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 suggesting 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. 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 marine and freshwater systems. In the oceans, SAR11 can represent 25-50% of total planktonic cells (Morris *et al.*, 2002, Schattenhofer *et al.*, 2009). Numerous subclades with unique spatio-temporal distributions comprise SAR11 (Carlson *et al.*, 2009, Giovannoni and Vergin, 2012, Morris *et al.*, 2002, Thrash *et al.*, 2014). At least nine subclades defined via 16S rRNA gene sequences occupy marine niches (Giovannoni and Vergin, 2012), and more likely exist (Tsementzi *et al.*, 2016). However, in spite of its global distribution (Morris *et al.*, 2002), massive predicted population size of 10^{28} cells (Morris *et al.*, 2002), and an estimated divergence time from its last common ancestor of 1.1 billion years ago (Luo *et al.*), the bulk of existing evidence suggests that SAR11 has only successfully colonized freshwater environments once in its natural history (Eiler *et al.*, 2014, Logares *et al.*, 2010, Salcher *et al.*, 2011). Traditionally, all known freshwater SAR11 belong to subclade IIIb, a.k.a. LD12. However, a recent report challenges this assertion: a genome sister to subclade I was recovered in Lake Baikal metagenomic data (Cabello-Yeves *et al.*, 2018). Regardless, the limited evolutionary diversification into less saline habitats has not prevented LD12 from achieving prominence in the ecosystems it inhabits. In many lotic and lentic environments, LD12 occupies similar relative abundances as its marine cousins (Dupont *et al.*, 2014, Garcia *et al.*, 2017, Henson *et al.*, 2016, Newton *et al.*, 2011, Salcher *et al.*, 2011). Study of LD12 is important to understand SAR11 evolution, specifically, and how successful transitions between marine and freshwater environments occur in bacterioplankton (Logares *et al.*, 2009), 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 (Eiler *et al.*, 2016, Luo *et al.*, 2013, Salcher *et al.*, 2011, Zaremba-Niedzwiedzka *et al.*, 2013). 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 (Dupont *et al.*, 2014, Eiler *et al.*, 2014, Eiler *et al.*, 2016, Giovannoni *et al.*, 2005a, Giovannoni *et al.*, 2005b, Grote *et al.*, 2012, Salcher *et al.*, 2011, Schwalbach *et al.*, 2010, Sun *et al.*, 2011, Thrash *et al.*, 2014, Tripp *et al.*, 2008, Zaremba-Niedzwiedzka *et al.*, 2013). Previous research suggests that LD12 differ from their marine counterparts in specific elements of metabolic potential that indicate a greater emphasis on production, rather than uptake, of osmolytes, and important metabolic changes related to energy production (Dupont *et al.*, 2014, Eiler *et al.*, 2016). 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 (Dupont *et al.*, 2014). Comparative genomic analyses of SAGs from different SAR11 strains

concurred: LD12 genomes contained the EMP pathway that is not found in most marine SAR11 (Eiler *et al.*, 2016, Grote *et al.*, 2012). SAG data also suggested that LD12 lacks the glyoxylate shunt and some single carbon (C1) metabolism (Eiler *et al.*, 2016).

Despite 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 (Zaremba-Niedzwiedzka *et al.*, 2013), and their population dynamics recently described with five-year time series data in freshwater lakes (Garcia *et al.*, 2017). However, we cannot delineate what distinguishes ecotypes without better physiological and genomic data. Similarly, 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 EMP-based glycolysis and the freshwater lifestyle, or what other adaptations might explain LD12 evolution away from salt water.

The next steps in translating 'omics-based predictions into measured data for integration with ecosystem models require living experimental subjects. For example, cultures of marine SAR11, such as HTCC1062 and HTCC7211, have facilitated testing of metabolism and growth (Carini *et al.*, Carini *et al.*, Giovannoni *et al.*, 2005a, Schwalbach *et al.*, 2010, Smith *et al.*, Smith *et al.*, Steindler *et al.*, 2011, Tripp *et al.*, 2008), structural organization (Zhao *et al.*, 2017), and virus-host interactions (Zhao *et al.*, 2013). We need cultivated representatives to provide this kind of understanding of other important bacterioplankton like LD12. In service of this goal, we pursued a systematic high throughput cultivation effort from coastal regions in the northern Gulf of Mexico. Here, we report the first isolation of a member of the LD12 clade, strain LSUCC0530, its complete genome sequence, preliminary physiological data, and an examination of the ecological distribution of this organism compared with other LD12 clade members. Our results provide evidence for temperature-dependent ecotype diversification within LD12 and a hypothesis about the evolutionary trajectory that led to LD12 colonization of fresh water.

Results

Isolation and growth of LSUCC0530

We isolated strain LSUCC0530 as part of a series of high throughput dilution-to-extinction cultivation experiments (HTC) conducted with inoculation water obtained across the southern Louisiana coast. The particular experiment that yielded LSUCC0530 utilized surface water from the coastal lagoon of Lake Borgne, 39 km southeast of New Orleans. At the time of collection, the surface water had a salinity of 2.39 and was 30.5°C. Subsequent community analysis demonstrated that LD12-type organisms represented 8.7% of the bacterioplankton population in the inoculum (see below). We diluted 2.7 µm-filtered Lake Borgne surface water for inoculation to an estimated 2 cells well^{-1} into our JW5 medium (salinity 1.45, Table S1), and incubated cultures for 30 days at 23°C (room temperature in

124 the Thrash Lab). LSUCC0530 grew slowly, reaching 8.6 x 10^5 cells mL^{-1} after 24 days. It had
125 a low fluorescence/low side scatter flow cytometric signature characteristic of other
126 SAR11 cells (Fig. S1).
127
128 *Taxonomy and morphology*
129 The LSUCC0530 16S rRNA gene sequence had > 99% identity with eight SAG sequences
130 from members of LD12 (Zaremba-Niedzwiedzka *et al.*, 2013) and only 91% and 92%
131 sequence identity with sequences from the sister subclade IIIa representatives IMCC9063
132 and HIMB114, respectively. Of note for taxonomic incorporation of genomic data (Chun *et*
133 *al.*, 2018), the genome-generated 16S rRNA gene sequence matched the PCR-generated 16S
134 rRNA gene sequence (GenBank accession number KY290650.1) at 1316 of 1320 sites. Two
135 of these mismatches were single bp gaps the genome sequence, and three of the four
136 mismatches occurred in the first 20 bp of the PCR-generated sequence. LSUCC0530 also
137 had average amino acid identities (AAI) of 58.1% and 59.7% with HIMB114 and IMCC9063,
138 respectively (Table S1). In contrast, AAI between LSUCC0530 and the LD12 SAGs ranged
139 from 81.9% to 88.0%, corroborating that LSUCC0530 belongs to separate species from
140 subclade IIIa (Konstantinidis and Tiedje, 2007). Phylogenetic inference using 16S rRNA
141 genes placed LSUCC0530 within the Order *Pelagibacteriales* (a.k.a. SAR11 (Thrash *et al.*,
142 2014)), in monophyly with the original LD12 clone library sequence (Fig. S2).
143 Phylogenomic inference using 83 single copy marker genes from 49 SAR11 genomes also
144 strongly supported placement of LSUCC0530 within LD12 as an early diverging member,
145 with only the SAG AAA280-B11 from Lake Damariscotta diverging earlier (Fig. 1A). This
146 tree also recovered the three microclusters within LD12 (A, B, C) observed previously
147 (Zaremba-Niedzwiedzka *et al.*, 2013) (Fig. 1A). Cells of strain LSUCC0530 were small,
148 curved rods, approximately 1 µm x 0.1 µm (Fig. 1B, C), showing conserved morphology
149 with the distantly related SAR11 strain HTCC1062 in subclade Ia (Rappé *et al.*, 2002, Zhao
150 *et al.*, 2017).
151
152 *Genome characteristics*
153 Genome assembly resulted in a single, circularized scaffold that we tested for completeness
154 using multiple tools (Supplemental Information). The completed genome is 1,160,202 bp
155 with a GC content of 29.02% and 1271 predicted genes (Fig. 2). There are 1231 putative
156 protein coding genes and 40 RNA genes, including one each of the 5S, 16S, and 23S rRNA
157 genes. We found no genomic evidence of lysogenic bacteriophage or a CRISPR-*cas* system.
158 LSUCC0530 has the characteristics of genome streamlining previously reported for other
159 SAR11 genomes, with an estimated coding percentage of 96.36% and the smallest complete
160 SAR11 genome reported thus far (Giovannoni *et al.*, 2005b, Grote *et al.*, 2012). Despite its
161 smaller overall genome size, intergenic spacer regions are not significantly smaller than
162 those of the sister subclade IIIa (Fig. S3). All LD12 genomes had much higher AAI and
163 synteny with each other than with the subclade IIIa (Table S1), mirroring their

164 phylogenomic relationships (Fig. 1A). However, within the LD12 clade, AAI and synteny
165 percentages did not clearly delineate the same patterns we observed in the phylogenomic
166 tree, which may have resulted from the LD12 SAGs being fragmented and incomplete.
167 The LSUCC0530 genome has the conserved architecture for hypervariable region 2
168 (HVR2) that has been observed in most other complete SAR11 genomes (Grote *et al.*, 2012,
169 Rodriguez-Valera *et al.*, 2009, Wilhelm *et al.*, 2007) and other LD12 SAGs (Zaremba-
170 Niedzwiedzka *et al.*, 2013), between the 16S rRNA-tRNA-tRNA-23S rRNA operon and the
171 5S rRNA gene operon. This region is 54,755 bp and has a strong variation in GC content
172 from neighboring parts of the genome (Fig. 2). Metagenomic recruitment also verified this
173 region as hypervariable based on the lack of matching sequences in community data (Fig.
174 S4). The predicted gene annotations in HVR2 resemble those in the same location in other
175 SAR11 genomes (Grote *et al.*, 2012, Zaremba-Niedzwiedzka *et al.*, 2013), namely genes
176 likely functioning in membrane modification such as glycosyltransferases, epimerases,
177 transaminases, and methyltransferases; as well as hypothetical proteins.
178
179 *Metabolic reconstruction*
180 Like most other SAR11s (Eiler *et al.*, 2016, Grote *et al.*, 2012, Thrash *et al.*, 2014), the
181 LSUCC0530 genome encodes for an obligate aerobic lifestyle, with a complete oxidative
182 phosphorylation pathway including a heme/copper-type cytochrome c oxidase, a proton-
183 translocating NADH dehydrogenase, and a proton-translocating ATP synthase (Fig. 3). As
184 previously detected in LD12 SAGs (Eiler *et al.*, 2016), the LSUCC0530 genome encodes a
185 complete EMP glycolysis pathway, including phosphofructokinase and pyruvate
186 dehydrogenase, and complete gluconeogenesis through nucleotide sugar production (Fig.
187 3). We recovered genes for the pentose phosphate pathway and a predicted fructokinase
188 gene for conversion of D-fructose to ß-D-fructose-6-P. As expected based on other SAR11
189 genomes, we did not find annotated genes for utilization of allose, fucose, fuculose,
190 galactose, maltose, mannose, rhamnose, sucrose, starch, trehalose, or xylose. LSUCC0530
191 has a complete TCA cycle and glycolate oxidase, and like other SAR11s, the non-mevalonate
192 pathway for isopentenyl diphosphate (Fig. 3).
193 We recovered the first isocitrate lyase in an LD12 genome, conferring a complete
194 glyoxylate bypass in combination with malate synthase. The isocitrate lyase shares its
195 ancestry with the majority of the SAR11 clade except subclade IIIa, where it appears to
196 have been lost (Eiler *et al.*, 2016, Grote *et al.*, 2012) (Fig. S5). However, the LSUCC0530
197 malate synthase shared with other LD12 SAGs belongs to isoform A, whereas all other
198 SAR11 clades contain isoform G (Anstrom *et al.*, 2003) (Figs. 3; S6). Comparison of gene
199 neighborhoods in a variety of SAR11 genomes revealed that both isoforms occur in the
200 same location (Fig. S7).
201 We predict production of eighteen amino acids (Fig. 3) and partial synthesis
202 pathways for three more (histidine, cysteine, and methionine), as well as the ability to
203 interconvert, but not synthesize *de novo*, phenylalanine and tyrosine. The LSUCC0530

genome has lost some of the C1 and methylated compound pathways found in other SAR11s (Sun *et al.*, 2011). For example, as in the LD12 SAGs (Eiler *et al.*, 2016), LSUCC0530 does not have DMSP, methylamine, or glycine-betaine metabolism. However, it retains tetrahydrafolate metabolism, formate oxidation, and the glycine cleavage pathway (Fig. 3).

The genome has the same PII nitrogen response system and *amtB* ammonium transporter found in other SAR11s and is predicted to utilize ammonia as a nitrogen source in amino acid synthesis (Fig. 3). The genome also encodes a putative thiosulfate/3-mercaptopyruvate sulfurtransferase that may be involved in sulfur and thiocyanate reactions (Fig. 3). A possible product of this enzyme is sulfite, which could be further oxidized to sulfate in the periplasm via putative *sorAB* genes. This gene pair occurs infrequently in the SAR11 clade, but sulfite oxidation may serve, in some strains, as a means to generate additional proton motive force (Denger *et al.*, 2008, Kappler, 2011). This is similar to the function conferred by proteorhodopsin (Béjà *et al.*, 2001, Giovannoni *et al.*, 2005a, Steindler *et al.*, 2011), which the LSUCC0530 genome also contains. Also like other SAR11s, a complete assimilatory sulfate reduction pathway is missing, making LSUCC0530 dependent on reduced sulfur compounds for cellular sulfur requirements (Tripp *et al.*, 2008). In our media, this was supplied in the form of sulfur-containing amino acids (Table S1). The genome encodes partial pathways for incorporation of sulfide into cysteine and methionine but is missing predicted serine O-acetyltransferase and homoserine O-succinyltransferase genes (Fig. 3).

In addition to the PII nitrogen sensor, LSUCC0530 also has putative two component systems handling phosphate limitation, pH, osmotic upshift via potassium, and redox state (Fig. 3). The LSUCC genome encodes ABC transporters for phosphate (*ptsABCS*), lipoprotein, and branched-chain and L amino acid uptake. We identified tripartite ATP-independent periplasmic (TRAP) transporter genes for C4-dicarboxylate and mannitol/chloroaromatic compounds. However, the gene cluster appears to be missing the substrate binding subunit, and we did not find predicted genes for mannitol utilization.

In general, LSUCC0530 and the LD12 SAGs share many of the same pathways for osmolyte synthesis as the subclade IIIa and other SAR11 organisms, such as N-acetyl-ornithine, glutamate, glutamine, and proline (Table 1). In some cases, there appear to be analogous substitutions in osmolyte transport systems. For example, LD12 genomes have potassium antiporters, but not the *trkAH* genes for potassium transport in other SAR11s. Similarly, a sodium/proton antiporter orthologus cluster comprised genes exclusively from LD12 genomes, but a separate sodium/proton antiporter occurred in all other SAR11 genomes (Table 1). Perhaps the most important difference in genomic content concerning osmolytes occurs in two key transporter losses: LSUCC0530 and the other LD12s have no genes for the glycine betaine/proline or ectoine/hydroxyectoine ABC transporters that are present in HIMB114 and IMCC9063 and other subclades of SAR11 (Table 1).

Physiology

244 Under isolation conditions (23°C, JW5 medium), strain LSUCC0530 grew to a density of 2 x
245 10^7 cells mL^{-1} with an average growth rate of 0.52 day^{-1}. Our physiological experiments to
246 determine the salinity tolerance and growth optima revealed that LSUCC0530 could grow
247 at salinities between 0.36 and 4.7 (Fig. 4A). The optimal salinity for LSUCC0530 was at or
248 below the isolation medium of 1.45. Growth rates decreased between salinities of 2.9 and
249 4.7 (Fig. 4A), and cells died (determined by rapid loss of flow cytometric signal) in salinities
250 above 8 (Fig. S8). We observed growth after a very extended lag phase in a subset of
251 cultures at salinity 5.8, but could not always reproduce this growth in repeated
252 experiments (Fig. S8). Among the temperatures tested, LSUCC0530 grew optimally at 23˚C,
253 and we observed growth up to 30˚C, but not at 12˚C or below, nor at 35˚C or above (Fig.
254 4B). However, we did not observe a clear loss of signal at 4, 12, 35, or even 40˚C (Fig. S8),
255 even after over 33 days, suggesting that LSUCC0530 can endure extended periods at these
256 temperatures in a non-growth state. Notably during growth experiments, one of the five
257 replicates during any given experiment frequently grew at a reduced rate than the others
258 (Fig. S8).
259
260 *Ecology*
261 We quantified the abundance of LD12 taxa in the coastal Louisiana ecosystems using 16S
262 rRNA gene sequences clustered into operational taxonomic units (OTUs). These sequences
263 were obtained as part of the combined sampling and HTC campaign that included the
264 successful isolation of LSUCC0530. Based on BLASTn sequence similarity, the
265 representative sequence for OTU7 and LSUCC0530 shared 100% identity over the 250bp
266 v4 region of the 16S rRNA gene, confirming the Mothur classification of OTU7 as belonging
267 to the SAR11 LD12 clade. Over the three-year sampling period, OTU7 was the fifth most
268 abundant OTU based on total read abundance and sixth ranked OTU based on median read
269 abundance. At the time of sampling for the experiment that resulted in the isolation of
270 LSUCC0530, OTU7 had a relative abundance of 8.7% in Lake Borgne. In some sites, such as
271 the Atchafalaya River Delta (salinity 0.18), LD12 represented as much as 16% of total
272 bacterioplankton (Fig. S9). Generally, the LD12 OTU7 relative abundance correlated
273 negatively with salinity. At seven sites with salinities below 6, OTU7 had a relative
274 abundance of 4.0% or greater with a median relative abundance of 8.8% (Fig. S9). At the
275 remaining ten sites above a salinity of 6 (median 19.44), OTU7 had relative abundances
276 around 2% or lower (Fig. S9).
277 The phylogenetic separation of LSUCC0530 from B11 and the IIIb A-C microclusters
278 raised the question of whether LSUCC0530 represents a novel ecotype with unique
279 spatiotemporal distribution compared to other LD12 taxa. Therefore, we explored the
280 distribution of LD12 genomes in 90 fresh and brackish water environments through
281 competitive recruitment of metagenomic sequences to all the SAR11 genomes in the
282 phylogenomic tree and the outgroup sequence HIMB59. We quantified relative abundance
283 via recruited reads (per kpb of metagenomic sequence per Mbp of genome sequence-

284 RPKM), and aggregated LD12 data based on the A, B, and C microclusters (Fig. 1A), the B11
285 SAG, and LSUCC0530. Figure 5A displays RPKM values for reads with ≥ 95% identity,
286 roughly corresponding to a species delineation via average nucleotide identity
287 (ANI)(Konstantinidis and Tiedje, 2005), but we also provide additional comparisons at ≥
288 85, 90, 92, 98, and 100% identity (Fig. S10).
289 All LD12 subgroups had their highest RPKM values in very low salinity sites,
290 although we could detect LD12 across various brackish salinities (Fig. 5A). In general, IIIb.A
291 and LSUCC0530 averaged the highest relative abundances across all sites with salinities
292 between 0 and 5.75. All subgroups showed site-specific variation in abundances (Fig. 5A).
293 Microcluster IIIb.A clearly dominated recruitment at Lake Michigan, whereas LSUCC0530
294 represented the most abundant genome at Feitsui Reservoir and Lake Gatun (Fig. 5A).
295 We repeated the previous examination of recruitment density to the different
296 microclusters across a range of percent identities, now including LSUCC0530 (Garcia *et al.*,
297 2017). We corroborate observed patterns indicating additional LD12 diversity beyond that
298 represented in the current genomes, despite the inclusion of our new representative. For
299 example, whereas the LSUCC0530 genome recruited the majority of high percent identity
300 reads in the Feitsui Reservoir and Lake Gatun samples (Fig. 5B), in some locations the
301 majority of recruitment to this genome centered around lower percent identity hits (Fig.
302 5C). We observed the same kind of alternative patterns for the other microclusters as well
303 (Fig. 5D). As noted our colleagues, since reads were competitively recruited against all
304 close relative genomes, these cases where a microcluster primarily recruited reads ~90-
305 92% identity likely mean that we lack a representative genome for that fraction of the
306 metagenomic data (Garcia *et al.*, 2017).
307
308 **Discussion**
309 Our successful cultivation of an LD12 strain for the first time has demonstrated that this
310 organism shares many physiological traits with cultured marine SAR11 representatives,
311 and also has some important distinctions. LSUCC0530 doubles roughly once every two
312 days, similar to taxa in subclade Ia (Rappé *et al.*, 2002) and Ib (Jimenez-Infante *et al.*, 2017).
313 It has an aerobic, chemoorganoheterotrophic lifestyle, thriving in oligotrophic media with
314 simple carbon compounds and likely depends on reduced organosulfur compounds and
315 ammonium as its sulfur and nitrogen sources, respectively. Cell size and shape showed
316 considerable morphological conservation (small, curved rods roughly 1 x 0.1 μm) across
317 large evolutionary distances (LD12 vs. subclade Ia ~89% 16S rRNA gene identity).
318 Although a systematic survey of salinity tolerance has not been conducted for other
319 cultured SAR11 representatives, we have shown that LD12 cannot survive in media with
320 salinities over eight, whereas all other cultured SAR11 grow in seawater salinities (Carini *et*
321 *al.*, 2013, Jimenez-Infante *et al.*, 2017, Rappé *et al.*, 2002, Song *et al.*, 2009, Stingl *et al.*,
322 2007). LSUCC0530 thrives in relatively high temperatures (23-30˚C), and we could not

323 confirm growth at low temperatures in which other SAR11 have been isolated (Song *et al.*,
324 2009).
325 LSUCC0530 shares many genomic characteristics with marine SAR11, such as a
326 small, streamlined genome with few paralogs, short intergenic spaces, and a conserved
327 hypervariable region, confirming predictions from LD12 SAGs and metagenomes (Eiler *et*
328 *al.*, 2016, Newton *et al.*, 2011, Zaremba-Niedzwiedzka *et al.*, 2013). The LSUCC0530
329 genome also encodes much of the same metabolic potential as what has been reported
330 through LD12 SAGs and metagenomic data. For example, we confirm the EMP glycolysis
331 pathway, TCA cycle, non-mevalonate terpenoid biosynthesis, numerous amino acid
332 synthesis pathways, TRAP transporters for C4 dicarboxylate compounds, and ABC
333 transporters for various amino acids. We can confirm the lack of a phosphoenolpyruvate
334 carboxylase (ppc) and missing pathways for DMSP, glycine-betaine, and methylamine
335 metabolism.
336 We have also recovered important genetic information previously unknown for
337 LD12. At 1.16 Mbp, LSUCC0530 has a smaller genome, with fewer genes, than any other
338 SAR11. LSUCC0530 has a complete glyoxylate bypass, which differs from previous
339 inferences using SAGs (Eiler *et al.*, 2016). The unusual LD12 malate synthase isoform A
340 gene occupies the same genomic location in LSUCC0530 as the isoform G copy in most
341 subclade Ia, Ic, and IIIa genomes, with the exception of HIMB114 (Fig. S7). Therefore, the
342 most parsimonious explanation for the presence of isoform A in LSUCC0530 is a
343 homologous recombination event that replaced the isoform G homolog. The LSUCC0530
344 genome also contains putative pathways for oxidation of sulfide, thiosulfate, and sulfite,
345 which have not yet been explored in any cultured SAR11.
346 The relative abundance data from coastal Louisiana and the isolation of an LD12
347 strain from a coastal environment supports previous work indicating that LD12 occupies a
348 much wider range of aquatic habitats than inland freshwater systems (Dupont *et al.*, 2014,
349 Herlemann *et al.*, 2014, Piwosz *et al.*, 2013, Salcher *et al.*, 2011). Questions have been raised
350 as to whether these organisms occur in brackish waters as active and growing community
351 members, or whether there are present simply as a result of freshwater input. LD12 cells
352 have been shown to uptake thymidine in brackish water (Piwosz *et al.*, 2013). Our
353 observation that LSUCC0530 can grow in slightly brackish water provides further evidence
354 that such habitats may support LD12 cells as active community members. However, the
355 relative competitiveness of LD12 within brackish microbial assemblages requires further
356 research.
357 Metagenomic recruitment provides evidence that LD12 has diversified into multiple
358 ecotypes with unique distributions (Fig. 5A) (Garcia *et al.*, 2017, Zaremba-Niedzwiedzka *et*
359 *al.*, 2013). Furthermore, our data suggests that LSUCC0530 likely represents a novel LD12
360 ecotype because it preferentially recruits metagenomic sequences at high percent identity
361 in different locations than the other microclusters of LD12 (Fig. 5). What distinguishes the
362 distribution of these ecotypes remains unclear. The observed patterns may represent

363 temporal variation captured in these sampling snapshots (Garcia *et al.*, 2017). Another
364 possibility is that small variations in salinity optima for different ecotypes may drive
365 diversification. However, given the paucity of low-brackish sites in available datasets, we
366 cannot establish whether salinity plays a role in ecotype diversification. Furthermore,
367 without complete genomes from other LD12 ecotypes, we cannot rule out whether variant
368 ionic strength management strategies, or some other metabolic characteristic, help define
369 evolutionary differences associated with ecotypes. For example, the ability to utilize
370 different sulfur sources may distinguish LSUCC0530 from other LD12. The *sorAB* sulfite
371 dehydrogenase and the thiosulfate/3-mercaptopyruvate sulfurtransferase genes have
372 uneven distribution among SAR11 genomes, including within the LD12 subclade.
373 An alternative hypothesis is that temperature has driven diversification between
374 ecotypes. A striking feature of LSUCC0530 is its growth preference for relatively high
375 temperatures and inability to grow at low temperatures. While robust growth at 30˚C may
376 seem intuitive given that this represents typical Lake Borgne surface water temperature,
377 the inability to grow at 12˚C or below would likely place LSUCC0530-type LD12 cells at a
378 fitness disadvantage in places like Lake Michigan, where annual water temperatures rarely
379 get above 22˚C and average temperatures remain below 15˚C for the majority of the year
380 (https://coastwatch.glerl.noaa.gov/). The metagenomic recruitment data support the
381 temperature hypothesis: the samples in which LSUCC0530 dominated recruitment (Lake
382 Gatun, Feitsui Reservoir) were all collected from locations having average temperatures
383 above 19˚C. Conversely, other microclusters dominated recruitment of samples from colder
384 environments (e.g. IIIb.A in Lake Michigan)(Fig. 5). More sequence data may strengthen the
385 relationship between temperature and ecotypes, but since temperature optima are
386 extremely hard to infer from genomes alone, additional cultivars will be needed to test the
387 temperature hypothesis for ecotype diversification.
388 A critical question remains for LD12. What has made these cells uniquely adapted to
389 low salinity waters? Genomic analyses here and elsewhere (Dupont *et al.*, 2014, Eiler *et al.*,
390 2016) have helped define the differences in LD12 metabolic potential compared to other
391 SAR11, but very few adaptations appear related to salinity. It has been noted that the EMP
392 pathway provides more ATP and NADH than the ED pathway, thus conferring greater
393 energy conservation through glycolysis for LD12 cells than other SAR11s (Dupont *et al.*,
394 2014, Eiler *et al.*, 2016). If or how this energy conservation relates to salinity adaptation
395 has not been pursued. The presence of unique LD12 cation-proton antiporters (Fig. 3)
396 could provide a connection. LD12 cells may benefit from additional NADH for the
397 production of a proton motive force that drives the potassium-proton antiporters, thus
398 improving active transport to lower internal ionic strength. However, while this adaptation
399 may improve the ability of LD12 cells to live in fresh water, it doesn't explain why they
400 have lost the ability to survive in salt water where their marine SAR11 cousins thrive.
401 The most likely explanation comes in the distribution of genes for osmolyte uptake.
402 Critically, LSUCC0530, and LD12 in general, have lost the *proVWX* glycine-betaine/proline

403 ABC transporter shared by all other marine SAR11 (Table 1). Although a relatively minimal
404 number of genes, this uptake capability likely represents a significant loss of function for
405 LD12 salinity tolerance. In *Escherichia coli*, the *proVWX* operon has a strong promoter that
406 upregulates these genes two orders of magnitude in response to increased salinity
407 (Dattananda and Gowrishankar, 1989). Indeed, rapid proline uptake is a common response
408 to osmotic upshift in many bacteria (Empadinhas and da Costa, 2008). Furthermore,
409 LSUCC0530 and other LD12 genomes have lost the ABC transporters for
410 ectoine/hydroxyectoine present in the other SAR11 genomes. Given the slow growth of
411 LD12, rapid production of new intracellular osmolytes in response to increased salinity
412 seems unlikely. Thus, the loss of these transport systems for quickly equilibrating
413 intracellular osmolarity probably prevents successful dispersal of this organism into higher
414 salinity waters.
415 This finding allows us to speculate further about the evolutionary trajectory that led
416 to successful colonization of freshwater environments by LD12. All evidence suggests that
417 the last common SAR11 ancestor was a marine organism with a streamlined genome
418 (Logares *et al.*, 2010, Luo *et al.*, 2013, Zaremba-Niedzwiedzka *et al.*, 2013). During further
419 diversification into subclade III, additional genes and pathways were lost, for example, ppc
420 and genes for DMSP and some C1 compound metabolisms (Eiler *et al.*, 2016, Sun *et al.*,
421 2016). Therefore, the common ancestor of subclade IIIa and LD12 would have had to adapt
422 to freshwater environments starting with an already reduced genome and limited
423 metabolic flexibility. A simple way to inhabit lower salinity environments would have been
424 to decrease expression of, or altogether lose, genes for uptake and/or production of
425 osmolytes. If the adaptation were reversible, we would expect to see examples of LD12
426 organisms growing in marine systems. However, existing evidence suggests that while
427 LD12 organisms can grow (established herein) and be active (Piwosz *et al.*, 2013) in low
428 brackish conditions, they are not successful in higher salinity brackish or marine water
429 (Dupont *et al.*, 2014, Herlemann *et al.*, 2014, Logares *et al.*, 2010, Salcher *et al.*, 2011). Thus,
430 we propose that LD12 adaptation to fresh water involved irreversible loss of function,
431 which decreased osmotic response capability in LD12 cells, thereby preventing
432 recolonization of higher salinity environments. The missing ABC transporters for
433 proline/glycine betaine and ectoine/hydroxyectoine in LD12 are likely candidates for this
434 loss of function. Experimental tests on the differences in production and uptake of
435 osmolytes among subclade IIIa and LD12 representatives will provide critical insight into
436 how these different evolutionary paths have resulted in unique salinity ranges.
437
438 For the first cultured representative of the LD12 clade, we propose the provisional
439 taxonomic assignment for strain LSUCC0530 as '*Candidatus* Fonsibacter ubiquis',
440
441 **Description of *Fonsibacter* gen. nov.**

442 *Fonsibacter* (*Fons* L. noun; fresh water, spring water, -bacter, Gr. adj.; rod, bacterium.
443 Fonsibacter referring to the preferred low salinity habitat of this organism).
444
445 Aerobic, chemoorganoheterotrophic, and oligotrophic. Cells are small, curved rods roughly
446 1 x 0.1µm. Non-motile. Based on phylogenomics and 16S rRNA gene phylogenetics,
447 subclade IIIb/LD12 occurs on a separate branch within the *Pelagibacteraceae* (SAR11),
448 sister to subclade IIIa containing strains HIMB114 and IMCC9063. Due to the depth of
449 branching between these clades using concatenated protein-coding genes, and the 92%
450 and 91% 16S rRNA gene sequence identity with HIMB114 and IMCC9063, respectively, we
451 propose LSUCC0530 and subclade IIIb/LD12 as a novel genus in the *Pelagibacteraceae*.
452
453 **Description of *Fonsibacter ubiquis* sp. nov.**
454 *Fonsibacter ubiquis* (L. adv. ubiquitous).
455
456 In addition to the properties given in the genus description, the species is described as
457 follows. Growth occurs at temperatures between 23˚C and 30˚C, but not at 11˚C or below,
458 nor at 35˚C or above. Optimal salinity is 1.45 and below, and growth occurs between 0.36
459 and 4.68. At optimal temperature and salinity, cells divide at an average rate of 0.52 day^{-1}.
460 Genome size is 1.16 Mbp, with 1271 predicted genes, and a GC content of 29.02%
461 (calculated). LSUCC0530 had an ANI of 75.3% and 75.9% and an AAI of 58.1% and 59.7%
462 with HIMB114 and IMCC9063, respectively. The genome is available on GenBank under
463 accession number CP024034.
464
465 The type strain, LSUCC0530^{T}, was isolated from brackish water from the coastal lagoon
466 Lake Borgne off the coast of southeastern Louisiana, USA.
467
468
469 **Materials and Methods**
470 *Isolation and Identification of LSUCC0530*
471 Isolation, propagation, and identification of strain LSUCC0530 (subclade IIIb) was
472 completed as previously reported (Henson *et al.*, 2016). Briefly, water was collected from
473 the surface of the coastal lagoon Lake Borgne (Shell Beach, Louisiana) (29.872035 N, -
474 89.672819 W) on 1 July, 2016. We recorded salinity on site using a YSI 556 MPS (YSI, Solon,
475 OH, USA) (Table S1). Whole water was filtered through a 2.7 µm filter (Whatman), stained
476 with 1X SYBR Green (Lonza, Basal, Switzerland), and enumerated using the Guava EasyCyte
477 5HT flow cytometer (Millipore, Massachusetts, USA). After serial dilution, water was
478 inoculated into five 96 x 2.1 mL well PTFE plates (Radleys, Essex, UK) containing 1.7 mL of
479 JW5 medium at an estimated 2 cells $well^{-1}$. Cultures were incubated at room temperature in
480 the dark for three weeks and evaluated for growth using flow cytometry. Positive wells (>
481 10^4 cells mL^{-l}) were transferred in duplicate to capped, 125 mL polycarbonate flasks

482 (Corning, New York, USA) containing 50 mL of JW5 medium. Upon reaching a density of >
483 5.0 x 10^5 cells mL^{-1}, cells were filtered onto 25 mm 0.22 µm polycarbonate filters (Millipore,
484 Massachusetts, USA), and DNA extractions were performed using the MoBio PowerWater
485 DNA kit (QIAGEN, Massachusetts, USA) following the manufacturer's instructions. DNA was
486 amplified as previously reported(Henson *et al.*, 2016) and sequenced at Michigan State
487 University RTSF Genomics Core. Sanger sequences were evaluated using the freely
488 available software 4Peaks (v. 1.7.1)(http://nucleobytes.com/4peaks/). Forward and
489 reverse sequences were assembled using the CAP3 webserver
490 (http://doua.prabi.fr/software/cap3), after reverse reads were converted to their reverse
491 complement at http://www.bioinformatics.org/sms/rev_comp.html. The PCR-generated
492 16S rRNA gene sequence for strain LSUCC0530 is accessible on NCBI GenBank under the
493 accession number KY290650.
494
495 *Community iTag analysis*
496 Community DNA from 17 sites of various salinities was filtered, extracted, and analyzed
497 following our previously reported protocol (Henson *et al.*, 2016). The sites sampled were
498 Lake Borgne (LKB; Shell Beach, LA), Bay Pomme D'or (JLB; Buras, LA), Terrebonne Bay
499 (TBon; Cocodrie, LA), Atchafalaya River Delta (ARD; Franklin, LA), Freshwater City (FWC;
500 Kaplan, LA), Calcasieu Jetties (CJ; Cameron, LA). Each site was visited three times with the
501 exception of TBon, which was visited only twice. Briefly, extracted DNA was sequenced at
502 the 16S rRNA gene V4 region (515F, 806R) (Parada *et al.*, 2015), with Illumina MiSeq
503 2x250bp paired-end sequencing at Argonne National Laboratories. 16S rRNA gene
504 amplicon data was analyzed and classified ($OTU_{0.03}$) with Mothur v.1.33.3 (Schloss *et al.*,
505 2009) using the Silva v119 database (Pruesse *et al.*, 2007). To determine the best matching
506 OTU for LSUCC0530, sequences from the OTU representative fasta files, provided by
507 mothur using *get.oturep*(), were used to create a blast database (*formatdb*) against which
508 the LSUCC isolate 16S rRNA genes could be searched via blastn (BLAST v 2.2.26). All best
509 hits ranked at 100% identity. OTU abundance analysis was conducted within the R
510 statistical environment v3.2.3, within the package PhyloSeq (v1.14.0) (McMurdie and
511 Holmes, 2013). Sequences were rarefied using the command *rarefy_even_depth*() and then
512 trimmed for OTUs with at least 2 reads in at least 90% of the samples. Abundances of an
513 OTU were averaged across biological duplicates and a rank abundance matrix was
514 calculated using the command *transform_sample_counts*() with the argument function of
515 function(x) x / sum(x). Completed rarefied rank abundance OTU tables were plotted using
516 the graphing program GGPLOT2 (v. 2.0.0) (Wickham, 2011). All iTag sequences are
517 available at the Short Read Archive with accession number SRR6235382-SRR6235415.
518
519 *Growth experiments*
520 Salinity tolerance was tested by growing LSUCC0530 in different media distinguished by
521 proportional changes in the concentration of major ions. All other media elements (carbon,

522 iron, phosphate, nitrogen, vitamins, and trace metals) were kept consistent. Media included
523 JW1, JW2, JW3, and JW4, previously published (Henson *et al.*, 2016), and JW3.5, JW4.5, JW5,
524 and JW6. Recipes for all media used in this study are provided in Table S1. Salinity
525 tolerance experiments were conducted in quintuplicate at room temperature. Temperature
526 range and sole carbon substrate use were tested using the isolation medium, JW5.
527 Temperature range experiments were also conducted in quintuplicate. For all experiments,
528 growth was measured using flow cytometry as described above.
529
530 *Microscopy*
531 Strain LSUCC0530 cells were grown to near max density ($1x10^7$ cells mL^{-1}) in JW5 medium.
532 Cells for transmission electron microscopy were prepared as previously described (Rappé
533 *et al.*, 2002) with one minor change: centrifuged cells were concentrated in JW5 medium
534 with no added fixative. Cells were visualized under a JEOL JSM-2011 High Resolution
535 transmission electron microscope at the LSU Socolofsky Microscopy Center, Baton Rouge,
536 LA. For scanning electron microscopy, cultured cells were fixed for four hours in 4%
537 Glutaraldehyde buffered with 0.2M Cacodylate buffer and filtered onto a 0.2 µm nylon filter
538 (Pall, Michigan, USA). Samples were dehydrated in an ethanol series (30%, 50%, 70%, 96
539 and 100%) for 15 minutes each, followed by critical point drying. Cells were visualized
540 under a JEOL JSM-6610 scanning electron microscope at the LSU Socolofsky Microscopy
541 Center, Baton Rouge, LA.
542
543 *Genome sequencing, assembly, and annotation*
544 Cells were grown in 1 L JW5 medium (Table S1) and filtered onto 0.2 µm nylon filters (Pall,
545 Michigan, USA). DNA was obtained via phenol-chloroform extraction and sequenced using
546 both Illumina HiSeq and MiSeq approaches. *HiSeq*. 350 ng of DNA was sheared to ~500 bp
547 with a Qsonica Q800R (QSonica) using 16 cycles with a amplitude of 25% and pulse rate of
548 20 seconds on, 20 seconds off. DNA was visualized on a 1.5% agrose gel stained with Gel
549 Red (Biotum) to determine average sheared length and quantified using Qubit high
550 sensitivity dsDNA assay kit (Invitrogen). Library prep was conducted following a modified
551 Kapa Hyper Prep Kit protocol (Kapa Biosystems, Inc.). Briefly, we ligated end-repaired and
552 adenylated DNA to universal Y-yoke oligonucleotide adapters and custom Illumina iTru
553 dual-indexed primers (Glenn *et al.*, 2016). Prior to index amplification, post-ligation
554 LSUCC0530 DNA was cleaned using 1.2x SPRI (Roland and Reich, 2012). Following a brief,
555 8 cycle PCR amplification of the library to add indexes, the genomic library was quantified
556 using a Qubit broad range dsDNA assay kit (Invitrogen). Illumina HiSeq 3000 sequencing
557 was performed at Oklahoma Medical Research Facility using SBS kit chemistry to generate
558 150 bp paired-end reads, insert size 400 bp. *MiSeq*. DNA was sent to the Argonne National
559 Laboratory Environmental Sample Preparation and Sequencing Facility which performed
560 library prep and sequencing, generating 250 bp paired-end reads, insert size 550 bp.

561 Due to the large number of reads from the HiSeq sequencing, we created a random
562 subset using 14% of the total reads by using the program *seqtk* (v. 1.0-r75-dirty) and the
563 flag –s100. We trimmed reads for adapter contamination and quality using Trimmomatic
564 (Bolger *et al.*). Initial assembly with the subset of reads was done with SPAdes (Bankevich
565 *et al.*, 2012). This generated a large single scaffold with overlapping ends. We performed a
566 quality assessment of the single scaffold via iterations of Reapr (Hunt *et al.*, 2013)-advised
567 scaffold breaks and SSPACE (Boetzer *et al.*, 2011) extensions using both the full set of HiSeq
568 reads and the MiSeq reads. Final assessment of the scaffold with overlaps removed from
569 the ends was performed with Pilon (Walker *et al.*, 2014) using all *HiSeq* reads mapped to
570 the scaffold with BWA (Li and Durbin, 2009). No issues were reported. An analysis of the
571 GC skew was performed on the final scaffold using the gc_skew script provided via Brown
572 et al. 2015 (Brown *et al.*, 2015). Contamination and completion were also evaluated with
573 CheckM (Parks *et al.*, 2015). Detailed commands and outputs for the assembly and quality
574 assessment are provided in Supplementary Information. The final circular chromosome
575 scaffold was annotated by IMG (Markowitz *et al.*, 2014) and is publicly available with IMG
576 Taxon ID number 2728369501. It is also publicly available in GenBank with accession
577 number CP024034. Genome sequencing reads are available in SRA with accession numbers
578 SRR6257553 (MiSeq) and SRR6257608 (HiSeq).
579
580 *Comparative genomics*
581 Orthologus clusters were determined using the Get Homologues pipeline (Contreras-
582 Moreira and Vinuesa, 2013) using the LSUCC0530 genome and 48 publicly available SAR11
583 genomes from IMG, including the 10 LD12 SAGs analyzed previously (Eiler *et al.*, 2016,
584 Zaremba-Niedzwiedzka *et al.*, 2013) (Table S1). Clusters were determined from amino acid
585 sequences using the OrthoMCL option in Get Homologues. All clusters are provided in
586 cluster_list.txt as part of Supplemental Information. Average amino acid identity (AAI) and
587 synteny were calculated as previously reported (Dick *et al.*, 2009). Average nucleotide
588 identity (ANI) was calculated with the Konstantinidis Lab website ANI calculator
589 (http://enve-omics.ce.gatech.edu/ani/index) as reported (Goris *et al.*, 2007). Compatible
590 solutes identified in Table S1 were identified in SAR11 genomes using KO numbers,
591 annotations, and cross-referenced with orthologous clusters.
592
593 *Bacteriophage searches*
594 We looked for signatures of bacteriophage in the LSUCC0530 genome using two methods,
595 VirSorter (Roux *et al.*, 2015) and PHASTER (Arndt *et al.*, 2016) with default settings.
596 CRISPR regions are identified as part of the standard IMG annotation process, and none
597 were found in LSUCC0530.
598
599 *Phylogenomics*

600 We selected single-copy genes present in at least 39 of the 49 genomes as determined via
601 Get Homologues (above). Amino acid sequences from these 83 markers were separately
602 aligned with MUSCLE (Edgar, 2004), culled with Gblocks (Castresana, 2000), and
603 concatenated into a superalignment totaling 23,356 positions. Inference was performed
604 with RAxML (Stamatakis *et al.*, 2008) using the PROTCATLG model and 100 bootstrapping
605 runs. All scripts (with settings) used in the alignment, culling, concatenation, and inference
606 processes are available in Supplemental Information. Node labels were renamed with
607 Newick Utilities (Junier and Zdobnov, 2010) *nw_rename*, and the tree was visualized with
608 Archaeopteryx (Han and Zmasek, 2009).
609
610 *Single gene phylogeny*
611 The 16S rRNA gene sequence from the LSUCC0530 genome was aligned with 284
612 alphaproteobacterial sequences and seven outgroup sequences. We completed
613 phylogenetic inference for *aceA* and *malA* in an analogous manner. For each, the
614 LSUCC0530 gene was searched against the NCBI nr database using blastp (v. 2.2.28). The
615 top 100 sequence hits were combined with all homologus SAR11 sequences identified with
616 Get Homologues for the *aceA* phylogeny, and both the *malA* and *malG* sequences in all
617 SAR11s for the *malA* phylogeny. Sequences from redundant taxa were removed. For all
618 trees, alignment with MUSCLE (Edgar, 2004), culling with Gblocks (Castresana, 2000), and
619 tree inference with FastTree2 (Price *et al.*, 2010) was completed with the FT_pipe script,
620 provided in Supplemental Information, along with the initial fasta file containing all
621 sequences and the newick tree output file. Node labels were changed with Newick Utilities
622 (Junier and Zdobnov, 2010) *nw_rename* and visualization was performed with
623 Archaeopteryx (Han and Zmasek, 2009).
624
625 *Metagenomic recruitment*
626 Scaffolds for the 49 SAR11 genomes, including strain LSUCC0530, used in the
627 phylogenomic analyses were used for competitive recruitment of metagenomic sequences
628 from 89 different samples. Prior to recruitment, multi-scaffold assemblies were condensed
629 into single scaffolds, and all single scaffolds were used to generate a BLAST database with
630 the command *makeblastdb* (BLAST v. 2.2.28). Metagenomic datasets from 90 different
631 environmental samples of various salinities from the Baltic Sea (Dupont *et al.*, 2014),
632 Amazon River, freshwater lakes (Eiler *et al.*, 2014, Martinez-Garcia *et al.*, 2011), Lagoon
633 Albufera (Ghai *et al.*, 2012), Columbia River (Fortunato and Crump, 2015, Smith *et al.*,
634 2013b), Lake Lanier (Oh *et al.*, 2011), Lake Michigan (Denef *et al.*, 2016), Feitsui Reservoir
635 (Tseng *et al.*), and GOS (Rusch *et al.*, 2007) were downloaded from the Short Read Archive
636 (SRA) and European Nucleotide Archive (ENA) using *ftp* and SRA toolkit (v.2.8.2-1),
637 respectively. Reads from pyrosequencing were dereplicated with the program CD-HIT (v
638 4.6) (Fu *et al.*, 2012) using the *cd-hit-454* command and flag -c 0.95. Illumina sequence data
639 was converted from fastq to fasta files with the FastX toolkit (v. 0.0.13.2)

640 (http://hannonlab.cshl.edu/fastx_toolkit) using the command *fastq_to_fa*sta with the flag –
641 Q33. Metagenomic sequences were separated into 1 million sequence files and recruited
642 against the SAR11 database with blastn (v. 2.2.28) with –max_target_seqs 1. Best hits for
643 each metagenomic sequence were kept if the alignment length was > 90% of the median
644 length of all metagenomic sequences for a given sample. Metagenomic recruitment was
645 expressed in read density per genome using ggplot2 *geom_density*() and reads per kilobase
646 of genome per million mapped reads (RPKM) as previously reported (Thrash *et al.*, 2017).
647
648 *Script availability*
649 All scripts used in this work can be found on the Thrash Lab website with the manuscript
650 link at http://thethrashlab.com/publications.
651
652 **Acknowledgements**
653 Funding for this work was provided in part by a Louisiana Board of Regents (grant
654 LEQSF(2014-17)-RD-A-06) grant to JCT and the LSU Department of Biological Sciences
655 (startup funds to JCT and BCF). The authors thank the LSU Socolofsky Microscopy Center
656 for assistance with electron microscopy imaging. Portions of this research were conducted
657 with high performance computing resources provided by Louisiana State University
658 (http://www.hpc.lsu.edu).
659
660 **Author Contributions**
661 JCT designed the study; MWH conducted cultivation experiments, nucleic acid extraction,
662 microscopy, and iTag and metagenomics analyses; JCT and BCF assembled the genome and
663 conducted quality assessment; MWH, VCL, and JCT conducted comparative genomics
664 analyses, metabolic reconstruction, phylogenetic, and phylogenomic analyses; JCT lead
665 manuscript writing and all authors contributed to and reviewed the text.
666
667 **Competing Financial Interests**
668 The authors declare no competing financial interests.
669
670
671

672 **Tables and Figures**

673

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

675

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	-	-	-	-	+

676

677 + genes present; - genes absent; +/- partial pathway; *B11 LD12 SAG only. Table shows
678 only systems with predicted genes in at least one SAR11 genome. Table only shows
679 systems for which genes could be identified in the SAR11 genomes examined in this study.
680 For a complete list of all systems searched, see Table S1.

681

682 **Figure Legends**

683

684 **Figure 1.** Phylogeny and morphology of LSUCC0530. **A)** Phylogenetic position of
685 LSUCC0530 within the SAR11 clade using a concatenation of 83 single copy marker genes
686 and 23,356 amino acid positions. The tree was inferred using RAxML with 100 bootstrap
687 runs. Values at the nodes indicate 100% bootstrap support unless otherwise noted. The
688 tree was rooted on HIMB59. Subclades are indicated on the right, and LD12 microclusters
689 are indicated A-C. Asterisk- we have included AAA028-C07 in microcluster A, even though
690 it was excluded previously (Zaremba-Niedzwiedzka *et al.*, 2013). **B)** Scanning electron
691 micrograph of LSUCC0530 cells at 35,000x magnification. Scale bar represents 0.5 μm. **C)**
692 Transmission electron micrograph of a LSUCC0530 cell thin section. Scale bar represents

693 0.5 μm.
694
695 **Figure 2.** Circular diagram of the LSUCC0530 genome. Rings indicate, from outside to the
696 center: Position relative to the replication start site; Forward strand genes, colored by COG
697 category; Reverse strand genes, colored by COG category; RNA genes (tRNAs green, rRNAs
698 red, other RNAs black); GC content; GC skew.
699
700 **Figure 3.** Metabolic reconstruction of LSUCC0530. Solid boxes indicate recovered genes,
701 dashed boxes indicate missing genes. Color shading distinguishes major metabolic sub-
702 systems, which are also labeled in bold text. ABC transporters are depicted with circles
703 indicating subunits. Symporters and antiporters are depicted with ovals. Two component
704 systems are depicted with large and small rectangles. The C4 TRAP transporter is depicted
705 separately. Major subsystems are colored for easier identification. Multiple arrows indicate
706 all genes present in a given pathway. Numbers identify genes according to the key in Table
707 S1. Question marks indicate a single missing gene in an otherwise complete pathway (e.g.,
708 PRPP -> His). Light blue fill indicates genes with no SAR11 orthologs outside of the LD12
709 subclade.
710
711 **Figure 4.** Growth rates for LSUCC0530 according to **A)** salinity and **B)** temperature. The
712 boxes indicate the interquartile range (IQR) of the data, with vertical lines indicating the
713 upper and lower extremes according to 1.5 x IQR. Horizontal lines within each box indicate
714 the median. The underlying data points of individual biological replicates, calculated from
715 corresponding growth curves in Figure S8, are plotted on top of each box.
716
717 **Figure 5.** Results of competitive metagenomic recruitment to all genomes used in the
718 phylogenomic tree. **A)** 95% identity RPKM values for LD12 genomes only, by site, with data
719 aggregated for all genomes in microclusters A-C, according to the key. Colors indicate broad
720 environmental categories. **B-D)** Density plots of metagenomic sequence recruitment
721 according to percent identity of the best hit. Data was aggregated according to the key, with
722 LD12 separated as in A, HIMB59 separately depicted as "V", and all other SAR11 genomes
723 together as "Other." **B)** Two sites where LSUCC0530 dominated recruitment with high
724 percent identity. **C)** Two sites where the LSUCC0530 only recruited sequences at ~90%
725 identity and no other group had high recruitment at higher identity. **D)** Two sites where
726 other LD12 microclusters showed similar patterns as LSUCC0530 in B and C.
727
728

729 **References Cited**
730

731 Anstrom DM, Kallio K, Remington JS (2003). Structure of the Escherichia coli malate
732 synthase G:pyruvate:acetyl‐coenzyme A abortive ternary complex at 1.95 Å resolution.
733 *Protein Science* **12:** 1822-1832.
734
735 Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y *et al.* (2016). PHASTER: a better, faster
736 version of the PHAST phage search tool. *Nucleic Acids Research* **44**.
737
738 Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS *et al.* (2012). SPAdes:
739 A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal*
740 *of Computational Biology* **19:** 455-477.
741
742 Béjà O, Spudich EN, Spudich JL, Leclerc M, DeLong EF (2001). Proteorhodopsin
743 phototrophy in the ocean. *Nature* **411:** 786-789.
744
745 Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2011). Scaffolding pre-assembled
746 contigs using SSPACE. *Bioinformatics* **27:** 578-579.
747
748 Bolger AM, Lohse M, Usadel B (2014). Trimmomatic: a flexible trimmer for Illumina
749 sequence data. *Bioinformatics* **30:** 2114-2120.
750
751 Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A *et al.* (2015). Unusual biology
752 across a group comprising more than 15% of domain Bacteria. *Nature* **523:** 208-211.
753
754 Cabello-Yeves PJ, Zemskaya TI, Rosselli R, Coutinho FH, Zakharenko AS, Blinov VV *et al.*
755 (2018). Genomes of novel microbial lineages assembled from the sub-ice waters of Lake
756 Baikal. *Appl Environ Microbiol* **84:** e02132-02117.
757
758 Carini P, Steindler L, Beszteri S, Giovannoni SJ (2013). Nutrient requirements for growth of
759 the extreme oligotroph 'Candidatus Pelagibacter ubique' HTCC1062 on a defined medium.
760 *ISME J* **7:** 592-602.
761
762 Carini P, Campbell EO, Morré J, Sañudo-Wilhelmy SA, Thrash CJ, Bennett SE *et al.* (2014).
763 Discovery of a SAR11 growth requirement for thiamin's pyrimidine precursor and its
764 distribution in the Sargasso Sea. *ISME J* **8:** 1727-1738.
765
766 Carlson CA, Morris R, Parsons R, Treusch AH, Giovannoni SJ, Vergin KL (2009). Seasonal
767 dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern
768 Sargasso Sea. *ISME J* **3:** 283-295.
769
770 Castresana J (2000). Selection of conserved blocks from multiple alignments for their use in
771 phylogenetic analysis. *Mol Biol Evol* **17:** 540-552.
772

773 Chun J, Oren A, Ventosa A, Christensen H, Arahal D, da Costa MS *et al.* (2018). Proposed
774 minimal standards for the use of genome data for the taxonomy of prokaryotes. *IJSEM* **68:**
775 461-466.
776
777 Contreras-Moreira B, Vinuesa P (2013). GET_HOMOLOGUES, a versatile software package
778 for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol* **79:** 7696-
779 7701.
780
781 Dattananda CS, Gowrishankar J (1989). Osmoregulation in Escherichia coli:
782 complementation analysis and gene-protein relationships in the proU locus. *J Bacteriol*
783 **171:** 1915-1922.
784
785 Denef VJ, Mueller RS, Chiang E, Liebig JR, Vanderploeg HA (2016). Chloroflexi CL500-11
786 Populations That Predominate Deep-Lake Hypolimnion Bacterioplankton Rely on
787 Nitrogen-Rich Dissolved Organic Matter Metabolism and C1 Compound Oxidation. *Appl*
788 *Environ Microbiol* **82:** 1423-1432.
789
790 Denger K, Weinitschke S, Smits THM, Schleheck D, Cook AM (2008). Bacterial sulfite
791 dehydrogenases in organotrophic metabolism: separation and identification in Cupriavidus
792 necator H16 and in Delftia acidovorans SPH-1. *Microbiology* **154:** 256-263.
793
794 Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP *et al.* (2009).
795 Community-wide analysis of microbial genome sequence signatures. *Genome Biol* **10:** R85.
796
797 Dupont CL, Larsson J, Yooseph S, Ininbergs K, Goll J, Asplund-Samuelsson J *et al.* (2014).
798 Functional Tradeoffs Underpin Salinity-Driven Divergence in Microbial Community
799 Composition. *PLOS ONE* **9:** e89549.
800
801 Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high
802 throughput. *Nucleic Acids Research* **32:** 1792-1797.
803
804 Eiler A, Zaremba-Niedźwiedzka K, Martinez-Garcia M, McMahon KD, Stepanauskas R,
805 Andersson SGE *et al.* (2014). Productivity and salinity structuring of the microplankton
806 revealed by comparative freshwater metagenomics. *Environmental Microbiology* **16:** 2682-
807 2698.
808
809 Eiler A, Mondav R, Sinclair L, Fernandez-Vidal L, Scofield DG, Schwientek P *et al.* (2016).
810 Tuning fresh: radiation through rewiring of central metabolism in streamlined bacteria.
811 *ISME J* **10:** 1902-1914.
812
813 Empadinhas N, da Costa M (2008). Osmoadaptation mechanisms in prokaryotes:
814 distribution of compatible solutes. *International Microbiology*.
815
816 Fortunato CS, Crump BC (2015). Microbial Gene Abundance and Expression Patterns across
817 a River to Ocean Salinity Gradient. *PLOS ONE* **10**.
818

819 Fu L, Niu B, Zhu Z, Wu S, Li W (2012). CD-HIT: accelerated for clustering the next-
820 generation sequencing data. *Bioinformatics* **28:** 3150-3152.
821
822 Garcia SL, Stevens SLR, Crary B, Martinez-Garcia M, Stepanauskas R, Woyke T *et al.* (2017).
823 Contrasting patterns of genome-level diversity across distinct co-occurring bacterial
824 populations. *ISME J* **AOP doi:10.1038/s41396-017-0001-0**.
825
826 Ghai R, Hernandez C, Picazo A, Mizuno C, Ininbergs K, Díez B *et al.* (2012). Metagenomes of
827 Mediterranean Coastal Lagoons. *Sci Rep* **2**.
828
829 Giovannoni SJ, Bibbs L, Cho J-C, Stapels MD, Desiderio R, Vergin KL *et al.* (2005a).
830 Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438:** 82-85.
831
832 Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D *et al.* (2005b). Genome
833 streamlining in a cosmopolitan oceanic bacterium. *Science* **309:** 1242-1245.
834
835 Giovannoni SJ, Vergin KL (2012). Seasonality in ocean microbial communities. *Science* **335:**
836 671-676.
837
838 Glenn TC, Nilsen R, Kieran TJ, Finger JW, Pierson TW, Bentley KE *et al.* (2016). Adapterama
839 I: Universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru &
840 iNext). *bioRxiv*.
841
842 Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM (2007).
843 DNA–DNA hybridization values and their relationship to whole-genome sequence
844 similarities. *IJSEM* **57:** 81-91.
845
846 Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ *et al.* (2012). Streamlining
847 and Core Genome Conservation among Highly Divergent Members of the SAR11 Clade.
848 *mBio* **3:** e00252-00212.
849
850 Han MV, Zmasek CM (2009). phyloXML: XML for evolutionary biology and comparative
851 genomics. *BMC Bioinformatics* **10:** 1-6.
852
853 Henson MW, Pitre DM, Weckhorst J, Lanclos VC, Webber AT, Thrash JC (2016). Artificial
854 Seawater Media Facilitate Cultivating Members of the Microbial Majority from the Gulf of
855 Mexico. *mSphere* **1:** e00028-00016. doi:00010.01128/mSphere.00028-00016.
856
857 Herlemann D, Woelk J, Labrenz M, Jürgens K (2014). Diversity and abundance of
858 "Pelagibacterales" (SAR11) in the Baltic Sea salinity gradient. *Syst Appl Microbiol* **37:** 601-
859 604.
860
861 Hunt M, Kikuchi T, Sanders M, Newbold C, Berriman M, Otto TD (2013). REAPR: a universal
862 tool for genome assembly evaluation. *Genome Biol* **14:** 1-10.
863

Jimenez-Infante F, Ngugi D, Vinu M, Blom J, Alam I, Bajic VB *et al.* (2017). Genomic Characterization of Two Novel SAR11 Isolates From the Red Sea, Including the First Strain of the SAR11 Ib clade. *FEMS Microbiol Ecol*.

Junier T, Zdobnov EM (2010). The Newick utilities: high-throughput phylogenetic tree processing in the Unix shell. *Bioinformatics* **26:** 1669-1670.

Kappler U (2011). Bacterial sulfite-oxidizing enzymes. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1807:** 1-10.

Konstantinidis KT, Tiedje JM (2005). Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* **102:** 2567-2572.

Konstantinidis KT, Tiedje JM (2007). Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Current Opinion in Microbiology* **10:** 504-509.

Li H, Durbin R (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25:** 1754-1760.

Logares R, Bråte J, Bertilsson S, Clasen JL, Shalchian-Tabrizi K, Rengefors K (2009). Infrequent marine–freshwater transitions in the microbial world. *Trends in Microbiology* **17:** 414-422.

Logares R, Brate J, Heinrich F, Shalchian-Tabrizi K, Bertilsson S (2010). Infrequent Transitions between Saline and Fresh Waters in One of the Most Abundant Microbial Lineages (SAR11). *Mol Biol Evol* **27:** 347-357.

Luo H, Csűros M, Hughes AL, Moran M (2013). Evolution of Divergent Life History Strategies in Marine Alphaproteobacteria. *mBio* **4:** e00373-00313.

Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Pillay M *et al.* (2014). IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Research* **42:** D560-567.

Martinez-Garcia M, Swan BK, Poulton NJ, Gomez M, Masland D, Sieracki ME *et al.* (2011). High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. *ISME J* **6:** 113-123.

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

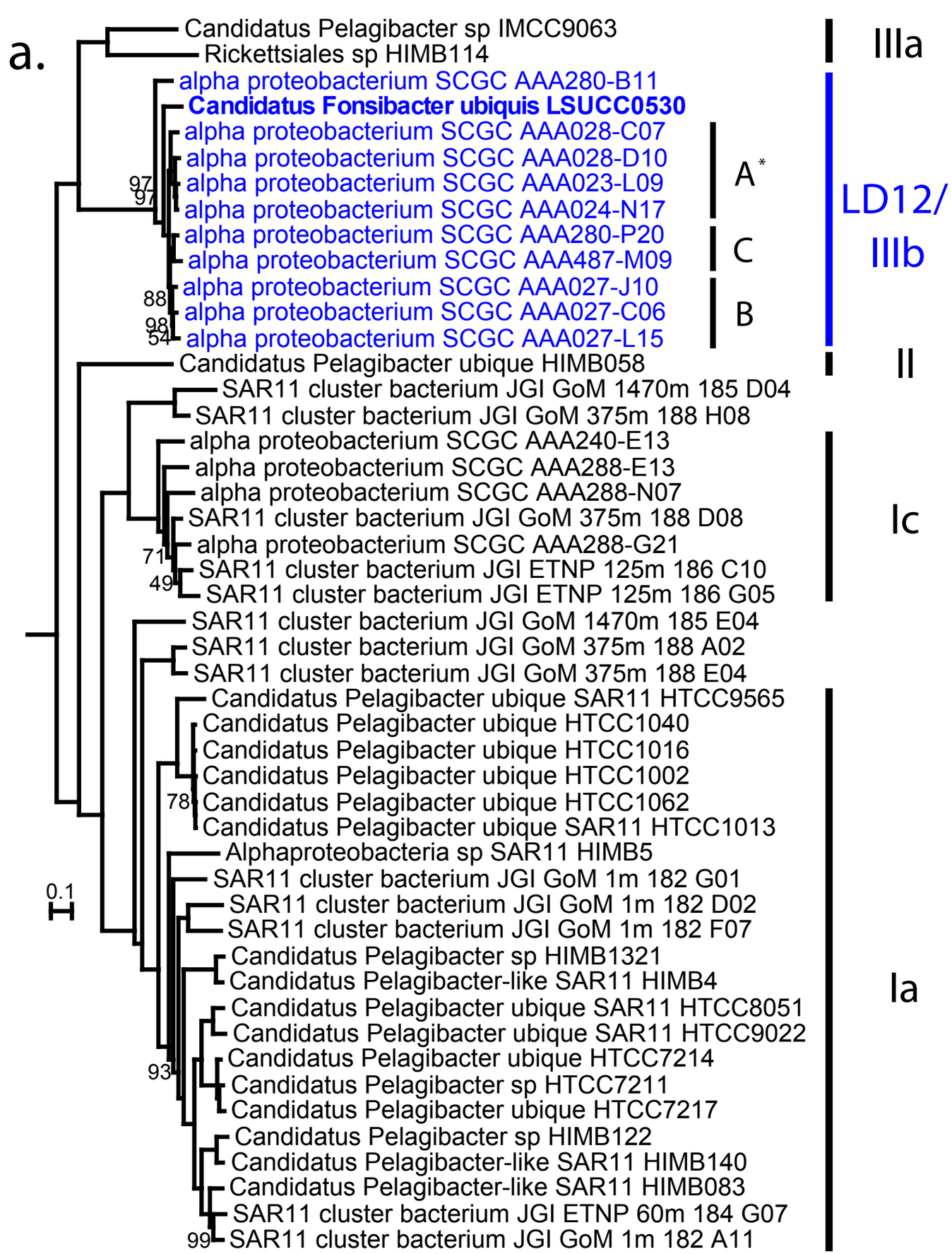
Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA *et al.* (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420:** 806-810.

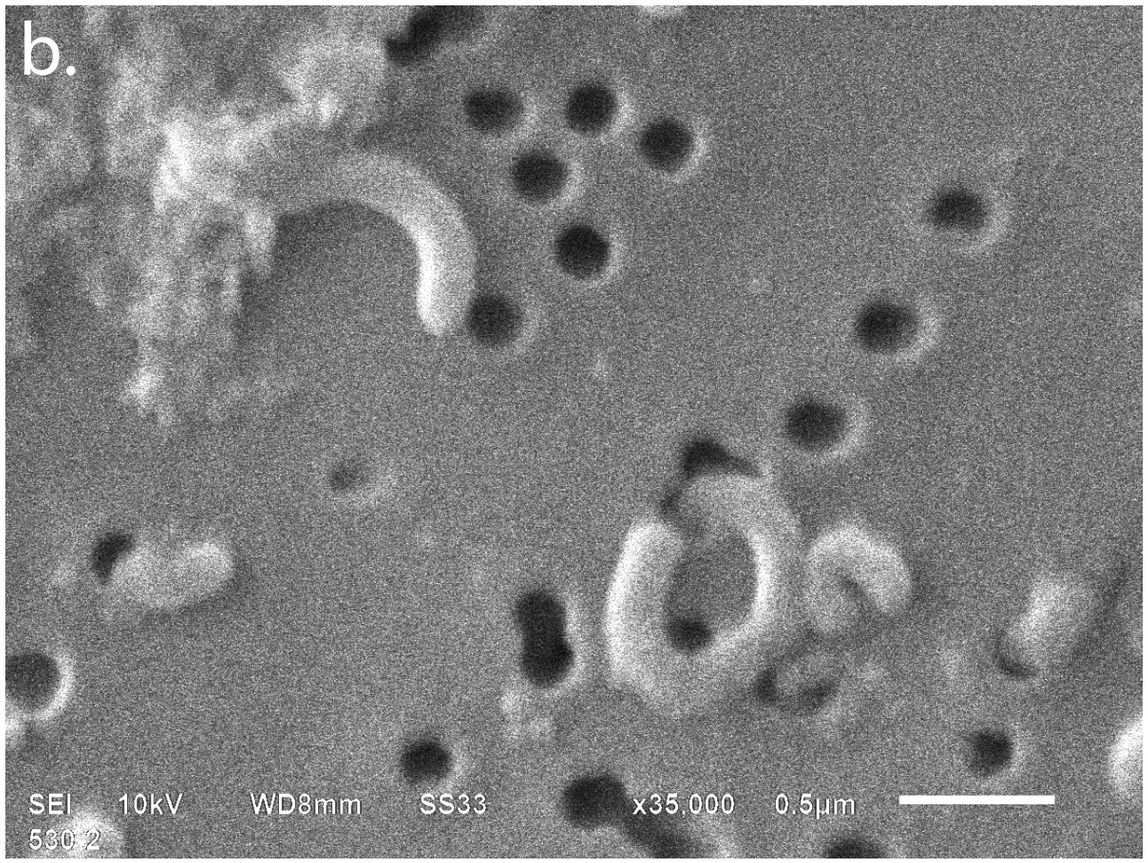
Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S (2011). A Guide to the Natural History of Freshwater Lake Bacteria. *Microbiol Mol Biol Rev* **75:** 14-49.

910
911 Oh S, Caro-Quintero A, Tsementzi D, DeLeon-Rodriguez N, Luo C, Poretsky R *et al.* (2011).
912 Metagenomic Insights into the Evolution, Function, and Complexity of the Planktonic
913 Microbial Community of Lake Lanier, a Temperate Freshwater Ecosystem. *Appl Environ*
914 *Microbiol* **77:** 6000-6011.
915
916 Parada AE, Needham DM, Fuhrman JA (2015). Every base matters: assessing small subunit
917 rRNA primers for marine microbiomes with mock communities, time series and global field
918 samples. *Environmental Microbiology* **18**.
919
920 Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015). CheckM: assessing
921 the quality of microbial genomes recovered from isolates, single cells, and metagenomes.
922 *Genome Research* **25:** 1043-1055.
923
924 Piwosz K, Salcher MM, Zeder M, Ameryk A, Pernthaler J (2013). Seasonal dynamics and
925 activity of typical freshwater bacteria in brackish waters of the Gulf of Gdańsk. *Limnology*
926 *and Oceanography* **58:** 817-826.
927
928 Price MN, Dehal PS, Arkin AP (2010). FastTree 2--approximately maximum-likelihood trees
929 for large alignments. *PLOS ONE* **5:** e9490.
930
931 Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J *et al.* (2007). SILVA: a
932 comprehensive online resource for quality checked and aligned ribosomal RNA sequence
933 data compatible with ARB. *Nucleic Acids Research* **35:** 7188-7196.
934
935 Rappé MS, Connon SA, Vergin KL, Giovannoni SJ (2002). Cultivation of the ubiquitous
936 SAR11 marine bacterioplankton clade. *Nature* **418:** 630-633.
937
938 Rodriguez-Valera F, Martin-Cuadrado A-B, Rodriguez-Brito B, Pašić L, Thingstad FT,
939 Rohwer F *et al.* (2009). Explaining microbial population genomics through phage
940 predation. *Nat Rev Micro* **7:** 828-836.
941
942 Roland N, Reich D (2012). Cost-effective, high-throughput DNA sequencing libraries for
943 multiplexed target capture. *Genome Research* **22:** 939-946.
944
945 Roux S, Enault F, Hurwitz BL, Sullivan MB (2015). VirSorter: mining viral signal from
946 microbial genomic data. *PeerJ* **3**.
947
948 Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S *et al.* (2007). The
949 Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical
950 Pacific. *Plos Biol* **5:** e77.
951
952 Salcher MM, Pernthaler J, Posch T (2011). Seasonal bloom dynamics and ecophysiology of
953 the freshwater sister clade of SAR11 bacteria 'that rule the waves' (LD12). *ISME J* **5:** 1242-
954 1252.
955

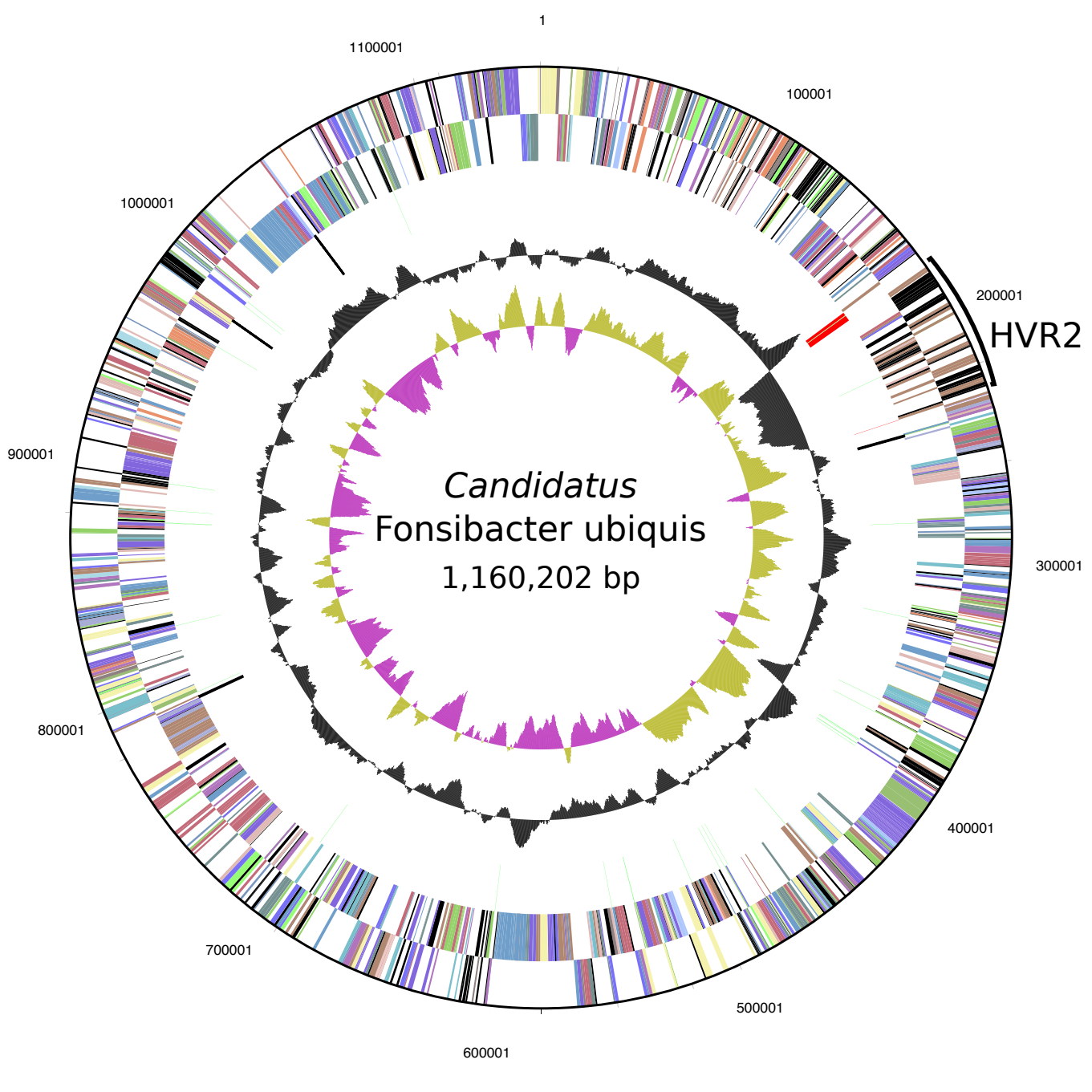
956 Schattenhofer M, Fuchs BM, Amann R, Zubkov MV, Tarran GA, Pernthaler J (2009).
957 Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean.
958 *Environmental Microbiology* **11:** 2078-2093.
959
960 Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.* (2009).
961 Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software
962 for Describing and Comparing Microbial Communities. *Appl Environ Microbiol* **75:** 7537-
963 7541.
964
965 Schwalbach MS, Tripp HJ, Steindler L, Smith DP, Giovannoni SJ (2010). The presence of the
966 glycolysis operon in SAR11 genomes is positively correlated with ocean productivity.
967 *Environmental Microbiology* **12:** 490-500.
968
969 Smith DP, Thrash JC, Nicora CD, Lipton MS, Burnum-Johnson KE, Carini P *et al.* (2013a).
970 Proteomic and Transcriptomic Analyses of "Candidatus Pelagibacter ubique" Describe the
971 First PII-Independent Response to Nitrogen Limitation in a Free-Living
972 Alphaproteobacterium. *mBio* **4:** 12.
973
974 Smith DP, Nicora CD, Carini P, Lipton MS, Norbeck AD, Smith RD *et al.* (2016). Proteome
975 Remodeling in Response to Sulfur Limitation in "Candidatus Pelagibacter ubique".
976 *mSystems* **1:** 16.
977
978 Smith MW, Allen L, Allen AE, Herfort L, Simon HM (2013b). Contrasting genomic properties
979 of free-living and particle-attached microbial assemblages within a coastal ecosystem.
980 *Front Microbiol* **4**.
981
982 Song J, Oh H-M, Cho J-C (2009). Improved culturability of SAR11 strains in dilution-to-
983 extinction culturing from the East Sea, West Pacific Ocean. *FEMS Microbiology Letters* **295:**
984 141-147.
985
986 Stamatakis A, Hoover P, Rougemont J (2008). A Rapid Bootstrap Algorithm for the RAxML
987 Web Servers. *Systematic Biology* **57:** 758-771.
988
989 Steindler L, Schwalbach MS, Smith DP, Chan F, Giovannoni SJ (2011). Energy Starved
990 Candidatus Pelagibacter Ubique Substitutes Light-Mediated ATP Production for
991 Endogenous Carbon Respiration. *PLOS ONE* **6:** e19725.
992
993 Stingl U, Tripp HJ, Giovannoni SJ (2007). Improvements of high-throughput culturing
994 yielded novel SAR11 strains and other abundant marine bacteria from the Oregon coast
995 and the Bermuda Atlantic Time Series study site. *ISME J* **1:** 361-371.
996
997 Sun J, Steindler L, Thrash JC, Halsey KH, Smith DP, Carter AE *et al.* (2011). One Carbon
998 Metabolism in SAR11 Pelagic Marine Bacteria. *PLOS ONE* **6:** e23973.
999

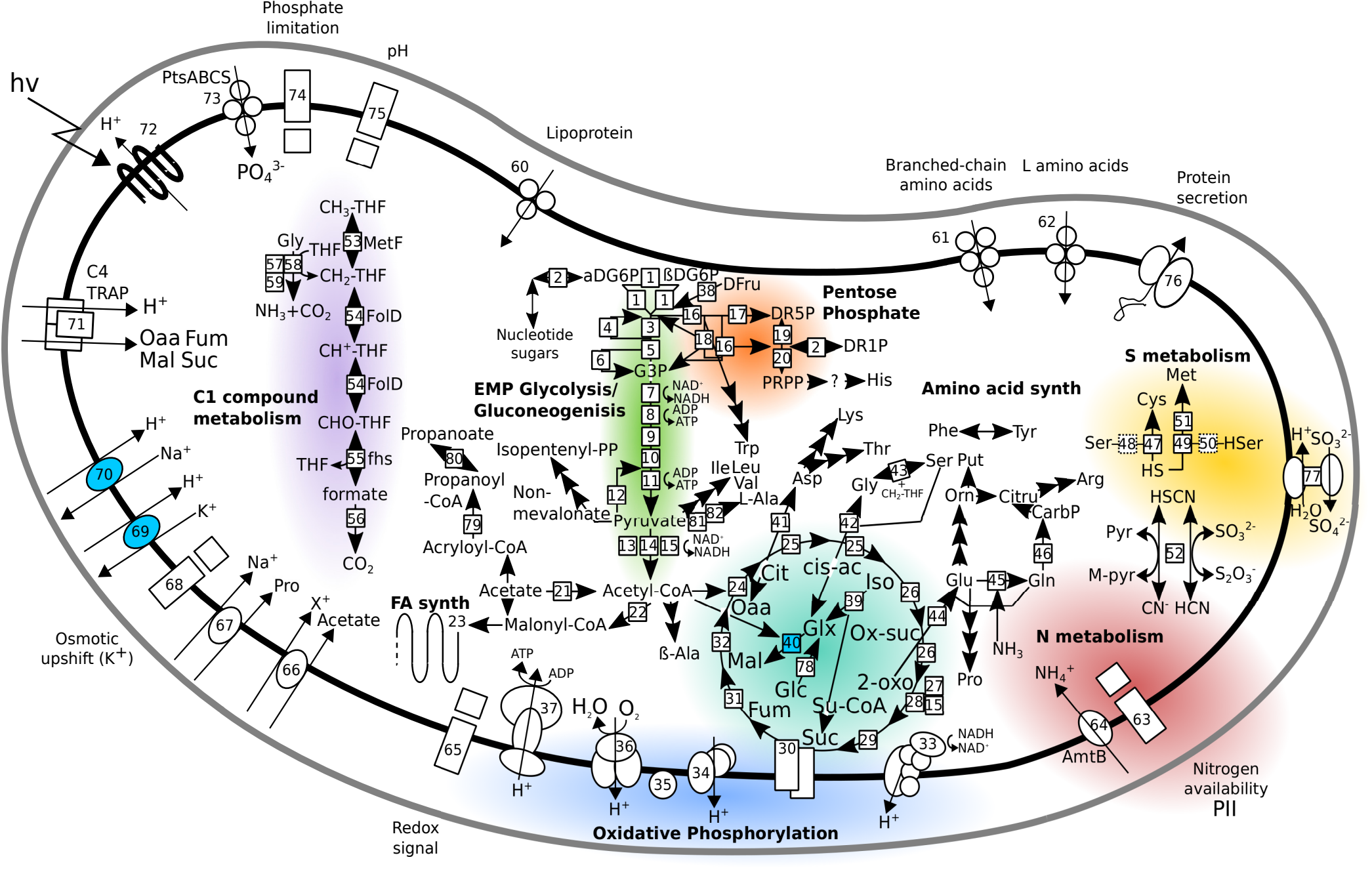
1000 Sun J, Todd JD, Thrash CJ, Qian Y, Qian MC, Temperton B *et al.* (2016). The abundant marine
1001 bacterium Pelagibacter simultaneously catabolizes dimethylsulfoniopropionate to the
1002 gases dimethylsulfide and methanethiol. *Nature Microbiology* **1**.
1003
1004 Thrash J, Seitz KW, Baker BJ, Temperton B, Gillies LE, Rabalais NN *et al.* (2017). Metabolic
1005 Roles of Uncultivated Bacterioplankton Lineages in the Northern Gulf of Mexico "Dead
1006 Zone". *mBio* **8**.
1007
1008 Thrash JC, Temperton B, Swan BK, Landry ZC, Woyke T, DeLong EF *et al.* (2014). Single-cell
1009 enabled comparative genomics of a deep ocean SAR11 bathytype. *ISME J* **8:** 1440-1451.
1010
1011 Tripp HJ, Kitner JB, Schwalbach MS, Dacey JWH, Wilhelm LJ, Giovannoni SJ (2008). SAR11
1012 marine bacteria require exogenous reduced sulphur for growth. *Nature* **452:** 741-744.
1013
1014 Tsementzi D, Wu J, Deutsch S, Nath S, Rodriguez-R LM, Burns AS *et al.* (2016). SAR11
1015 bacteria linked to ocean anoxia and nitrogen loss. *Nature* **536:** 179-183.
1016
1017 Tseng C-H, Chiang P-W, Shiah F-K, Chen Y-L, Liou J-R, Hsu T-C *et al.* (2013). Microbial and
1018 viral metagenomes of a subtropical freshwater reservoir subject to climatic disturbances.
1019 *ISME J* **7:** 2374-2386.
1020
1021 Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S *et al.* (2014). Pilon: An
1022 Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly
1023 Improvement. *PLOS ONE* **9**.
1024
1025 Wickham H (2011). ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics* **3:**
1026 180-185.
1027
1028 Wilhelm LJ, Tripp JH, Givan SA, Smith DP, Giovannoni SJ (2007). Natural variation in SAR11
1029 marine bacterioplankton genomes inferred from metagenomic data. *Biology Direct* **2:** 1-19.
1030
1031 Zaremba-Niedzwiedzka K, Viklund J, Zhao W, Ast J, Sczyrba A, Woyke T *et al.* (2013). Single-
1032 cell genomics reveal low recombination frequencies in freshwater bacteria of the SAR11
1033 clade. *Genome Biol* **14:** 1-14.
1034
1035 Zhao X, Schwartz CL, Pierson J, Giovannoni SJ, McIntosh RJ, Nicastro D (2017). Three-
1036 Dimensional Structure of the Ultraoligotrophic Marine Bacterium "Candidatus Pelagibacter
1037 ubique". *Appl Environ Microbiol* **83:** 16.
1038
1039 Zhao Y, Ben T, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC *et al.* (2013). Abundant
1040 SAR11 viruses in the ocean. *Nature* **494:** 357-360.
1041
1042


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
IIIa
LD12/
IIIb
A*
C
B
II
Ic
Ia

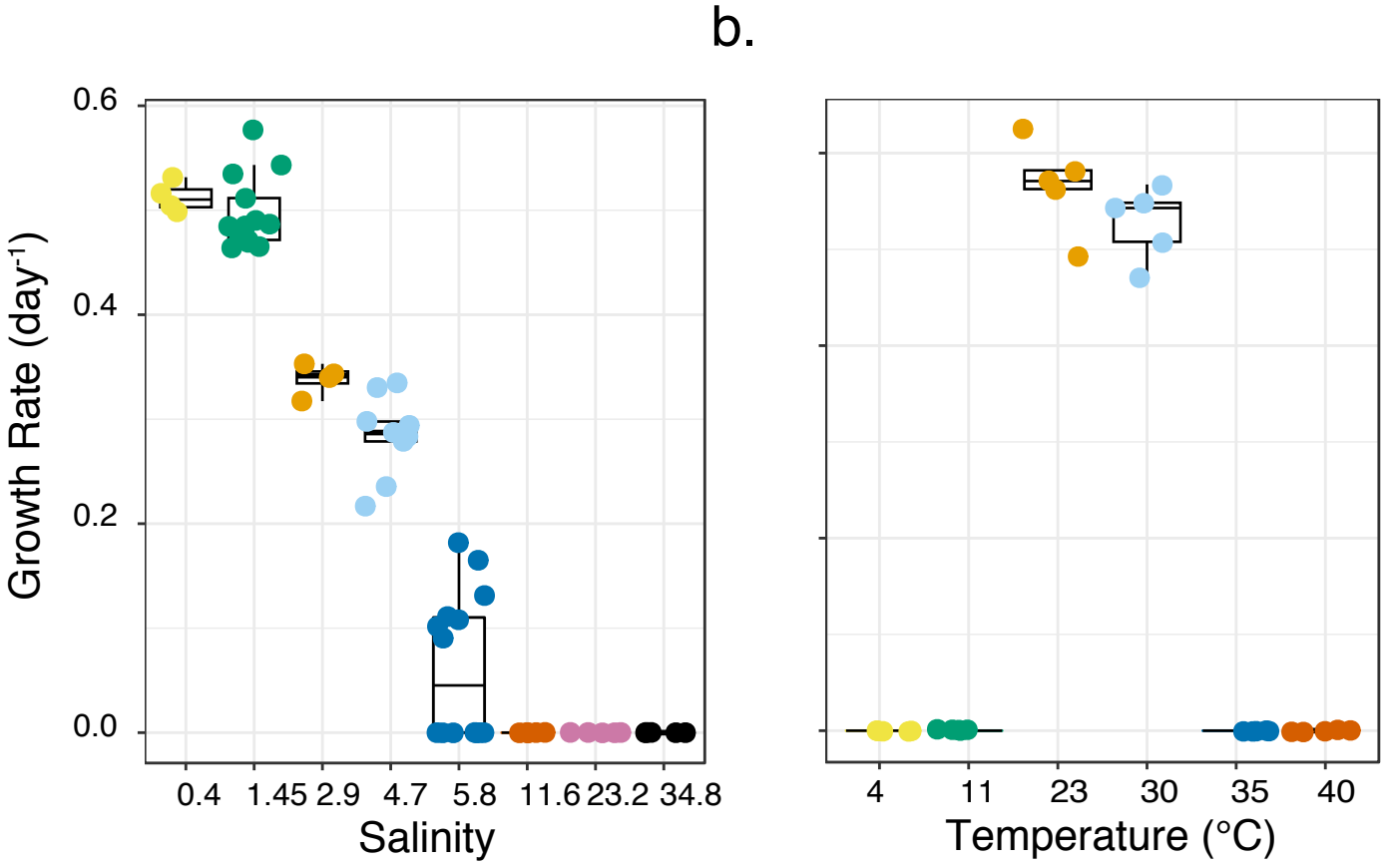

b.
SEI 10kV WD8mm SS33 x35,000 0.5µm
530 2

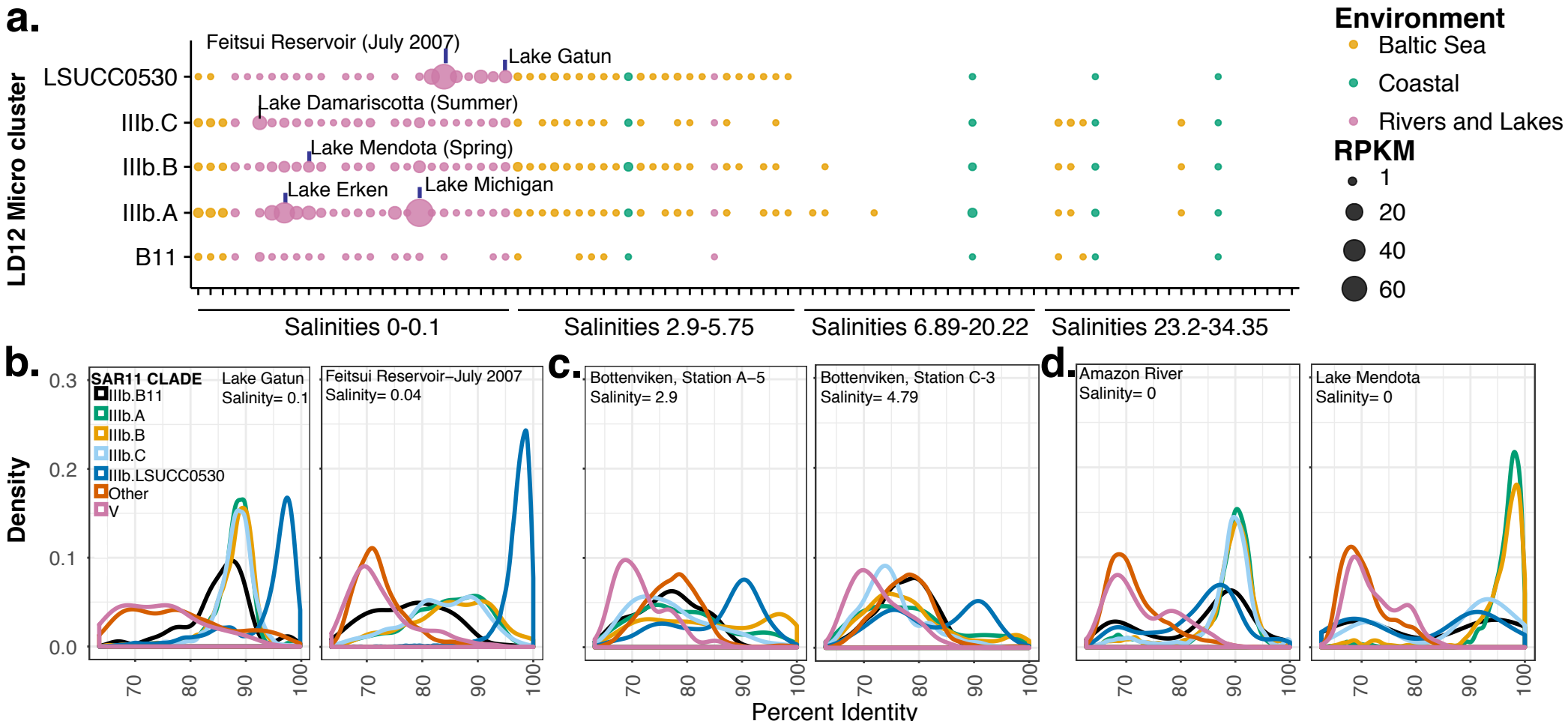
c.
500 nm


1
100001
200001
300001
400001
500001
600001
700001
800001
900001
1000001
1100001
HVR2
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
CH3-THF
Gly
THF
53 MetF
57 58
59
CH2-THF
NH3+CO2
54 FolD
CH+-THF
54 FolD
CHO-THF
THF
55 fhs
formate
56
CO2
C4 TRAP
71
H+
Oaa Fum Mal Suc
C1 compound metabolism
H+
Na+
70
H+
K+
69
68
Na+
Pro
X+
Acetate
67
66
Osmotic upshift (K+)
2 aDG6P
1 ßDG6P
38 DFru
Nucleotide sugars
4
3
5
6
G3P
16
17 DR5P
Pentose Phosphate
18
16
19
20
2 DR1P
PRPP
? His
EMP Glycolysis/ Gluconeogenisis
7 NAD+ NADH
8 ADP ATP
9
10
11 ADP ATP
12
Pyruvate
13 14 15 NAD+ NADH
Propanoate
80
Propanoyl -CoA
79
Acryloyl-CoA
Isopentenyl-PP
Non-mevalonate
Trp
Ile Leu Val
L-Ala
82
81
Lys
Thr
Asp
Phe
Tyr
Ser Put
43
Gly
CH2-THF
Orn
Citru
Arg
CarbP
46
Glu
45
Gln
NH3
Pro
Amino acid synth
S metabolism
Met
Cys
51
Ser 48 47 49 50 HSer
HS
HSCN
Pyr
52
M-pyr
SO3 2-
S2O3-
CN- HCN
H+ SO3 2-
77
H2O
SO4 2-
Acetate 21 Acetyl-CoA
22
Malonyl-CoA
23
FA synth
ß-Ala
24
Cit
25
cis-ac
25
Iso
26
Oaa
39
Glx
Ox-suc
44
26
32
Mal
40
78
Glc
2-oxo
27
28 15
31
Fum
Su-CoA
Suc
29
30
33 NADH NAD+
H+
N metabolism
NH4+
64
63
AmtB
Nitrogen availability
PII
ATP
ADP
37
H2O
O2
36
35
34
H+
H+
H+
H+
65
Redox signal
Oxidative Phosphorylation


a.
b.
Growth Rate (day-1)
0.6
0.4
0.2
0.0
0.4
1.45
2.9
4.7
5.8
11.6
23.2
34.8
Salinity
4
11
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