC0530 within LD12 as an
125 early diverging member with only the SAG AAA280-B11 from Lake Damariscotta diverging
126 earlier (Fig. 1A). This tree also recovered the three microclusters within LD12 (A, B, C)
127 observed previously[18] (Fig. 1A). Cells of strain LSUCC0530 were small, curved rods,
128 approximately 1 μm x 0.1 μm (Fig. 1B, C), showing conserved morphology with the
129 distantly related SAR11 strain HTCC1062 in subclade Ia[30,33].
130
131 *Genome characteristics*
132 Genome assembly resulted in a single, circularized scaffold that we tested for completeness
133 using multiple tools (Supplemental Information). The completed genome is 1,160,202 bp
134 with a GC content of 29.02% and 1271 predicted genes (Fig. 2). There are 1231 putative
135 protein coding genes and 40 RNA genes, including one each of the 5S, 16S, and 23S rRNA
136 genes. We found no genomic evidence of lysogenic bacteriophage or a CRISPR-*cas* system.
137 LSUCC0530 has the characteristics of genome streamlining previously reported for other
138 SAR11 genomes, with an estimated coding percentage of 96.36% and the smallest complete
139 SAR11 genome reported thus far[19,20]. However, in spite of its unusually small overall size,
140 intergenic spacer regions are not significantly smaller than those of the sister subclade IIIa
141 (Fig. S3). All LD12 genomes had much higher AAI and synteny with each other than with
142 the subclade IIIa (Table S1), mirroring their phylogenomic relationships (Fig. 1A).
143 However, within the LD12 clade, AAI and synteny percentages did not clearly delineate the
144 same patterns seen in the phylogenomic tree, which may have resulted from the LD12 SAGs
145 being fragmented and incomplete.
146 The LSUCC0530 genome has the conserved architecture for hypervariable region
147 (HVR) 2, observed in most other complete SAR11 genomes[19,34,35] and other LD12 SAGs[18],
148 between the 16S rRNA-tRNA-tRNA-23S rRNA operon and the 5S rRNA gene operon. This
149 region is 54,755 bp and has a clear variation in GC content from neighboring parts of the
150 genome (Fig. 2). Metagenomic recruitment also verified this region as hypervariable based
151 on the lack of matching sequences in community data (Fig. S4). The predicted gene
152 annotations in HVR2 resemble those in the same location in other SAR11 genomes[18,19],
153 namely genes likely functioning in membrane modification such as glycosyltransferases,
154 epimerases, transaminases, and methyltransferases; as well as hypothetical proteins.
155
156 *Metabolic reconstruction*
157 Like most other SAR11s[5,17,19], the LSUCC0530 genome encodes for an obligate aerobic
158 lifestyle, with a complete oxidative phosphorylation pathway including a heme/copper-
159 type cytochrome c oxidase, a proton-translocating NADH dehydrogenase, and a proton-
160 translocating ATP synthase (Fig. 3). As previously detected in LD12 SAGs[17], the LSUCC0530
161 genome encodes a complete EMP glycolysis pathway, including phosphofructokinase and
162 pyruvate dehydrogenase, and complete gluconeogenesis through nucleotide sugar
163 production (Fig. 3). We recovered genes for the pentose phosphate pathway and a

164 predicted fructokinase gene for conversion of D-fructose to ß-D-fructose-6-P. As expected
165 based on other SAR11 genomes, we did not find genes for utilization of allose, fucose,
166 fuculose, galactose, maltose, mannose, rhamnose, sucrose, starch, trehalose, or xylose.
167 LSUCC0530 has a complete TCA cycle and glycolate oxidase, and like other SAR11s, the
168 non-mevalonate pathway for isopentenyl diphosphate (Fig. 3).
169 We recovered the first isocitrate lyase in an LD12 genome, conferring a complete
170 glyoxylate bypass. The isocitrate lyase shares its ancestry with the majority of the SAR11
171 clade except subclade IIIa, where it appears to have been lost[17,19] (Fig. S5). However, the
172 LSUCC0530 malate synthase shared with other LD12 SAGs belongs to isoform A, whereas
173 all other SAR11 clades contain isoform G[36] (Figs. 3; S6). Comparison of gene neighborhoods
174 in a variety of SAR11 genomes revealed that both isoforms occur in the same location (Fig.
175 S7).
176 We predict production of eighteen amino acids (Fig. 3) and partial synthesis
177 pathways for three more (histidine, cysteine, and methionine), as well as the ability to
178 interconvert, but not synthesize *de novo*, phenylalanine and tyrosine. The LSUCC0530
179 genome has lost some of the C1 and methylated compound pathways found in other
180 SAR11s[24]. For example, as in the LD12 SAGs[17], LSUCC0530 does not have DMSP,
181 methylamine, or glycine-betaine metabolism. However, it retains tetrahydrafolate
182 metabolism, formate oxidation, and the glycine cleavage pathway (Fig. 3).
183 The genome has the same PII nitrogen response system and *amtB* ammonium
184 transporter found in other SAR11s, and is predicted to utilize ammonia as a nitrogen
185 source in amino acid synthesis (Fig. 3). The genome also encodes a putative thiosulfate/3-
186 mercaptopyruvate sulfurtransferase that may be involved in sulfur and thiocyanate
187 reactions (Fig. 3). A possible product of this enzyme is sulfite, which could be further
188 oxdized to sulfate in the periplasm via putative *sorAB* genes. This gene pair occurs
189 infrequently in the SAR11 clade, but sulfite oxidation may serve in some strains as a means
190 to generate additional proton motive force[37,38], similarly to proteorhodopsin[21,29,39], which
191 the LSUCC0530 genome also contains. Also like other SAR11s, a complete assimilatory
192 sulfate reduction pathway is missing, making LSUCC0530 dependent on reduced sulfur
193 compounds for cellular sulfur requirements[22]. In our media, this was supplied in the form
194 of sulfur-containing amino acids (Table S1). The genome encodes partial pathways for
195 incorporation of sulfide into cysteine and methionine, but is missing predicted serine O-
196 acetyltransferase and homoserine O-succinyltransferase genes (Fig. 3).
197 In addition to the PII nitrogen sensor, LSUCC0530 also has putative two component
198 systems handling phosphate limitation, pH, osmotic upshift via potassium, and redox state
199 (Fig. 3). The LSUCC genome encodes ABC transporters for phosphate (*ptsABCS*),
200 lipoprotein, as well as branched-chain and L amino acid uptake. We identified tripartite
201 ATP-independent periplasmic (TRAP) transporter genes for C4-dicarboxylate and
202 mannitol/chloroaromatic compounds. However, the gene cluster appears to be missing the
203 substrate binding subunit, and we did not find predicted genes for mannitol utilization.

204 In general, LSUCC0530 and the LD12 SAGs share many of the same pathways for
205 osmolyte synthesis as the subclade IIIa and other SAR11 organisms, such as N-acetyl-
206 ornithine, glutamate, glutamine, and proline (Table 1). In some cases, there appear to be
207 analogous substitutions in osmolyte transport systems. For example, LD12 genomes have
208 potassium antiporters, but not the *trkAH* genes for potassium transport in other SAR11s.
209 Similarly, a sodium/proton antiporter orthologus cluster comprised genes exclusively from
210 LD12 genomes, but a separate sodium/proton antiporter occurred in all other SAR11
211 genomes (Table 1). Perhaps the most important difference in genomic content concerning
212 osmolytes occurs in two key transporter losses: LSUCC0530 and the other LD12s have no
213 genes for the glycine betaine/proline or ectoine/hydroxyectoine ABC transporters that are
214 present in HIMB114 and IMCC9063 and other subclades of SAR11 (Table 1).
215
216 *Physiology*
217 Under isolation conditions (23°C, JW5 medium), strain LSUCC0530 grew to a density of 2 x
218 10^7 cells mL^{-1} with an average growth rate of 0.53 day^{-1}. Our physiological experiments to
219 determine the salinity tolerance and growth optima revealed that LSUCC0530 could grow
220 at salinities between 0.36 and 4.7 (Fig. 4A). The optimal salinity for LSUCC0530 was at or
221 below the isolation medium of 1.45. Growth rates decreased between salinities of 2.9 and
222 4.7 (Fig. 4A), and cells died (determined by rapid loss of flow cytometric signal) in salinities
223 above 8 (Fig. S8). We observed growth after a very extended lag phase in a subset of
224 cultures at salinity 5.8, but could not reproduce this growth in repeated experiments (Fig.
225 S8). Among the temperatures tested, LSUCC0530 grew optimally at 23˚C, and we observed
226 growth up to 30˚C, but not at 12˚C or below, nor at 35˚C or above (Fig. 4B). However, we
227 did not observe a clear loss of signal at 4, 12, 35, or even 40˚C (Fig. S8), even after over 33
228 days, suggesting that LSUCC0530 can endure extended periods at these temperatures in a
229 non-growth state. Notably during growth experiments, one of the five replicates during any
230 given experiment frequently grew at a reduced rate than the others (Fig. S8).
231
232 *Ecology*
233 We quantified the abundance of LD12 taxa in the coastal Louisiana ecosystems using 16S
234 rRNA gene sequences clustered into operational taxonomic units (OTUs). These sequences
235 were obtained as part of the combined sampling and HTC campaign that included the
236 successful isolation of LSUCC0530. Based on BLASTn sequence similarity, the
237 representative sequence for OTU7 and LSUCC0530 shared 100% identity over the 250bp
238 v4 region of the 16S rRNA gene, confirming the Mothur classification of OTU7 as belonging
239 to the SAR11 LD12 clade. Over the three year sampling period, OTU7 was the fifth most
240 abundant OTU based on total read abundance and sixth ranked OTU based on median read
241 abundance. At the time of sampling for the experiment that resulted in the isolation of
242 LSUCC0530, OTU7 had a relative abundance of 8.7% in Lake Borgne. In some sites, such as
243 the Atchafalaya River Delta (salinity 0.18), LD12 represented as much as 16% of total

244 bacterioplankton (Fig. S9). Generally, the LD12 OTU7 relative abundance correlated
245 negatively with salinity. At seven sites with salinities below 6, OTU7 had a relative
246 abundance of 4.0% or greater with a median relative abundance of 8.8% (Fig. S9). At the
247 remaining ten sites above a salinity of 6 (median 19.44), OTU7 had relative abundances
248 around 2% or lower (Fig. S9).
249 The phylogenetic separation of LSUCC0530 from B11 and the IIIb A-C microclusters
250 raised the question of whether LSUCC0530 represents a novel ecotype with unique
251 spatiotemporal distribution compared to other LD12 taxa. Therefore, we explored the
252 distribution of LD12 genomes in 90 fresh and brackish water environments through
253 competitive recruitment of metagnomic sequences to all the SAR11 genomes in the
254 phylogenomic tree and the outgroup sequence HIMB59. We quantified relative abundance
255 via recruited reads (per kpb of metagenomic sequence per Mbp of genome sequence-
256 RPKM), and aggregated LD12 data based on the A, B, and C microclusters (Fig. 1A), the B11
257 SAG, and LSUCC0530. Figure 5A displays RPKM values for reads with ≥ 95% identity,
258 roughly corresponding to a species delineation via average nucleotide identity (ANI)[40], but
259 additional comparisons at ≥ 85, 90, 92, 98, and 100% identity are provided in Figure S10.
260 All LD12 subgroups had their highest RPKM values in very low salinity sites,
261 although we could detect LD12 across various brackish salinities (Fig. 5A). In general, IIIb.A
262 and LSUCC0530 averaged the highest relative abundances across all sites with salinities
263 between 0 and 5.75. All subgroups showed site-specific variation in abundances (Fig. 5A),
264 indicating that the available genomes likely do represent specific ecotypes within LD12.
265 Microcluster IIIb.A clearly dominated recruitment at Lake Michigan, whereas LSUCC0530
266 represented the most abundant genome at Feitsui Reservoir and Lake Gatun (Fig. 5A).
267 By examining recruitment density to the different microclusters across a range of
268 percent identities, we also observed patterns indicating additional LD12 diversity beyond
269 that represented in the current genomes. For example, whereas the LSUCC0530 genome
270 recruited the majority of high percent identity reads in the Feitsui Reservoir and Lake
271 Gatun samples (Fig. 5B), in some locations the majority of recruitment to this genome
272 centered around lower percent identity hits (Fig. 5C). We observed the same kind of
273 alternative patterns for the other microclusters as well (Fig. 5D). Since reads were
274 competitively recruited against all close relative genomes, these cases where a
275 microcluster primarily recruited reads ~90-92% identity likely mean that we lack a
276 representative genome for that fraction of the metagenomic data.
277
278 **Discussion**
279 The successful cultivation of and LD12 strain has demonstrated for the first time that this
280 organism shares many physiological traits with cultured marine SAR11 representatives.
281 LSUCC0530 doubles roughly once every two days, similarly to taxa in subclade Ia[33] and
282 Ib[41]. It has an aerobic, chemoorganoheterotrophic lifestyle, thriving in oligotrophic media
283 with simple carbon compounds, and likely depends on reduced organosulfur compounds

284 and ammonium as its sulfur and nitrogen sources, respectively. Cell size and shape
285 demonstrated considerable morphological conservation (small, curved rods roughly 1 x 0.1
286 μm) across large evolutionary distances (LD12 vs. subclade Ia ~89% 16S rRNA gene
287 identity).
288 Cultivation of LSUCC0530 has also helped to constrain important physiological
289 differences between LD12 and marine SAR11. Although a systematic survey of salinity
290 tolerance has not been conducted for other cultured SAR11 representatives, we have
291 shown that LD12 cannot survive in media with salinities over eight, whereas all other
292 cultured SAR11 grow in seawater salinities[25,33,41-43]. LSUCC0530 thrives in relatively high
293 temperatures (23-30˚C), and we could not confirm growth at low temperatures where
294 other SAR11 have been isolated[42].
295 LSUCC0530 shares many genomic characteristics with marine SAR11, such as a
296 small, streamlined genome with few paralogs, short intergenic spaces, and a conserved
297 hypervariable region, confirming predictions from LD12 SAGs and metagenomes[12,17,18]. The
298 LSUCC0530 genome also encodes much of the same metabolic potential as what has been
299 reported through LD12 SAGs and metagenomic data. For example, we confirm the EMP
300 glycolysis pathway, TCA cycle, non-mevalonate terpenoid biosynthesis, numerous amino
301 acid synthesis pathways, TRAP transporters for C4 dicarboxylate compounds, and ABC
302 transporters for various amino acids. We can confirm the lack of a phosphoenolpyruvate
303 carboxylase (ppc) and missing pathways for DMSP, glycine-betaine, and methylamine
304 metabolism.
305 We have also recovered important genetic information previously unknown for
306 LD12. At 1.16 Mbp, LSUCC0530 has a smaller genome, with fewer genes, than any other
307 SAR11. LSUCC0530 has a complete glyoxylate bypass, which differs from previous
308 inferences using SAGs[17]. The unusual LD12 malate synthase isoform A gene occupies the
309 same genomic location in LSUCC0530 as the isoform G copy in most subclade Ia, Ic, and IIIa
310 genomes, with the exception of HIMB114 (Fig. S7). Therefore, the most parsimonious
311 explanation for the presence of isoform A in LSUCC0530 is a homologous recombination
312 event that replaced the isoform G homolog. The LSUCC0530 genome also contains putative
313 pathways for oxidation of sulfide, thiosulfate, and sulfite, which have not yet been explored
314 in any cultured SAR11.
315 The relative abundance data from coastal Louisiana, the isolation of an LD12 strain
316 from a coastal environment, and the ability of LSUCC0530 to grow in slightly brackish
317 water supports previous work indicating that LD12 occupies a much wider range of aquatic
318 habitats than inland freshwater systems[9,15,44,45]. However, metagenomic recruitment
319 provides supporting evidence that LD12 has diversified into multiple ecotypes with unique
320 distributions (Fig. 5A), and new evidence that in many environments, additional LD12
321 diversity exists beyond that found in current genomes (Fig. 5C-D). In addition, our data
322 suggests that LSUCC0530 likely represents a novel LD12 ecotype since it preferentially

323 recruits metagenomic sequences at high percent identity in different environments than
324 for the other microclusters of LD12 (Fig. 5).
325 What distinguishes the distribution of these ecotypes remains unclear. Small
326 variations in salinity optima for different ecotypes may drive diversification. However,
327 given the paucity of low brackish sites in available datasets, we cannot establish whether
328 salinity plays a role in ecotype diversification. Furthermore, without complete genomes
329 from other LD12 ecotypes, we cannot rule out whether alternative ionic strength
330 management strategies, or some other metabolic characteristic, help define evolutionary
331 differences associated with ecotypes. For example, the ability to utilize alternative sulfur
332 sources may distinguish LSUCC0530 from other LD12. The *sorAB* sulfite dehydrogenase
333 and the thiosulfate/3-mercaptopyruvate sulfurtransferase genes have uneven distribution
334 among SAR11 genomes, including within the LD12 subclade.
335 An alternative hypothesis is that temperature has driven diversification between
336 ecotypes. A striking feature of LSUCC0530 is its growth preference for relatively high
337 temperatures and inability to grow at low temperatures. While robust growth at 30˚C may
338 seem intuitive given that this represents typical Lake Borgne surface water temperature,
339 the inability to grow at 12˚C or below would likely place LSUCC0530-type LD12 cells at a
340 fitness disadvantage in places like Lake Michigan, where annual temperatures rarely get
341 above 22˚C and average temperatures remain below 15˚C for the majority of the year
342 (https://coastwatch.glerl.noaa.gov/). The metagenomic recruitment data support this
343 hypothesis: the samples in which LSUCC0530 dominated recruitment (Lake Gatun, Feitsui
344 Reservoir) were all above 19˚C. Conversely, other microclusters dominated recruitment of
345 samples from colder environments (e.g. IIIb.A in Lake Michigan)(Fig. 5). More sequence
346 data may strengthen the relationship between temperature and ecotypes, but since
347 temperature optima is extremely hard to infer from genomes alone, additional cultivars
348 will be needed to test the temperature hypothesis for ecotype diversification.
349 A critical question remains for LD12. What has made these cells uniquely adapted to
350 low salinity waters? Genomic analyses here and elsewhere have helped define the
351 differences in LD12 metabolic potential with that of other SAR11, but very few adaptations
352 appear related to salinity. It has been noted that the EMP pathway provides more ATP and
353 NADH than the ED pathway, thus conferring a greater energy conservation through
354 glycolysis for LD12 cells than other SAR11s[15,17]. If or how this energy conservation relates
355 to salinity adaptation has not been pursued. The presence of unique LD12 cation-proton
356 antiporters (Fig. 3) could provide a connection. LD12 cells may benefit from additional
357 NADH for the production of a proton motive force that drives the potassium-proton
358 antiporters, thus improving active transport to lower internal ionic strength. However,
359 while this adaptation may improve the ability of LD12 cells to live in fresh water, it doesn't
360 explain why they have lost the ability to survive in salt water where marine SAR11 cousins
361 thrive.

362 The most likely explanation comes in the distribution of genes for osmolyte uptake.
363 Critically, LSUCC0530 and LD12 in general have lost the *proVWX* glycine-betaine/proline
364 ABC transporter shared by all other marine SAR11 (Table 1). Although a relatively minimal
365 number of genes, this uptake capability likely represents a significant loss of function for
366 LD12 salinity tolerance. In *Escherichia coli*, the *proVWX* operon has a strong promoter that
367 upregulates these genes two orders of magnitude in response to increased salinity[46].
368 Indeed, rapid proline uptake is a common response to osmotic upshift in many bacteria[47].
369 Furthermore, LSUCC0530 and other LD12 genomes have lost the ABC transporters for
370 ectoine/hydroxyectoine present in the other SAR11 genomes. Given the slow growth of
371 LD12, rapid production of new intracellular osmolytes in response to increased salinity
372 seems unlikely. Thus, the loss of these transport systems for quickly equilibrating
373 intracellular osmolarity probably prevents successful dispersal of this organism into higher
374 salinity waters.
375 This finding allows us to speculate further about the evolutionary trajectory that led
376 to successful colonization of freshwater environments by LD12. All evidence suggests that
377 the last common SAR11 ancestor was a marine organism with a streamlined genome[7,8,18].
378 During further diversification into subclade III, additional genes and pathways were lost,
379 for example, ppc and genes for DMSP and some C1 compound metabolisms[17,48]. Therefore,
380 the common ancestor of subclade IIIa and LD12 would have had to adapt to freshwater
381 environments starting with an already reduced genome and limited metabolic flexibility. A
382 simple way to inhabit lower salinity environments would have been to decrease expression
383 of, or altogether lose, genes for uptake and/or production of osmolytes. If the adaptation
384 were reversible, we would expect to see examples of LD12 organisms growing in marine
385 systems. However, existing evidence suggests that while LD12 organisms can grow
386 (established herein) and be active[44] in low brackish conditions, they are not successful in
387 higher salinity brackish or marine water[8,9,15,45]. Thus, we propose that LD12 adaptation to
388 freshwater involved irreversible loss of function, decreasing osmotic response capability in
389 LD12 cells, and preventing recolonization of higher salinity environments. The missing ABC
390 transporters for proline/glycine betaine and ectoine/hydroxyectoine in LD12 are likely
391 candidates for this loss of function. Experimental tests on the differences in production and
392 uptake of osmolytes among subclade IIIa and LD12 representatives will also provide
393 critical insight into how these different evolutionary paths have resulted in unique salinity
394 ranges.
395
396 For the first cultured representative of the LD12 clade, we propose the provisional
397 taxonomic assignment for strain LSUCC0530 as '*Candidatus* Fonsibacter ubiquis',
398
399 **Description of *Fonsibacter* gen. nov.**
400 *Fonsibacter* (*Fons* L. noun; fresh water, spring water, -bacter, Gr. adj.; rod, bacterium.
401 Fonsibacter referring to the preferred low salinity habitat of this organism).

Aerobic, chemoorganoheterotrophic, and oligotrophic. Cells are small, curved rods roughly 1 x 0.1µm. Non-motile. Based on phylogenomics and 16S rRNA gene phylogenetics, subclade IIIb/LD12 occurs on a separate branch within the *Pelagibacteraceae* (SAR11), sister to subclade IIIa containing strains HIMB114 and IMCC9063. Due to the depth of branching between these clades using concatenated protein-coding genes, and the 92% and 91% 16S rRNA gene sequence identity with HIMB114 and IMCC9063, respectively, we propose LSUCC0530 and subclade IIIb/LD12 as a novel genus in the *Pelagibacteraceae*.

Description of *Fonsibacter ubiquis* sp. nov.

Fonsibacter ubiquis (L. adv. ubiquitous).

In addition to the properties given in the genus description, the species is described as follows. Growth occurs at temperatures between 23˚C and 30˚C, but not at 11˚C or below, nor at 35˚C or above. Optimal salinity is 1.45 and below, and growth occurs between 0.36 and 4.68. At optimal temperature and salinity, cells divide at an average rate of 0.53 day^{-1}. Genome size is 1.16 Mbp, with 1271 predicted genes, and a GC content of 29.02% (calculated). LSUCC0530 had an ANI of 75.3% and 75.9% and an AAI of 88.7% and 89.8% with HIMB114 and IMCC9063, respectively. The genome is available on GenBank under accession number CP024034.

The type strain, LSUCC0530^T, was isolated from brackish water from the coastal lagoon Lake Borgne off the coast of southeastern Louisiana, USA.

Materials and Methods

Isolation and Identification of LSUCC0530

Isolation, propagation, and identification of strain LSUCC0530 (subclade IIIb) was completed as previously reported[14]. Briefly, water was collected from surface water in the coastal lagoon Lake Borgne (Shell Beach, Louisiana) (29.872035 N, -89.672819 W) on 1 July, 2016. We recorded salinity on site using a YSI 556 MPS (YSI, Solon, OH, USA) (Table S1). Whole water was filtered through a 2.7 µm filter (Whatman), stained with 1X SYBR Green (Lonza, Basal, Switzerland), and enumerated using the Guava EasyCyte 5HT flow cytometer (Millipore, Massachusetts, USA). After serial dilution, water was inoculated into five 96 x 2.1 mL well PTFE plates (Radleys, Essex, UK) containing 1.7 mL of JW5 medium at an estimated 2 cells $well^{-1}$. Cultures were incubated at room temperature in the dark for three weeks and evaluated for growth using flow cytometry. Positive wells (> 10^4 cells mL^{-1}) were transferred in duplicate to capped, 125 mL polycarbonate flasks (Corning, New York, USA) containing 50 mL of JW5 medium. Upon reaching a density of > 5.0 x 10^5 cells mL^{-1}, cells were filtered onto 25 mm 0.22 µm polycarbonate filters (Millipore,

442 Massachusetts, USA), and DNA extractions were performed using the MoBio PowerWater
443 DNA kit (QIAGEN, Massachusetts, USA) following the manufacturer's instructions. DNA was
444 amplified as previously reported[14] and sequenced at Michigan State University RTSF
445 Genomics Core. Sanger sequences were evaluated using the freely available software
446 4Peaks (v. 1.7.1)(http://nucleobytes.com/4peaks/). Forward and reverse sequences were
447 assembled using the CAP3 webserver (http://doua.prabi.fr/software/cap3), after reverse
448 reads were converted to their reverse complement at
449 http://www.bioinformatics.org/sms/rev_comp.html.
450
451 *Community iTag analysis*
452 Community DNA from 17 sites of various salinities was filtered, extracted, and analyzed
453 following our previously reported protocol[14]. The sites sampled were Lake Borgne (LKB;
454 Shell Beach, LA), Bay Pomme D'or (JLB; Buras, LA), Terrebonne Bay (TBon; Cocodrie, LA),
455 Atchafalaya River Delta (ARD; Franklin, LA), Freshwater City (FWC; Kaplan, LA), Calcasieu
456 Jetties (CJ; Cameron, LA). Each site was visited three times with the exception of TBon,
457 which was visited only twice. Briefly, extracted DNA was sequenced at the 16S rRNA gene
458 V4 region (515F, 806R)[49], with Illumina MiSeq 2x250bp paired-end sequencing at Argonne
459 National Laboratories. 16S rRNA gene amplicon data was analyzed and classified ($OTU_{0.03}$)
460 with Mothur v.1.33.3[50] using the Silva v119 database[51]. To determine the best matching
461 OTU for LSUCC0530, sequences from the OTU representative fasta files, provided by
462 mothur using *get.oturep*(), were used to create a blast database (*formatdb*) against which
463 the LSUCC isolate 16S rRNA genes could be searched via blastn (BLAST v 2.2.26). All best
464 hits ranked at 100% identity. OTU abundance analysis was done within the R statistical
465 environment v3.2.3, within the package PhyloSeq (v1.14.0)[52]. Sequences were rarefied
466 using the command *rarefy_even_depth*() and then trimmed for OTUs with at least 2 reads in
467 at least 90% of the samples. Abundances of an OTU were averaged across biological
468 duplicates and a rank abundance matrix was calculated using the command
469 *transform_sample_counts*() with the argument function of function(x) x / sum(x).
470 Completed rarefied rank abundance OTU tables were plotted using the graphing program
471 GGPLOT2 (v. 2.0.0)[53]. All iTag sequences are available at the Short Read Archive with
472 accession number SRR6235382-SRR6235415.
473
474 *Growth experiments*
475 Salinity tolerance was tested by growing LSUCC0530 in different media distinguished by
476 proportional changes in the concentration of major ions. All other media elements (carbon,
477 iron, phosphate, nitrogen, vitamins, and trace metals) were kept consistent. Media included
478 JW1, JW2, JW3, and JW4, previously published[14], and JW3.5, JW4.5, JW5, and JW6. Recipes
479 for all media used in this study are provided in Table S1. Salinity tolerance experiments
480 were conducted in quintuplicate at room temperature. Temperature range and sole carbon
481 substrate use were tested using the isolation medium, JW5. Temperature range

482 experiments were also conducted in quintuplicate. For all experiments, growth was
483 measured using flow cytometry as described above.
484
485 *Microscopy*
486 Strain LSUCC0530 cells were grown to near max density ($1x10^7$ cells mL^{-1}) in JW5 medium.
487 Cells for transmission electron microscopy were prepared as previously described[33] with
488 one minor change: centrifuged cells were concentrated in JW5 medium with no added
489 fixative. Cells were visualized under a JEOL JSM-2011 High Resolution transmission
490 electron microscope at the LSU Socolofsky Microscopy Center, Baton Rouge, LA. For
491 scanning electron microscopy, cultured cells were fixed for four hours in 4%
492 Glutaraldehyde buffered with 0.2M Cacodylate buffer and filtered onto a 0.2 µm filter (Pall,
493 Florida, USA). Samples were dehydrated in an ethanol series (30%, 50%, 70%, 96 and
494 100%) for 15 minutes each, followed by critical point drying. Cells were visualized under a
495 JEOL JSM-6610 scanning electron microscope at the LSU Socolofsky Microscopy Center,
496 Baton Rouge, LA.
497
498 *Genome sequencing, assembly, and annotation*
499 Cells were grown in 1 L JW5 medium (Table S1) and filtered onto 0.2 µm nylon filters
500 (Manuf, STATE). DNA was obtained via phenol-chloroform extraction and sequenced using
501 both Illumina HiSeq and MiSeq approaches. *HiSeq*. 350 ng of DNA was sheared to ~500 bp
502 with a Qsonica Q800R (QSonica) using 16 cycles with a amplitude of 25% and pulse rate of
503 20 seconds on, 20 seconds off. DNA was visualized on a 1.5% agrose gel stained with Gel
504 Red (Biotum) to determine average sheared length and quantified using Qubit high
505 sensitivity dsDNA assay kit (Invitrogen). Library prep was conducted following a modified
506 Kapa Hyper Prep Kit protocol (Kapa Biosystems, Inc.). Briefly, we ligated end-repaired and
507 adenylated DNA to universal Y-yoke oligonucleotide adapters and custom Illumina iTru
508 dual-indexed primers[54]. Prior to index amplification, post ligated LSUCC0530 DNA was
509 cleaned using a 1.2x SPRI cleanup to increase ligated DNA retention. Following a brief, 8
510 cycle PCR amplification of the index library, the genomic library was quantified using a
511 Qubit broad range dsDNA assay kit (Invitrogen). Illumina HiSeq 3000 sequencing was
512 performed at Oklahoma Medical Research Facility using SBS kit chemistry to generate 150
513 bp paired-end reads, insert size 400 bp. *MiSeq*. DNA was sent to the Argonne National
514 Laboratory Environmental Sample Preparation and Sequencing Facility which performed
515 library prep and sequencing, generating 250 bp paired-end reads, insert size 550 bp. Due
516 to the large number of reads from the HiSeq sequencing, we created a random subset using
517 14% of the total reads by using the program *seqtk* (v. 1.0-r75-dirty) and the flag –s100.
518 Initial assembly with the subset of reads was done with SPAdes[55]. This generated a large
519 single scaffold with overlapping ends. We performed a quality assessment of the single
520 scaffold via iterations of Reapr[56]-advised scaffold breaks and SSPACE[57] extensions using
521 both the full set of HiSeq reads and the MiSeq reads. Final assessment of the scaffold with

522 overlaps removed from the ends was performed with Pilon[58] using all *HiSeq* reads mapped
523 to the scaffold with BWA[59]. No issues were reported. An analysis of the GC skew was
524 performed on the final scaffold using the gc_skew script provided via Brown et al. 2015[60].
525 Contamination and completion were also evaluated with CheckM[61]. Detailed commands
526 and outputs for the assembly and quality assessment are provided in Supplementary
527 Information. The final circular chromosome scaffold was annotated by IMG[62] and is
528 publicly available with IMG Taxon ID number 2728369501. It is also publicly available in
529 GenBank with accession number CP024034.
530
531 *Comparative genomics*
532 Orthologus clusters were determined using the Get Homologues pipeline[63] using the
533 LSUCC0530 genome and 48 publicly available SAR11 genomes from IMG, including the 10
534 LD12 SAGs analyzed previously[17,18] (Table S1). Clusters were determined from amino acid
535 sequences using the OrthoMCL option in Get Homologues. All clusters are provided in
536 cluster_list.txt as part of Supplemental Information. Average amino acid identity (AAI) and
537 synteny were calculated as previously reported[64]. Average nucleotide identity (ANI) was
538 calculated with the Konstantinidis Lab website ANI calculator (http://enve-
539 omics.ce.gatech.edu/ani/index) as reported[65]. Compatible solutes identified in Table S1
540 were identified in SAR11 genomes using KO numbers, annotations, and cross-referenced
541 with orthologous clusters.
542
543 *Bacteriophage searches*
544 We looked for signatures of bacteriophage in the LSUCC0530 genome using two methods,
545 VirSorter[66] and PHASTER[67] with default settings. CRISPR regions are identified as part of
546 the standard IMG annotation process, and none were found in LSUCC0530.
547
548 *Phylogenomics*
549 We selected single-copy genes present in at least 39 of the 49 genomes as determined via
550 Get Homologues (above). Amino acid sequences from these 83 markers were separately
551 aligned with MUSCLE[68], culled with Gblocks[69], and concatenated into a superalignment
552 totaling 23,356 positions. Inference was performed with RAxML[70] using the PROTCATLG
553 model and 100 bootstrapping runs. All scripts (with settings) used in the alignment, culling,
554 concatenation, and inference processes are available in Supplemental Information. Node
555 labels were renamed with Newick Utilities[71] *nw_rename*, and the tree was visualized with
556 Archaeopteryx[72].
557
558 *Single gene phylogeny*
559 The 16S rRNA gene sequence from the LSUCC0530 genome was aligned with 284
560 alphaproteobacterial sequences and seven outgroup sequences. We completed
561 phylogenetic inference for *aceA* and *malA* in an analogous manner. For each, the

562 LSUCC0530 gene was searched against the NCBI nr database using blastp (v. 2.2.28). The
563 top 100 sequence hits were combined with all homologus SAR11 sequences identified with
564 Get Homologues for the *aceA* phylogeny, and both the *malA* and *malG* sequences in all
565 SAR11s for the *malA* phylogeny. Sequences from redundant taxa were removed. For all
566 trees, alignment with MUSCLE[68], culling with Gblocks[69], and tree inference with
567 FastTree2[73] was completed with the FT_pipe script, provided in Supplemental Information,
568 along with the initial fasta file containing all sequences and the newick tree output file.
569 Node labels were changed with Newick Utilities[71] *nw_rename* and visualization was
570 performed with Archaeopteryx[72].
571
572 *Metagenomic recruitment*
573 Scaffolds for the 49 SAR11 genomes, including strain LSUCC0530, used in the
574 phylogenomic analyses were used for competitive recruitment of metagenomic sequences
575 from 89 different samples. Prior to recruitment, multi-scaffold assemblies were condensed
576 into single scaffolds, and all single scaffolds were used to generate a BLAST database with
577 the command *makeblastdb* (BLAST v. 2.2.28). Metagenomic datasets from 90 different
578 environmental samples of various salinities from the Baltic Sea[15], Amazon River,
579 freshwater lakes[10,74], Lagoon Albufera[75], Columbia River [76,77], Lake Lanier[78], Lake
580 Michigan[79], Feitsui Reservoir[80], and GOS[81] were downloaded from the Short Read Archive
581 (SRA) and European Nucleotide Archive (ENA) using *ftp* and SRA toolkit (v.2.8.2-1),
582 respectively. Reads from pyrosequencing were dereplicated with the program CD-HIT(v
583 4.6)[82] using the *cd-hit-454* command and flag -c 0.95. Illumina sequence data was converted
584 from fastq to fasta files with the FastX toolkit (v. 0.0.13.2)
585 (http://hannonlab.cshl.edu/fastx_toolkit) using the command *fastq_to_fa*sta with the flag –
586 Q33. Metagenomic sequences were separated into 1 million sequence files and recruited
587 against the SAR11 database with blastn (v. 2.2.28) with –max_target_seqs 1. Best hits for
588 each metagenomic sequence were kept if the alignment length was > 90% of the median
589 length of all metagenomic sequences for a given sample. Metagenomic recruitment was
590 expressed in read density per genome using ggplot2 *geom_density*() and reads per kilobase
591 of genome per million mapped reads (RPKM) as previously reported[83].
592
593 *Script availability*
594 All scripts used in this work can be found on the Thrash Lab website with the manuscript
595 link at http://thethrashlab.com/publications.
596
597 **Acknowledgements**
598 Funding for this work was provided in part by a Louisiana Board of Regents (grant
599 LEQSF(2014-17)-RD-A-06) grant to JCT and the LSU Department of Biological Sciences. The
600 authors thank the LSU Socolofsky Microscopy Center for assistance with electron

microscopy imaging. Portions of this research were conducted with high performance computing resources provided by Louisiana State University (http://www.hpc.lsu.edu).

Author Contributions

JCT designed the study; MWH conducted cultivation experiments, nucleic acid extraction, microscopy, and iTag and metagenomics analyses; JCT and BCF assembled the genome and conducted quality assessment; MWH, VCL, and JCT conducted comparative genomics analyses, metabolic reconstruction, phylogenetic, and phylogenomic analyses; JCT lead manuscript writing and all authors contributed to and reviewed the text.

Competing Financial Interests

The authors declare no competing financial interests.

616 **Tables and Figures**

617

618 **Table 1.** Compatible solute production and transport in SAR11

619

Osmolyte system	LSUCC0530	LD12 SAGs	HIMB114	IMCC9063	Other SAR11
Alanine synthesis	+	+	+	+	+
Di-myo-inositol phosphate synthesis	+/-	+/-	+/-	+/-	+/-
Ectoine/hydroxyectoine transport	-	-	+	+	+
Glutamate synthesis	+	+	+	+	+
Glutamate transport	-	+*	-	-	-
Glutamine synthesis	+	+	+	+	+
Glycine synthesis	+	+	+	+	+
Glycine betaine synthesis	-	+/-*	+	+	+
Glycine betaine/proline ABC transporter	-	-	+	+	+
N-∂-Acetyl-ornithine synthesis	+	+	+	+	+
Potassium transport (antiporters K03549, K03455)	+	+	-	-	-
Potassium transport (trkAH + K16052)	-	-	+	+	+
Proline synthesis	+	+	+	+	+
Proline transport (symporter)	+	+	+	+	+
Sodium transport (antiporter K03313)	-	-	+	+	+
Sodium transport (other antiporter)	+	+	-	-	-
Sorbitol synthesis	-	-	+	-	+
Taurine transport	-	-	-	-	+
TMAO synthesis	-	-	-	-	+

620

621 + genes present; - genes absent; +/- partial pathway; *B11 LD12 SAG only. Table shows
622 only systems with predicted genes in at least one SAR11 genome. For a complete list of all
623 systems searched, see Table S1.

624

625 **Figure Legends**

626

627 **Figure 1.** Phylogeny and morphology of LSUCC0530. A. Phylogenetic position of
628 LSUCC0530 within the SAR11 clade using a concatenation of 83 single copy marker genes
629 and 23,356 amino acid positions. The tree was inferred using RAxML with 100 bootstrap
630 runs. Values at the nodes indicate 100% bootstrap support unless otherwise noted. The
631 tree was rooted on HIMB59. Subclades are indicated on the right, and LD12 microclusters
632 are indicated A-C. Asterisk- we have included AAA028-C07 in microcluster A, even though
633 it was excluded previously[18]. B. Scanning electron micrograph of LSUCC0530 cells at
634 35,000x magnification. Scale bar represents 0.5 μm. C. Transmission electron micrograph
635 of a LSUCC0530 cell thin section. Scale bar represents 0.5 μm.

636

637 **Figure 2.** Circular diagram of the LSUCC0530 genome. Rings indicate, from outside to the
638 center: Position relative to the replication start site; Forward strand genes, colored by COG
639 category; Reverse strand genes, colored by COG category; RNA genes (tRNAs green, rRNAs
640 red, other RNAs black); GC content; GC skew.
641
642 **Figure 3.** Metabolic reconstruction of LSUCC0530. Solid boxes indicate recovered genes,
643 dashed boxes indicate missing genes. ABC transporters are depicted with circles indicating
644 subunits. Symporters and antiporters are depicted with ovals. Two component systems are
645 depicted with large and small rectangles. The C4 TRAP transporter is depicted separately.
646 Major subsystems are colored for easier identification. Multiple arrows indicate all genes
647 present in a given pathway. Numbers identify genes according to the key in Table S1.
648 Question marks indicate a single missing gene in an otherwise complete pathway. Blue fill
649 indicates genes with no orthologs outside of the LD12 subclade.
650
651 **Figure 4.** Growth rates for LSUCC0530 according to salinity (A) and temperature (B).
652 Points indicate individual biological replicates, calculated from corresponding growth
653 curves in Figure S8.
654
655 **Figure 5.** Results of competitive metagenomic recruitment to all genomes depicted in Fig.
656 1. A. 95% identity RPKM values for LD12 genomes only, by site, with data aggregated for all
657 genomes in microclusters A-C, according to the key. Colors indicate broad environmental
658 categories. B-D. Density plots of metagenomic sequence recruitment according to percent
659 identity of the best hit. Data was aggregated according to the key, with LD12 separated as
660 in A, HIMB59 separately depicted as "V", and all other SAR11 genomes together as "Other."
661 B. Two sites where LSUCC0530 dominated recruitment with high percent identity. C. Two
662 sites where the LSUCC0530 only recruited sequences at ~90% identity and no other group
663 had high recruitment at higher identity. D. Two sites where other LD12 microclusters
664 showed similar patterns as LSUCC0530 in B and C.
665
666

667 **References Cited**

668

669 1 Morris, R. M. *et al.* SAR11 clade dominates ocean surface bacterioplankton
670 communities. *Nature* **420**, 806-810, (2002).
671 2 Schattenhofer, M. *et al.* Latitudinal distribution of prokaryotic picoplankton
672 populations in the Atlantic Ocean. *Environmental Microbiology* **11**, 2078-2093,
673 (2009).
674 3 Carlson, C. A. *et al.* Seasonal dynamics of SAR11 populations in the euphotic and
675 mesopelagic zones of the northwestern Sargasso Sea. *ISME J.* **3**, 283-295, (2009).
676 4 Giovannoni, S. J. & Vergin, K. L. Seasonality in ocean microbial communities. *Science*
677 **335**, 671-676, (2012).
678 5 Thrash, J. C. *et al.* Single-cell enabled comparative genomics of a deep ocean SAR11
679 bathytype. *ISME J.* **8**, 1440-1451, (2014).
680 6 Tsementzi, D. *et al.* SAR11 bacteria linked to ocean anoxia and nitrogen loss. *Nature*
681 **536**, 179-183, (2016).
682 7 Luo, H., Csűros, M., Hughes, A. L. & Moran, M. Evolution of Divergent Life History
683 Strategies in Marine Alphaproteobacteria. *mBio* **4**, e00373-00313, (2013).
684 8 Logares, R., Brate, J., Heinrich, F., Shalchian-Tabrizi, K. & Bertilsson, S. Infrequent
685 Transitions between Saline and Fresh Waters in One of the Most Abundant
686 Microbial Lineages (SAR11). *Mol. Biol. Evol.* **27**, 347-357, (2010).
687 9 Salcher, M. M., Pernthaler, J. & Posch, T. Seasonal bloom dynamics and
688 ecophysiology of the freshwater sister clade of SAR11 bacteria 'that rule the waves'
689 (LD12). *ISME J.* **5**, 1242-1252, (2011).
690 10 Eiler, A. *et al.* Productivity and salinity structuring of the microplankton revealed by
691 comparative freshwater metagenomics. *Environmental Microbiology* **16**, 2682-2698,
692 (2014).
693 11 Cabello-Yeves, P. J. *et al.* Genomes of novel microbial lineages assembled from the
694 sub-ice waters of Lake Baikal. *Appl. Environ. Microbiol.* **AOP**, (2017).
695 12 Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D. & Bertilsson, S. A Guide to the
696 Natural History of Freshwater Lake Bacteria. *Microbiol. Mol. Biol. Rev.* **75**, 14-49,
697 (2011).
698 13 Garcia, S. L. *et al.* Contrasting patterns of genome-level diversity across distinct co-
699 occurring bacterial populations. *bioRxiv*, (2017).
700 14 Henson, M. W. *et al.* Artificial Seawater Media Facilitate Cultivating Members of the
701 Microbial Majority from the Gulf of Mexico. *mSphere* **1**, e00028-00016.
702 doi:00010.01128/mSphere.00028-00016., (2016).
703 15 Dupont, C. L. *et al.* Functional Tradeoffs Underpin Salinity-Driven Divergence in
704 Microbial Community Composition. *PLOS ONE* **9**, e89549, (2014).
705 16 Logares, R. *et al.* Infrequent marine–freshwater transitions in the microbial world.
706 *Trends in Microbiology* **17**, 414-422, (2009).
707 17 Eiler, A. *et al.* Tuning fresh: radiation through rewiring of central metabolism in
708 streamlined bacteria. *ISME J.* **10**, 1902-1914, (2016).
709 18 Zaremba-Niedzwiedzka, K. *et al.* Single-cell genomics reveal low recombination
710 frequencies in freshwater bacteria of the SAR11 clade. *Genome Biol* **14**, 1-14, (2013).

711 19 Grote, J. *et al.* Streamlining and Core Genome Conservation among Highly Divergent
712 Members of the SAR11 Clade. *mBio* **3**, e00252-00212, (2012).
713 20 Giovannoni, S. J. *et al.* Genome streamlining in a cosmopolitan oceanic bacterium.
714 *Science* **309**, 1242-1245, (2005).
715 21 Giovannoni, S. J. *et al.* Proteorhodopsin in the ubiquitous marine bacterium SAR11.
716 *Nature* **438**, 82-85, (2005).
717 22 Tripp, H. J. *et al.* SAR11 marine bacteria require exogenous reduced sulphur for
718 growth. *Nature* **452**, 741-744, (2008).
719 23 Schwalbach, M. S., Tripp, H. J., Steindler, L., Smith, D. P. & Giovannoni, S. J. The
720 presence of the glycolysis operon in SAR11 genomes is positively correlated with
721 ocean productivity. *Environmental Microbiology* **12**, 490-500, (2010).
722 24 Sun, J. *et al.* One Carbon Metabolism in SAR11 Pelagic Marine Bacteria. *PLOS ONE* **6**,
723 e23973, (2011).
724 25 Carini, P., Steindler, L., Beszteri, S. & Giovannoni, S. J. Nutrient requirements for
725 growth of the extreme oligotroph 'Candidatus Pelagibacter ubique' HTCC1062 on a
726 defined medium. *ISME J.* **7**, 592-602, (2013).
727 26 Carini, P. *et al.* Discovery of a SAR11 growth requirement for thiamin's pyrimidine
728 precursor and its distribution in the Sargasso Sea. *ISME J.* **8**, 1727-1738, (2014).
729 27 Smith, D. P. *et al.* Proteome Remodeling in Response to Sulfur Limitation in
730 "Candidatus Pelagibacter ubique". *mSystems* **1**, 16, (2016).
731 28 Smith, D. P. *et al.* Proteomic and Transcriptomic Analyses of "Candidatus
732 Pelagibacter ubique" Describe the First PII-Independent Response to Nitrogen
733 Limitation in a Free-Living Alphaproteobacterium. *mBio* **4**, 12, (2013).
734 29 Steindler, L., Schwalbach, M. S., Smith, D. P., Chan, F. & Giovannoni, S. J. Energy
735 Starved Candidatus Pelagibacter Ubique Substitutes Light-Mediated ATP Production
736 for Endogenous Carbon Respiration. *PLOS ONE* **6**, e19725, (2011).
737 30 Zhao, X. *et al.* Three-Dimensional Structure of the Ultraoligotrophic Marine
738 Bacterium "Candidatus Pelagibacter ubique". *Appl. Environ. Microbiol.* **83**, 16,
739 (2017).
740 31 Zhao, Y. *et al.* Abundant SAR11 viruses in the ocean. *Nature* **494**, 357-360, (2013).
741 32 Konstantinidis, K. T. & Tiedje, J. M. Prokaryotic taxonomy and phylogeny in the
742 genomic era: advancements and challenges ahead. *Current Opinion in Microbiology*
743 **10**, 504-509, (2007).
744 33 Rappé, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. Cultivation of the
745 ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**, 630-633, (2002).
746 34 Wilhelm, L. J., Tripp, J. H., Givan, S. A., Smith, D. P. & Giovannoni, S. J. Natural
747 variation in SAR11 marine bacterioplankton genomes inferred from metagenomic
748 data. *Biology Direct* **2**, 1-19, (2007).
749 35 Rodriguez-Valera, F. *et al.* Explaining microbial population genomics through phage
750 predation. *Nat Rev Micro* **7**, 828-836, (2009).
751 36 Anstrom, D. M., Kallio, K. & Remington, J. S. Structure of the Escherichia coli malate
752 synthase G:pyruvate:acetyl‐coenzyme A abortive ternary complex at 1.95 Å
753 resolution. *Protein Science* **12**, 1822-1832, (2003).
754 37 Kappler, U. Bacterial sulfite-oxidizing enzymes. *Biochimica et Biophysica Acta (BBA)*
755 *- Bioenergetics* **1807**, 1-10, (2011).

38 Denger, K., Weinitschke, S., Smits, T. H. M., Schleheck, D. & Cook, A. M. Bacterial sulfite dehydrogenases in organotrophic metabolism: separation and identification in Cupriavidus necator H16 and in Delftia acidovorans SPH-1. *Microbiology* **154**, 256-263, (2008).

39 Béjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M. & DeLong, E. F. Proteorhodopsin phototrophy in the ocean. *Nature* **411**, 786-789, (2001).

40 Konstantinidis, K. T. & Tiedje, J. M. Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2567-2572, (2005).

41 Jimenez-Infante, F. *et al.* Genomic Characterization of Two Novel SAR11 Isolates From the Red Sea, Including the First Strain of the SAR11 Ib clade. *FEMS Microbiol Ecol*, (2017).

42 Song, J., Oh, H.-M. & Cho, J.-C. Improved culturability of SAR11 strains in dilution-to-extinction culturing from the East Sea, West Pacific Ocean. *FEMS Microbiology Letters* **295**, 141-147, (2009).

43 Stingl, U., Tripp, H. J. & Giovannoni, S. J. Improvements of high-throughput culturing yielded novel SAR11 strains and other abundant marine bacteria from the Oregon coast and the Bermuda Atlantic Time Series study site. *ISME J.* **1**, 361-371, (2007).

44 Piwosz, K., Salcher, M. M. & and …, Z.-M. Seasonal dynamics and activity of typical freshwater bacteria in brackish waters of the Gulf of Gdańsk. *Limnology and …*, (2013).

45 Herlemann, D., Woelk, J., Labrenz, M. & Jürgens, K. Diversity and abundance of "Pelagibacterales" (SAR11) in the Baltic Sea salinity gradient. *Syst. Appl. Microbiol.* **37**, 601-604, (2014).

46 Dattananda, C. S. & Gowrishankar, J. Osmoregulation in Escherichia coli: complementation analysis and gene-protein relationships in the proU locus. *J. Bacteriol.* **171**, 1915-1922, (1989).

47 Empadinhas, N. & da Costa, M. Osmoadaptation mechanisms in prokaryotes: distribution of compatible solutes. *International Microbiology*, (2008).

48 Sun, J. *et al.* The abundant marine bacterium Pelagibacter simultaneously catabolizes dimethylsulfoniopropionate to the gases dimethylsulfide and methanethiol. *Nature Microbiology* **1**, (2016).

49 Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology* **18**, (2015).

50 Schloss, P. D. *et al.* Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl. Environ. Microbiol.* **75**, 7537-7541, (2009).

51 Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research* **35**, 7188-7196, (2007).

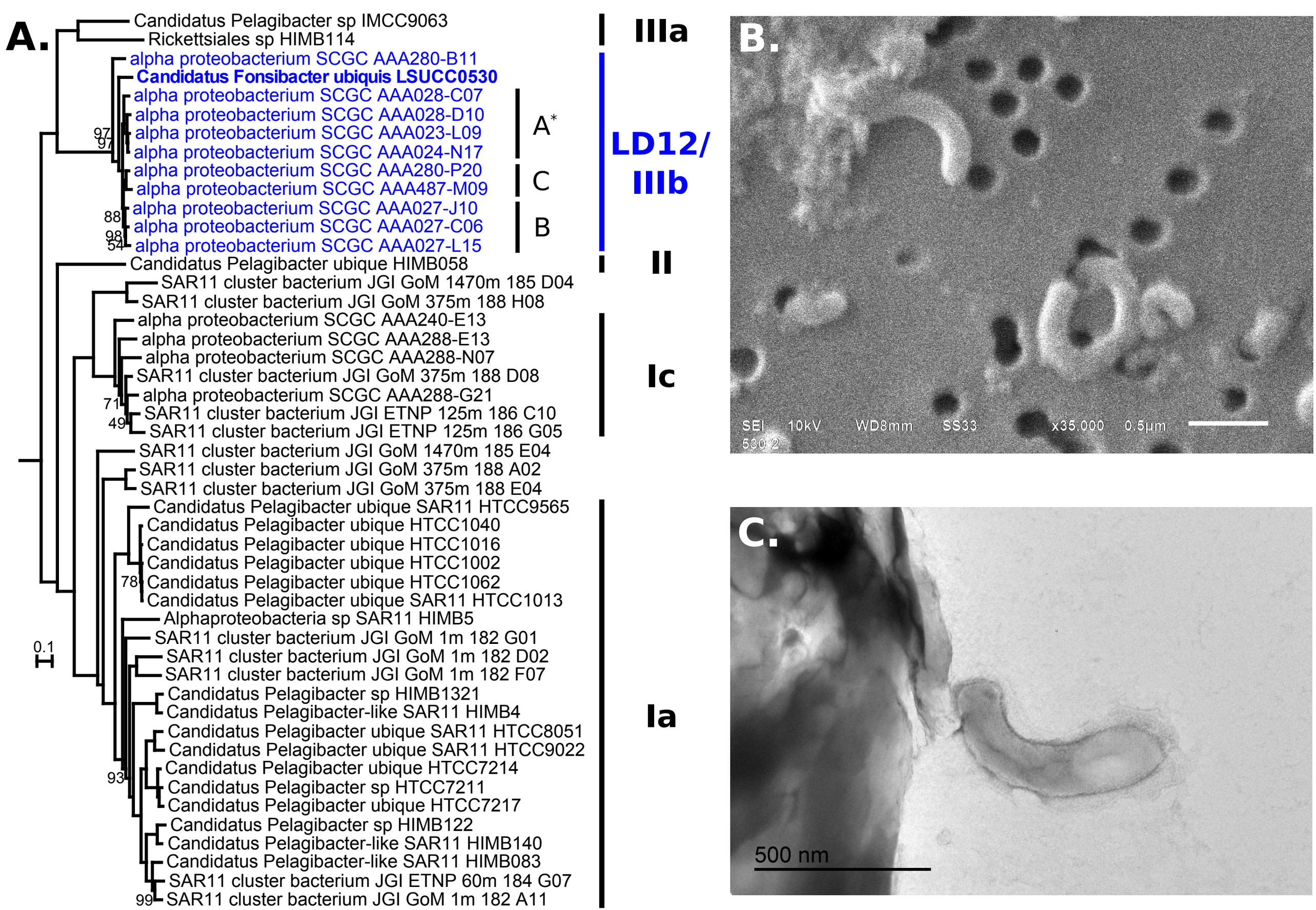
52 McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE* **8**, e61217, (2013).

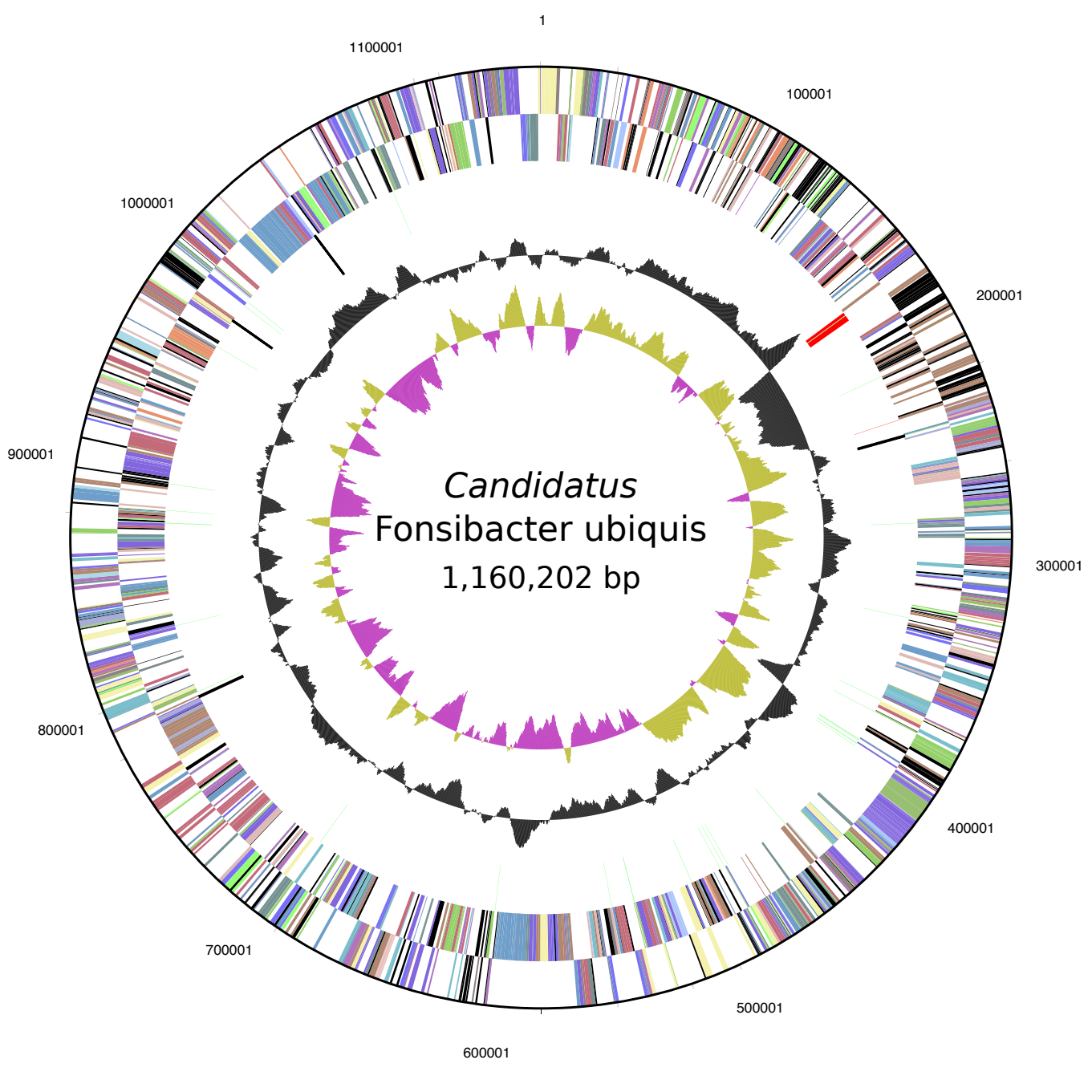
53 Wickham, H. ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics* **3**, 180-185, (2011).

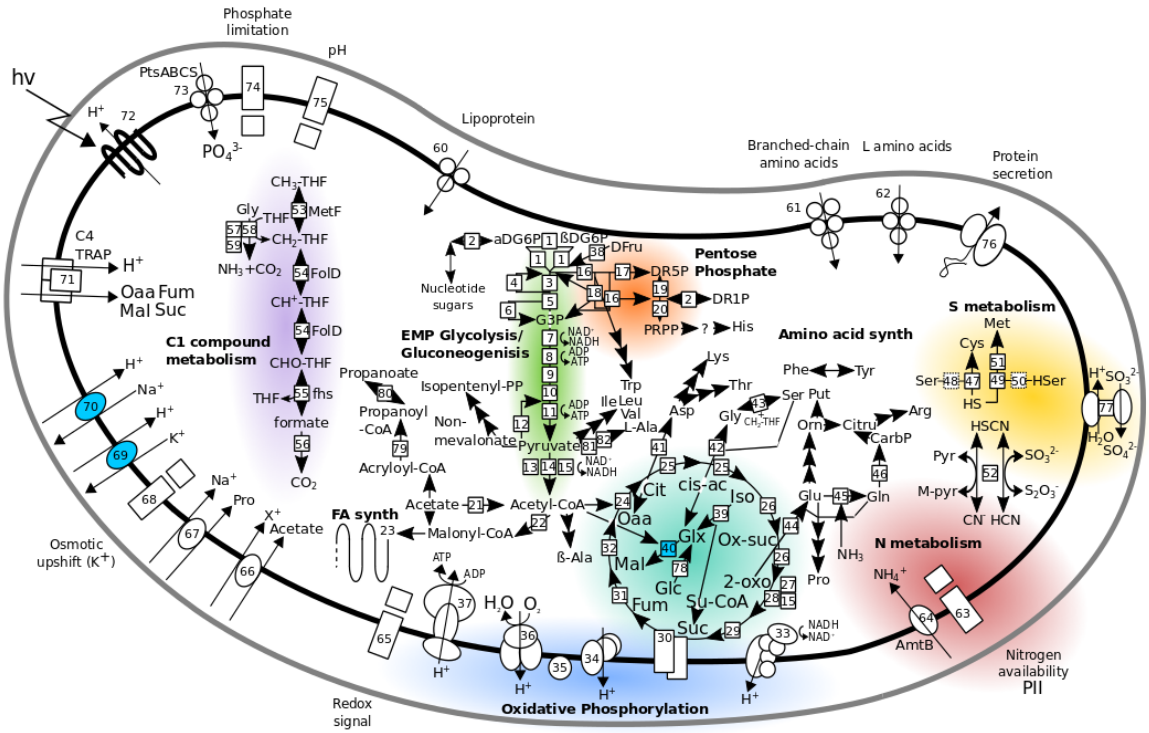
54 Glenn, T. C. *et al.* Adapterama I: Universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext). *bioRxiv*, (2016).

802 55 Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications
803 to Single-Cell Sequencing. *Journal of Computational Biology* **19**, 455-477, (2012).
804 56 Hunt, M. *et al.* REAPR: a universal tool for genome assembly evaluation. *Genome Biol*
805 **14**, 1-10, (2013).
806 57 Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-
807 assembled contigs using SSPACE. *Bioinformatics* **27**, 578-579, (2011).
808 58 Walker, B. J. *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant
809 Detection and Genome Assembly Improvement. *PLOS ONE* **9**, (2014).
810 59 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler
811 transform. *Bioinformatics* **25**, 1754-1760, (2009).
812 60 Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of
813 domain Bacteria. *Nature* **523**, 208-211, (2015).
814 61 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM:
815 assessing the quality of microbial genomes recovered from isolates, single cells, and
816 metagenomes. *Genome Research* **25**, 1043-1055, (2015).
817 62 Markowitz, V. M. *et al.* IMG 4 version of the integrated microbial genomes
818 comparative analysis system. *Nucleic Acids Research* **42**, D560-567, (2014).
819 63 Contreras-Moreira, B. & Vinuesa, P. GET_HOMOLOGUES, a versatile software
820 package for scalable and robust microbial pangenome analysis. *Appl. Environ.*
821 *Microbiol.* **79**, 7696-7701, (2013).
822 64 Dick, G. J. *et al.* Community-wide analysis of microbial genome sequence signatures.
823 *Genome Biol* **10**, R85, (2009).
824 65 Goris, J. *et al.* DNA–DNA hybridization values and their relationship to whole-
825 genome sequence similarities. *IJSEM* **57**, 81-91, (2007).
826 66 Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from
827 microbial genomic data. *PeerJ* **3**, (2015).
828 67 Arndt, D. *et al.* PHASTER: a better, faster version of the PHAST phage search tool.
829 *Nucleic Acids Research* **44**, (2016).
830 68 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high
831 throughput. *Nucleic Acids Research* **32**, 1792-1797, (2004).
832 69 Castresana, J. Selection of conserved blocks from multiple alignments for their use in
833 phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540-552, (2000).
834 70 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
835 large phylogenies. *Bioinformatics* **30**, 1312-1313, (2014).
836 71 Junier, T. & Zdobnov, E. M. The Newick utilities: high-throughput phylogenetic tree
837 processing in the Unix shell. *Bioinformatics* **26**, 1669-1670, (2010).
838 72 Han, M. V. & Zmasek, C. M. phyloXML: XML for evolutionary biology and comparative
839 genomics. *BMC Bioinformatics* **10**, 1-6, (2009).
840 73 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-
841 likelihood trees for large alignments. *PLOS ONE* **5**, e9490, (2010).
842 74 Martinez-Garcia, M. *et al.* High-throughput single-cell sequencing identifies
843 photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. *ISME J.* **6**,
844 113-123, (2011).
845 75 Ghai, R. *et al.* Metagenomes of Mediterranean Coastal Lagoons. *Sci. Rep.* **2**, (2012).

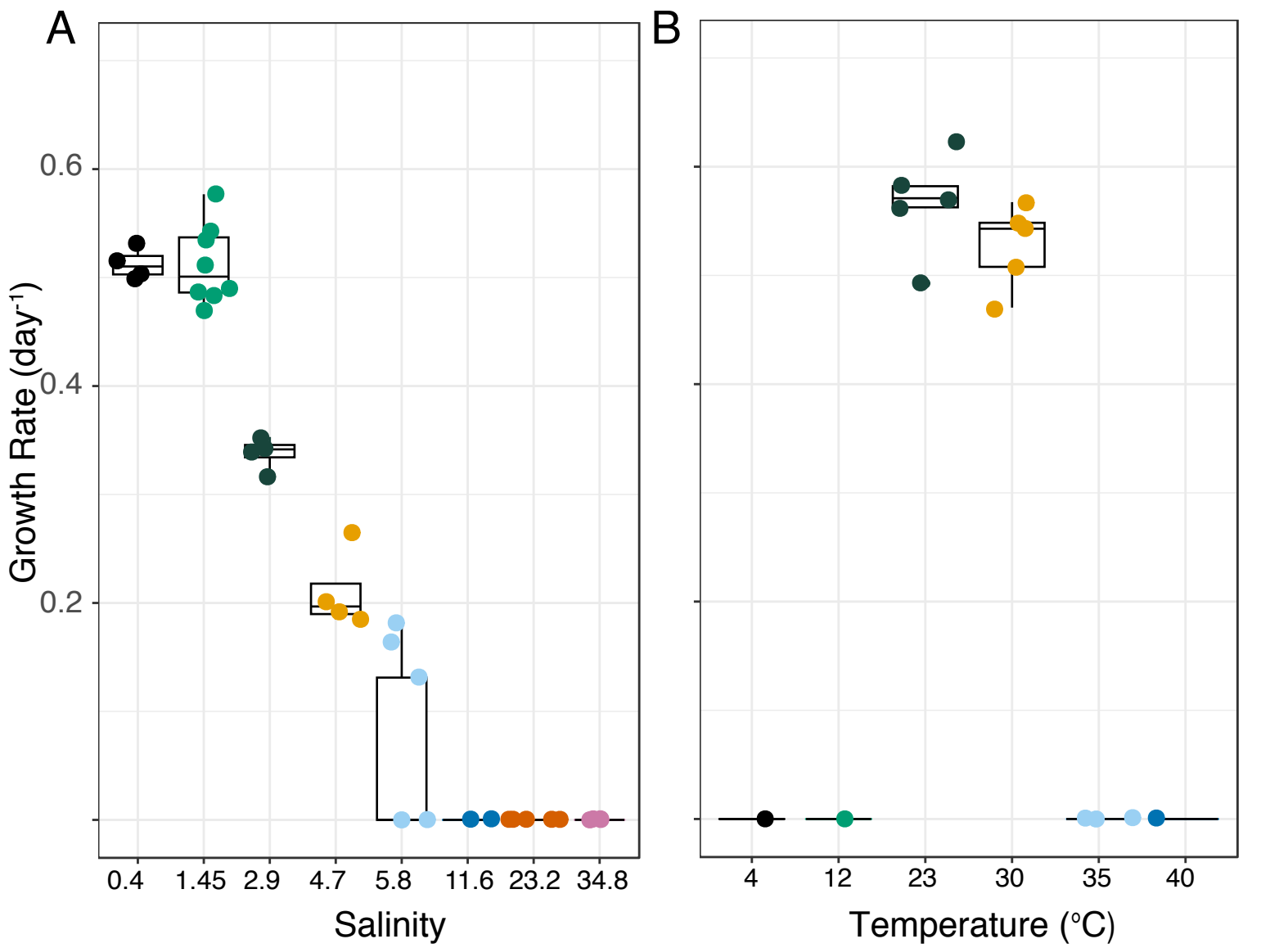
846 76 Smith, M. W., Allen, L., Allen, A. E., Herfort, L. & Simon, H. M. Contrasting genomic
847 properties of free-living and particle-attached microbial assemblages within a
848 coastal ecosystem. *Front. Microbiol.* **4**, (2013).
849 77 Fortunato, C. S. & Crump, B. C. Microbial Gene Abundance and Expression Patterns
850 across a River to Ocean Salinity Gradient. *PLOS ONE* **10**, (2015).
851 78 Oh, S. *et al.* Metagenomic Insights into the Evolution, Function, and Complexity of
852 the Planktonic Microbial Community of Lake Lanier, a Temperate Freshwater
853 Ecosystem. *Appl. Environ. Microbiol.* **77**, 6000-6011, (2011).
854 79 Denef, V. J., Mueller, R. S., Chiang, E., Liebig, J. R. & Vanderploeg, H. A. Chloroflexi
855 CL500-11 Populations That Predominate Deep-Lake Hypolimnion Bacterioplankton
856 Rely on Nitrogen-Rich Dissolved Organic Matter Metabolism and C1 Compound
857 Oxidation. *Appl. Environ. Microbiol.* **82**, 1423-1432, (2016).
858 80 Tseng, C.-H. *et al.* Microbial and viral metagenomes of a subtropical freshwater
859 reservoir subject to climatic disturbances. *ISME J.* **7**, 2374-2386, (2013).
860 81 Rusch, D. B. *et al.* The Sorcerer II Global Ocean Sampling Expedition: Northwest
861 Atlantic through Eastern Tropical Pacific. *Plos Biol* **5**, e77, (2007).
862 82 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-
863 generation sequencing data. *Bioinformatics* **28**, 3150-3152, (2012).
864 83 Thrash, J. *et al.* Metabolic Roles of Uncultivated Bacterioplankton Lineages in the
865 Northern Gulf of Mexico "Dead Zone". *mBio* **8**, (2017).
866

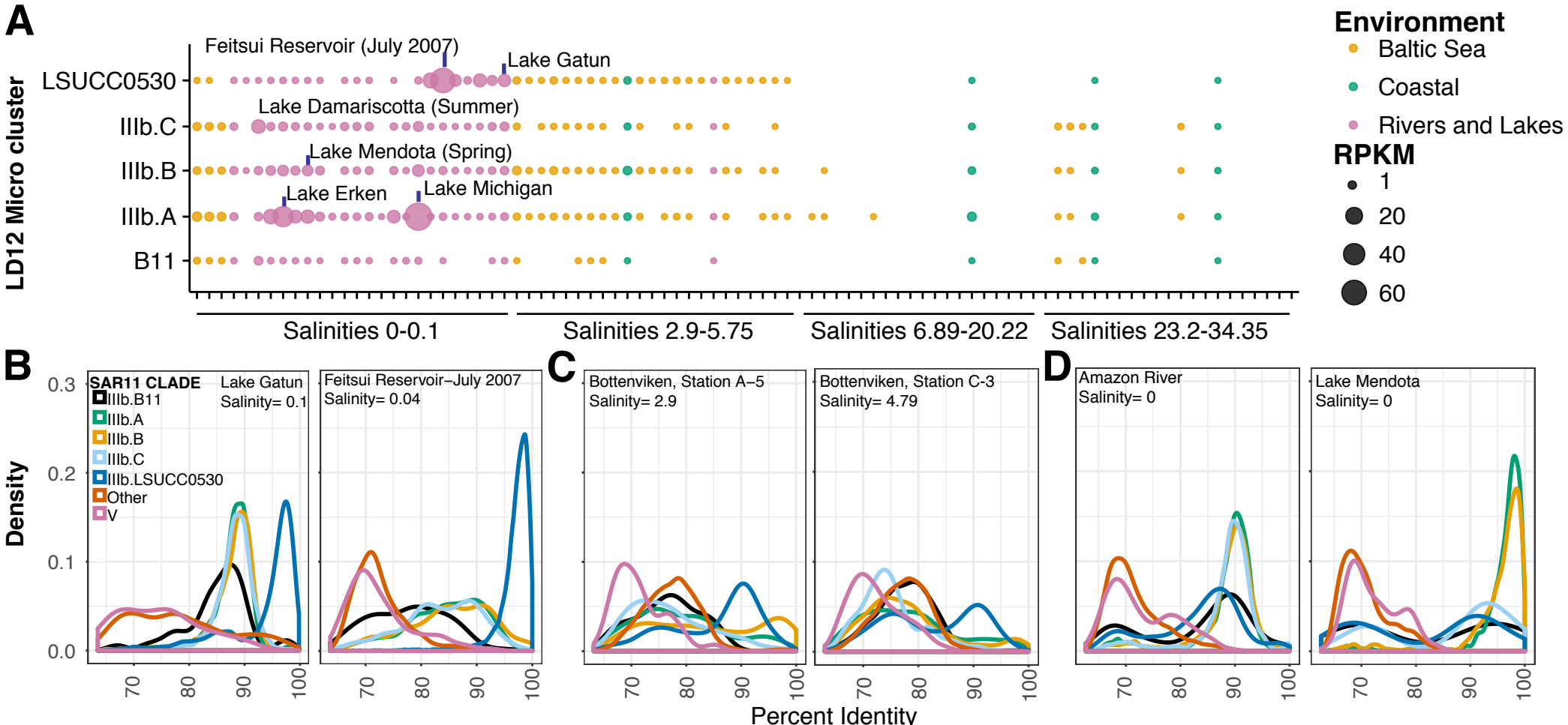

A.
Candidatus Pelagibacter sp IMCC9063
Rickettsiales sp HIMB114
alpha proteobacterium SCGC AAA280-B11
Candidatus Fonsibacter ubiquis LSUCC0530
alpha proteobacterium SCGC AAA028-C07
alpha proteobacterium SCGC AAA028-D10
alpha proteobacterium SCGC AAA023-L09
alpha proteobacterium SCGC AAA024-N17
alpha proteobacterium SCGC AAA280-P20
alpha proteobacterium SCGC AAA487-M09
alpha proteobacterium SCGC AAA027-J10
alpha proteobacterium SCGC AAA027-C06
alpha proteobacterium SCGC AAA027-L15
Candidatus Pelagibacter ubique HIMB058
SAR11 cluster bacterium JGI GoM 1470m 185 D04
SAR11 cluster bacterium JGI GoM 375m 188 H08
alpha proteobacterium SCGC AAA240-E13
alpha proteobacterium SCGC AAA288-E13
alpha proteobacterium SCGC AAA288-N07
SAR11 cluster bacterium JGI GoM 375m 188 D08
alpha proteobacterium SCGC AAA288-G21
SAR11 cluster bacterium JGI ETNP 125m 186 C10
SAR11 cluster bacterium JGI ETNP 125m 186 G05
SAR11 cluster bacterium JGI GoM 1470m 185 E04
SAR11 cluster bacterium JGI GoM 375m 188 A02
SAR11 cluster bacterium JGI GoM 375m 188 E04
Candidatus Pelagibacter ubique SAR11 HTCC9565
Candidatus Pelagibacter ubique HTCC1040
Candidatus Pelagibacter ubique HTCC1016
Candidatus Pelagibacter ubique HTCC1002
Candidatus Pelagibacter ubique HTCC1062
Candidatus Pelagibacter ubique SAR11 HTCC1013
Alphaproteobacteria sp SAR11 HIMB5
SAR11 cluster bacterium JGI GoM 1m 182 G01
SAR11 cluster bacterium JGI GoM 1m 182 D02
SAR11 cluster bacterium JGI GoM 1m 182 F07
Candidatus Pelagibacter sp HIMB1321
Candidatus Pelagibacter-like SAR11 HIMB4
Candidatus Pelagibacter ubique SAR11 HTCC8051
Candidatus Pelagibacter ubique SAR11 HTCC9022
Candidatus Pelagibacter ubique HTCC7214
Candidatus Pelagibacter sp HTCC7211
Candidatus Pelagibacter ubique HTCC7217
Candidatus Pelagibacter sp HIMB122
Candidatus Pelagibacter-like SAR11 HIMB140
Candidatus Pelagibacter-like SAR11 HIMB083
SAR11 cluster bacterium JGI ETNP 60m 184 G07
SAR11 cluster bacterium JGI GoM 1m 182 A11
97
97
88
98
54
71
49
78
93
99
0.1
A*
C
B
IIIa
LD12/
IIIb
II
Ic
Ia
B.
SEI 10kV WD8mm SS33 x35,000 0.5µm
530 2
C.
500 nm


1
100001
200001
300001
400001
500001
600001
700001
800001
900001
1000001
1100001
Candidatus
Fonsibacter ubiquis
1,160,202 bp



Phosphate limitation
pH
hv
PtsABCS
73
74
75
72
H+
PO4 3-
Lipoprotein
60
Branched-chain amino acids
L amino acids
Protein secretion
61
62
76
C4 TRAP
71
H+
Oaa Fum Mal Suc
CH3-THF
Gly
THF
53 MetF
57 58
59
CH2-THF
NH3 +CO2
54 FolD
CH+-THF
54 FolD
C1 compound metabolism
CHO-THF
THF
55 fhs
formate
56
CO2
aDG6P
ßDG6P
DFru
Pentose Phosphate
DR5P
DR1P
PRPP
? His
Nucleotide sugars
G3P
EMP Glycolysis/ Gluconeogenisis
NAD+
NADH
ADP
ATP
Trp
Ile Leu Val
L-Ala
Asp
Lys
Thr
Gly
CH2-THF
Amino acid synth
Phe
Tyr
Ser Put
Orn
Citru
Arg
CarbP
Glu
Gln
NH3
Pro
S metabolism
Met
Cys
Ser
HS
HSer
HSCN
Pyr
M-pyr
CN- HCN
SO3 2-
S2O3-
H+ SO3 2-
H2O
SO4 2-
N metabolism
NH4+
AmtB
Nitrogen availability PII
Propanoate
Isopentenyl-PP
Propanoyl -CoA
Non-mevalonate
Pyruvate
Acryloyl-CoA
Acetate
Acetyl-CoA
Malonyl-CoA
FA synth
ß-Ala
Cit
cis-ac
Iso
Oaa
Glx
Ox-suc
Mal
Glc
2-oxo
Fum
Su-CoA
Suc
ATP
ADP
H2O
O2
Osmotic upshift (K+)
Na+
K+
X+
Acetate
Redox signal
Oxidative Phosphorylation


A
B
Growth Rate (day-1)
0.6
0.4
0.2
0.4
1.45
2.9
4.7
5.8
11.6
23.2
34.8
Salinity
4
12
23
30
35
40
Temperature (°C)


A
LD12 Micro cluster
LSUCC0530
IIIb.C
IIIb.B
IIIb.A
B11
Feitsui Reservoir (July 2007)
Lake Gatun
Lake Damariscotta (Summer)
Lake Mendota (Spring)
Lake Erken
Lake Michigan
Salinities 0-0.1
Salinities 2.9-5.75
Salinities 6.89-20.22
Salinities 23.2-34.35
Environment
Baltic Sea
Coastal
Rivers and Lakes
RPKM
1
20
40
60
B
C
D
Density
SAR11 CLADE
IIIb.B11
IIIb.A
IIIb.B
IIIb.C
IIIb.LSUCC0530
Other
V
Lake Gatun
Salinity= 0.1
Feitsui Reservoir–July 2007
Salinity= 0.04
Bottenviken, Station A–5
Salinity= 2.9
Bottenviken, Station C-3
Salinity= 4.79
Amazon River
Salinity= 0
Lake Mendota
Salinity= 0
Percent Identity