Streamlined freshwater bacterioplankton *Nanopelagicales* (acI) and "*Ca*. Fonsibacter" (LD12) thrive in functional cohorts

Running title: Cooperation models for freshwater streamlined bacteria

Rhiannon Mondav[a,#], Stefan Bertilsson[a,b,c], Moritz Buck[a,c,d], Silke Langenheder[a], Eva S. Lindström[a], Sarahi L Garcia[a,e]

[a] Department of Ecology and Genetics, Limnology, Uppsala University, Sweden

[b] Science for Life Laboratory, Uppsala University, Sweden

[c] Swedish University of Agricultural Sciences, Uppsala, Sweden

[d] National Bioinformatics Infrastructure Sweden, Sweden

[e] Department of Ecology, Environment, and Plant Sciences, Science for Life Laboratory, Stockholm University, Stockholm, Sweden

Corresponding author: Rhiannon Mondav, rhiannon.mondav@gmail.com

ABSTRACT

While fastidious microbes can be abundant and ubiquitous in their natural communities, many fail to grow axenically in laboratories due to auxotrophies or other dependencies. To overcome auxotrophies these microbes rely on their surrounding cohort. A cohort may consist of kin (ecotypes) or more distantly related organisms (community) with the cooperation being reciprocal or non-reciprocal, and expensive (Black Queen hypothesis) or costless (byproduct). These metabolic partnerships (whether at single species population or community level) enable dominance by and coexistence of these lineages in nature. Here we examine the relevance of these cooperation models to explain the abundance and ubiquity of the dominant fastidious bacterioplankton of a dimictic mesotrophic freshwater lake. Using both culture dependent (minimalist mixed cultures) and culture independent (SSU rRNA gene time series and environmental metagenomics) methods we independently identified the primary cohorts of *Actinobacterial* genera "*Ca*. Planktophila*"* (acI-A) and "*Ca.* Nanopelagicus" (acI-B), and the *Proteobacterial* genus "*Ca.* Fonsibacter" (LD12). While "*Ca*. Planktophila*"* and "*Ca.* Fonsibacter" had no correlation in their natural habitat, they have the potential to be complementary in laboratory settings. We also investigated the bi-functional catalase-peroxidase enzyme KatG (a common good which "*Ca*. Planktophila*"* is dependent upon) and its most likely providers in the lake. Further we found that while ecotype and community cooperation combined may explain "*Ca*. Planktophila*"* population abundance, the success of "*Ca.* Nanopelagicus" and "*Ca.* Fonsibacter" is better explained as a community byproduct. Ecotype differentiation of "*Ca.* Fonsibacter" as a means of escaping predation was supported but not for overcoming auxotrophies.

40

41 **IMPORTANCE**

42 This study examines evolutionary and ecological relationships of three of the most
43 ubiquitous and abundant freshwater bacterial genera: "*Ca*. Planktophila*"* (acI-A), "*Ca.*
44 Nanopelagicus" (acI-B), and "*Ca.* Fonsibacter" (LD12). Due to high abundance, these
45 genera might have a significant influence on nutrient cycling in freshwaters worldwide and
46 this study adds a layer of understanding to how seemingly competing clades of bacteria
47 can co-exist by having different cooperation strategies. Our synthesis ties together network
48 and ecological theory with empirical evidence and lays out a framework for how the
49 functioning of populations within complex microbial communities can be studied.

50

51 **INTRODUCTION**

52 The stable coexistence of species is described by a continuum of inter-specific symbiotic
53 interactions from parasitism through neutralism to mutualism (1). The extended
54 coexistence of microbial strains is described by intra-specific relationship models limited
55 to niche theory (ecotype differentiation) (2) and predator control of dominant strains (3).
56 Combining these ecological theories of co-existence with phylogenetic information can
57 produce a simplified model with a new holistic perspective for understanding aquatic
58 microbial communities. In aquatic microbial communities, especially for non-particle-
59 associated bacterioplankton, physical structures and proximity as seen in soil, host-

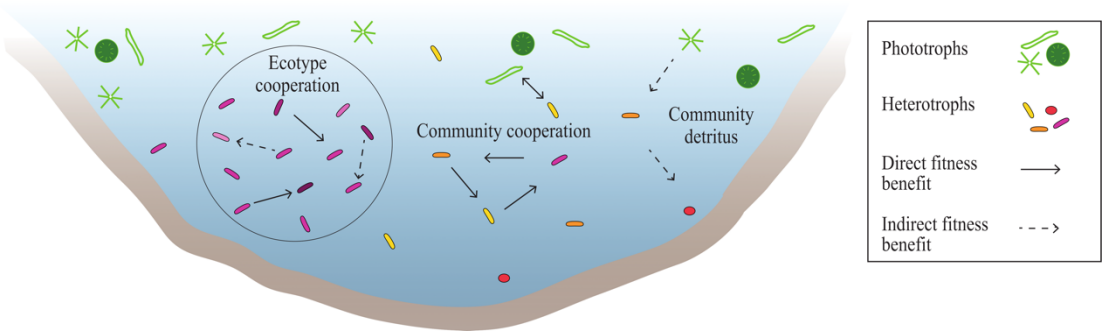
60 associated microbiome, and biofilm communities (4) are absent. As cooperation models
61 have been developed from the study of such tightly interwoven assemblages their models
62 may not be the most suitable. Absent also is the habitat isolation which limits dispersal as
63 seen in e.g. soil ecosystems. In aquatic systems interactions between nutrients and biota
64 can occur from micro up to macro-scale distances involving stochastic effectors from
65 Brownian motion up to whole system circulation currents. Direct but distant interactions
66 are therefore difficult to identify based purely on proximity as common goods and toxins
67 in the environment can come from multiple sources and be transported long distances. Data
68 from long-term monitoring of aquatic environments can be used to test the validity of
69 ecological models.
70 Lake Erken has been extensively studied for several decades, and ample background
71 information on nutrient cycling, planktonic communities and biogeochemical processes
72 combined with well-established infrastructure makes this ecological observatory an
73 excellent choice for the study of competing microbial assemblage processes (5–7). Several
74 studies have investigated the microbial component of this lake ecosystem (8–13), revealing
75 that the most abundant microbes in Lake Erken belonged to the LD12 (now described as
76 the candidate genus "*Ca.* Fonsibacter" (14)) and acI (ACK-M1, hgcI, but now described as
77 the candidate order *Nanopelagicales*) clades (11, 12, 15–20). "*Ca.* Fonsibacter" is a non-
78 marine genus within the ubiquitous and abundant *Pelagibacteraceae* family which includes
79 the marine *Pelagibacter* genus (SAR11). *Nanopelagicales* is a freshwater order with two

80 described genera, *"Ca.* Planktophila*"* and "*Ca*. Nanopelagicus". Both "*Ca.* Fonsibacter"
81 and the *Nanopelagicales* have small cell size, reduced genomes (1.16 to 1.48 Mbp),
82 multiple auxotrophies, a requirement for reduced sulfur compounds, have rhodopsins, and
83 have been recalcitrant to maintenance in axenic laboratory culture (14, 21–23). "*Ca.*
84 Fonsibacter" and "*Ca*. Nanopelagicus" genera both have additional amino-acid
85 auxotrophies (23, 24). Recently a single strain of "*Ca.* Fonsibacter" and two strains of the
86 *Nanopelagicales* genus *"Ca.* Planktophila*"* (acI-A and acI-A4) were successfully
87 maintained in axenic cultures by customized liquid media in the case of "*Ca.* Fonsibacter"
88 (14) and the addition of an active enzyme to filtered lake-water media for *"Ca.*
89 Planktophila*"* (25). While both have been cultured neither have been deposited in culture
90 collections and so retain *Candidatus* status. The difference in culture methods point
91 towards a difference in lifestyle strategies and therefore evolution and ecology in these two
92 abundant and ubiquitous freshwater bacterioplankton clades.
93 The ubiquity of the small-genome bacterial lineages, such as *Nanopelagicales* and "*Ca*.
94 Fonsibacter" in freshwaters, is thought to be a consequence of their streamlined genomes
95 and small cell size (26). On one hand, small cell size increases surface area to volume
96 ratio, and improves diffusion, therefore decreasing time for foraging for non-motile
97 organisms (4). On the other hand, reduced genomes have lost extraneous metabolic
98 functions while having very low gene redundancy, reduced regulatory components, and
99 high coding density (27). This combined streamlining and shrinkage reduces the energetic

100 and material cost of genome, proteome, and cellular maintenance. Because their small
101 genomes encode fewer metabolic functions than average, it has been proposed they require
102 a large effective population (large N_e) within which ecotypes diverge to provide the
103 genomic flexibility usually seen in individual cells of bacteria with larger genomes. A large
104 N_e is also required for purifying selection to be efficient at reducing and maintaining small
105 genome size. Such streamlined microbes were first noted in low nutrient aquatic
106 environments and streamlining has since then been suggested as being of selective
107 advantage under reduced nitrogen availability (28). *Nanopelagicales* and "*Ca*.
108 Fonsibacter" clades are dominant and successful in non-marine environments such as Lake
109 Erken, which far from being nutrient limited is classified as mesotrophic on a
110 eutrophication trajectory. Further, as Lake Erken is ice and snow covered most winters
111 (temperature and light limiting) and stratified in summer (thermo and oxy-clines), it is not
112 a stable environment where organisms with reduced regulatory capacity are expected to
113 thrive. It is unlikely that dominance in such productive lakes is made possible purely by
114 streamlining, as the subsequent loss of genomic flexibility required to respond to a dynamic
115 and variable habitat such as Lake Erken would counteract the benefits. Furthermore, large
116 populations attract predators though both clades may escape some grazing predation due
117 to their small size (29, 30). "*Ca*. Fonsibacter" may additionally have similar membrane
118 properties to its sister marine lineage SAR11(31) to further elude grazing. All three genera
119 encode cell membrane modification genes in hypervariable regions predicted to assist

120 escape from viral predation (23, 32). A large N_e where the divergence is focused in
121 immunogenic or predator-escape mechanisms (defense specialist) will help a species
122 survive population sweeps caused by predators, though it does not explain the intra nor
123 inter-specific diversity or the seeming lack of competition between these three genera.
124 We combined the continuum of symbiotic interactions with phylogenetic information to
125 produce a simplified model with a new holistic perspective for understanding aquatic
126 microbial communities. (Fig. 1). Ecotype cooperation model describes how anabolic
127 variability within a species population i.e. strain level variation, enables all the strains
128 working together to produce nutrients required for cellular function and replication. It also
129 allows for recombination and swapping of metabolic modules across strains. The
130 community cooperation model also has a basis in anabolic complementarity but places the
131 metabolic variation within a symbiotic group of different species whereby altruistic
132 members provide expensive (Black Queen Hypothesis) or costless common goods (33, 34).
133 The community detrital model is similar to community cooperation but focuses on uptake
134 and catabolism of nutrients released by cellular death mediated by toxicity, starvation, or
135 predation. Aquatic habitats in particular, due to transport and diffusion, are understood to
136 have a unique niche for detritivores that may select for streamlined microbes (35).
137 Here we examine an 8-year time-series along with 26 minimalist mixed-cultures and
138 selected lake metagenomes, to identify putative supporting microbes (cohort) and whether
139 that support is best described by ecotype, community, and/or detrital models. We began

140 with OTU network modelling (to define a candidate cohort that might provide for the target
141 three genera dependencies), examined the phylogenetic relatedness of each cohort (to
142 determine if a signal for ecotype or community existed), reviewed the metabolic diversity
143 of the target group (to establish its dependencies and diversity), and investigated potential
144 providers of the common-goods catalase-peroxidase (KatG) (a known dependency of some
145 *Nanopelagicales*) within the metagenomes. Our proposed framework focusses on neutral
146 to positive interactions as the laboratory experimental design does not permit detection of
147 negative interactions even though timeseries can expose potential negative interactions.
148



149

150 Fig. 1. Cooperation models shown in a lake setting. Ecotype Cooperation Model describes
151 anabolic variability as the main contributor to co-existence while allowing for
152 recombination and swapping of metabolic modules across strains. Community Cooperation
153 Model also encompasses anabolic complementarity but between rather than within species
154 and includes common goods ranging from expensive to costless. The Community Detrital

155 Model places complementarity both between and within species but describes catabolic
156 variation.
157
158 **RESULTS**
159 **Taxon abundance and environmental correlates**
160 The most abundant clades detected in the SSU rRNA gene 8-year time-series were the
161 *Actinobacterial* order *Nanopelagicales* at 29% average relative abundance (av.ra), the
162 eukaryote phytoplankton genus *Stramenopiles* 12% av.r.a, the *Actinobacterial* clade 'acIV'
163 (C111) 11% av.ra, the *β-proteobacterial* family *Comamonadaceae* 5% av.ra, the *α-*
164 *Proteobacterial* candidate genus "*Ca.* Fonsibacter" 4% av.ra, and the *Verrucomicrobial*
165 family *Cerasicoccaceae* 3% av.ra (Fig. S1, Table S1). There were few significant
166 correlations or differences (Pearsons correlation or Kruskal-Wallis (KW) with post-hoc
167 testing to identify which pairs were significantly different (KWmc)) between dominant
168 clade abundance and a broad range of environmental predictors such as lake-cycle/season
169 or physico-chemical conditions of the water (Fig. S2, Table S2). The three significant linear
170 correlations detected had only moderate strength (r >= 0.4) with cloud or wedge-shaped
171 distributions (Fig. S3).
172
173
174

175 **Microbial networks**

176 The OTU network from both the time-series and mixed-cultures (Fig. S4) shows
177 modularity and taxonomic assortativity (grouping of related organisms or under-
178 dispersion) at all taxonomic levels with greater assortativity in the mixed-cultures (Table
179 S3). Assortativity increased with taxonomic level peaking at phylum level in the timeseries,
180 while assortativity was lowest at phylum level in the cultures. First neighbours (the primary
181 cohort) of OTUs in networks provide evidence (but not proof) for preferential association
182 between phylotypes, or when negatively correlated may indicate mutual exclusion. The
183 primary cohorts of "*Ca.* Planktophila, "*Ca*. Nanopelagicus", and "*Ca*. Fonsibacter" (Fig.
184 2) are overlapping and include phylotypes from each other's cohorts. Primary cohorts from
185 the time-series and mixed-cultures are similar, though the correlation type and strength
186 vary. There was a moderate correlation between relative abundance of OTUs in the
187 timeseries (Fig. 2a) and relative abundance in cultures (Pearsons $r = 0.58$, $p_{adj} = 0$, Table
188 S4) but no other correlations between relative abundance or prevalence (percentage of
189 samples an OTU was detected in) were observed.

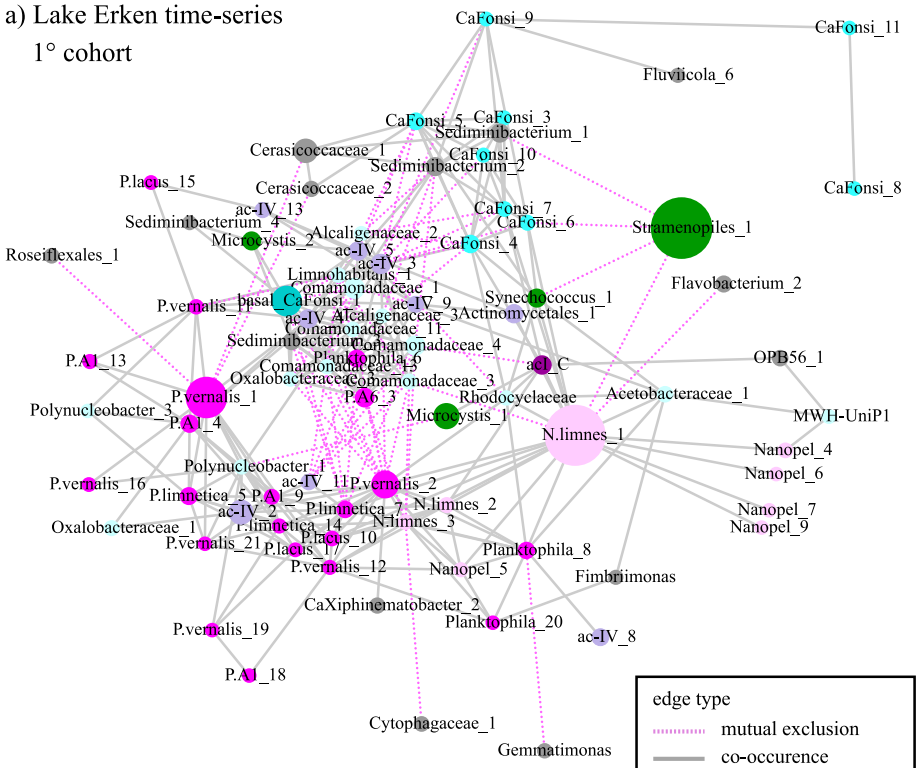
190

191 **Actinobacterial order Nanopelagicales (acI)**

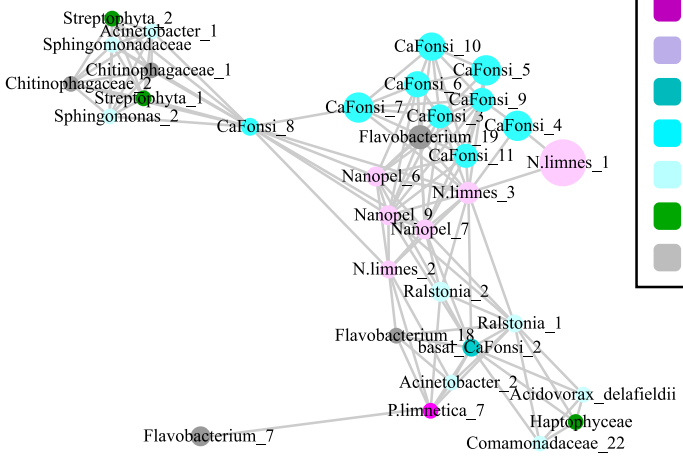
192 *Nanopelagicales* in Erken had higher relative abundances in the hypolimnion, (KW
193 $p<0.001$, KWmc $p<0.001$) and were negatively correlated with % O_2 (Pearsons $r = -$
194 0.4, $p<0.001$). *Nanopelagicales* were the most abundant clade in Lake Erken and were

195

a) Lake Erken time-series
1° cohort



b) Mixed-cultures from lake water
1° cohort



196 Fig. 2. OTU networks of the primary cohorts from amplicons of a) time-series aand b)
197 mixed-cultures. Nodes sized by average relative abundance in time-series and by
198 percentage of cultures detected in for mixed-cultures. Pink dotted lines indicate mutual
199 exclusion of OTUs while grey solid lines are co-occurrence.
200
201 represented by 31 OTUs in the lake time-series network with nine "*Ca*.
202 Nanopelagicus", 21 *"Ca.* Planktophila*"*, and one acI-c (Fig. 2a, Fig. S4). A subset of six
203 "*Ca*. Nanopelagicus" and one *"Ca.* Planktophila*"* phylotype were detected in the mixed-
204 culture based network (Fig. 2b, Fig. S4).
205 The time-series primary (1°) cohort (i.e. the cohort immediately surrounding) of *"Ca.*
206 Planktophila*"* included: other *Actinobacteria*, *Armatimonadetes*, *Chloroflexi*,
207 *Gemmatimonadetes*, *α*- and *β-Proteobacteria*, and *Verrucomicrobia*, while in the
208 mixed-cultures only *Actinobacteria*, *Bacteroidetes*, and *α*- and *β-Proteobacteria* co-
209 existed. A pairwise comparison of the phylogenetic distance and correlation
210 (averaged edge value from the correlation network) was conducted to see if there was
211 a relationship between OTU genetic relatedness and co-occurrence; to distinguish
212 between ecotype and directed reciprocation models. Within the *"Ca.* Planktophila*"* 1°
213 cohort, there was no support for a correlation between phylogenetic distance and co-
214 occurrence (Fig. 3) and *"Ca.* Planktophila*"* was the only genera tested that had

215 negative intra-genus correlations. *"Ca.* Planktophila*"* had more intra-phyla (i.e. non-
216 *Nanopegicales Actinobacteria*) correlations than the other genera tested. Unlike the
217 other genera, *"Ca.* Planktophila*"* had no correlations (positive or negative) with
218 phytoplankton in its 1° network (Fig. 3). Additionally, *"Ca.* Planktophila*"* were the
219 dominant Actinobacterial genus in the time-series with "*Ca*. P. vernalis*"* being most
220 abundant (Fig. 2a). Conversely, in the mixed-cultures, only one *Planktophila* phylotype
221 was detected, "*Ca*. P. limnetica*"* (Fig. 2b).
222 *"Ca.* Planktophila spp." were successfully maintained in axenic culture via the addition
223 of bovine catalase and therefore the presence, absence, and functionality of the
224 relevant *Nanopelagicales* catalase-peroxidase (KatG) is debated. We investigated the
225 presence of the *katG* gene, its gene neighbourhood, and the homology and topology of
226 its translated product, KatG, to establish if there is evidence to support its activity level or
227 effect on host abundance. "*Ca*. P. vernalis*"* (acI-A7) KatG had highest homology to the high
228 functioning *Synechoccocus* version (Fig. S5) and the gene is found near a tRNA (Fig. S6).
229 "*Ca*. P. limnetica*"*'s KatG had highest identity with the assayed low activity KatG from "*Ca*.
230 P. rubra*"* (IMCC 25003) and was located within a putative DNA repair region (Fig. S5 &
231 S6). When modelled these KatGs were most similar to *Burkholderia pseudomaleii*s with a
232 predicted β-sheet around one of the active site residues instead of an α-helix, and are 10 to
233 20 amino-acid residues longer than known high functioning versions i.e. *Burkholderia* type

234 A (Fig. S5). All known active site residues involved in heme interaction or enzyme

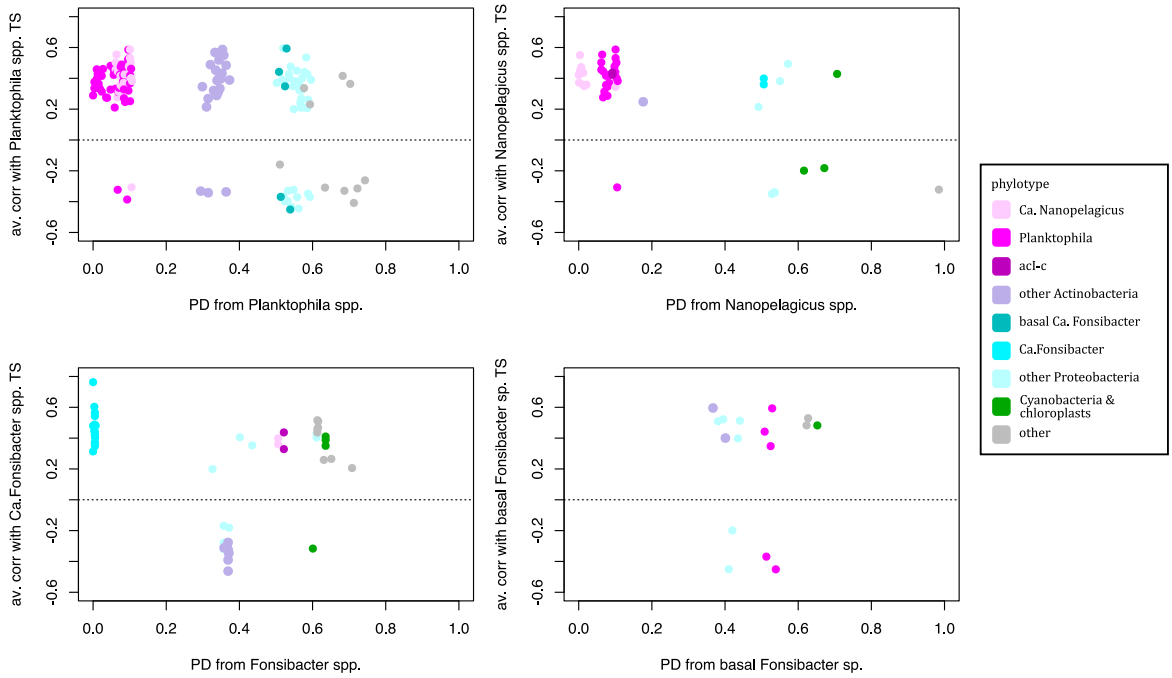
235 switching were however present. *"Ca*. Planktophila limnetica*", "Ca*. P. vernalis*", "Ca*. P.

236 lacus*", and "Ca*. N. limnes*"* all have the repair cassette without *katG*. Metagenomic *katG*

237 sequences were recovered matching *"Ca*. P. dulcis*", "Ca*. P. limnetica*", "Ca*. P.

238 versatilis*"*, and *"Ca*. P. vernalis*"*.

239



240

241 Fig. 3. The relationship between OTU phylogenetic distance and correlation of the primary

242 cohort phylotypes to selected genera: *"Ca. Planktophila"* upper-left, "*Ca*. Nanopelagicus"

243 upper-right, "*Ca*. Fonsibacter" lower-left, and basal "*Ca*. Fonsibacter "lower-right.

244

245 *"Ca*. Nanopelagicus" 1° cohort in the time-series included other *Actinobacteria*,
246 *Bacteroidetes*, *α*- and *β-Proteobacteria*, and the phytoplanktonic *Synechococcus* and
247 *Stramenopiles* (identified by chloroplast sequences), while in the mixed-cultures
248 there was a subset of these phyla with the addition of one *γ-proteobacterial* phylotype
249 (Fig 2). To calculate whether the presence of an OTU in the mixed-cultures could be
250 attributed to the probability of it being in the inoculum, the total and average relative
251 abundances, and the number of samples an OTU was detected in (prevalence) was
252 compared across the two datasets with Pearsons correlation. There was no
253 correlation between "*Ca*. Nanopelagicus" OTU relative abundance or prevalence in the
254 lake time-series as compared to the mixed-cultures (Table S4). "*Ca*. Nanopelagicus"
255 was overall more prevalent in mixed cultures than in the lake time-series. "*Ca*.
256 Nanopelagicus" had strong correlation with "Ca. Fonsibacter" in both datasets. There
257 were between 3 and 19 other phylotypes (non-*Nanopelgicales*, non-Fonsibacter)
258 detected in cultures with "*Ca*. Nanopelagicus" (Table S5).
259 The most abundant "*Ca*. Nanopelagicus" phylotypes in both datasets were assigned to the
260 *katG*$^{-}$ "*Ca*. Nanopelagicus limnes". Several metagenome-recovered putative "*Ca*.
261 Nanopelagicus" *katG* genes were identified in IMG databases and while these KatGs had
262 greatest homology with *Burkholderia* fold type B, they were not located within the repair

263 cassette and there was no synteny between them. No "*Ca*. Nanopelagicus" katG sequences
264 were recovered from our metagenomes.
265
266 **α-Proteobacterial genus "Ca. Fonsibacter" (LD12)**
267 "*Ca*. Fonsibacter spp." had higher relative abundance from winter to spring and moderate
268 negative correlation with water temperature (r = -0.4, p<0.001, Table S2). "*Ca*.
269 Fonsibacter" was represented by 10 OTUs in each of the time-series and mixed-culture
270 networks, with nine OTUs common to both and the alternating OTUs being the least
271 abundant and least prevalent in both environments. The "*Ca*. Fonsibacter" 1° cohort from
272 the time-series includes other "*Ca*. Fonsibacter", *Bacteroidetes*, *α*- and *β-Proteobacteria*,
273 *Actinobacteria* including "*Ca*. Nanopelagicus" and acI-c, and phytoplankton
274 *Synechococcus* and *Stramenopiles*. The "*Ca*. Fonsibacter" 1° cohort from mixed-culture
275 experiments also included other "*Ca*. Fonsibacter" phylotypes, "*Ca*. Nanopelagicus",
276 *Bacteroidetes*, other *α*- and *β-Proteobacteria* and macro-algal eukaryotes of the
277 *Streptophyta* order. There was a strong correlation detected between "*Ca*. Fonsibacter"
278 OTU prevalence in the lake time-series compared to the prevalence of the respective OTU
279 in the mixed-cultures (corr = 0.82, n=11, p adj = 0, Table S4). Based on OTU identity,
280 phylogenetic branch placement, and network connections, two "*Ca*. Fonsibacter"
281 phylotypes were designated as basal and analysed separately for 1° cohort. Unlike the core

282 "*Ca*. Fonsibacter" phylotypes, these basal phylotypes had no negative correlations with
283 phytoplankton and no correlations with other "*Ca*. Fonsibacter" phylotypes (Fig. 3).
284
285 **Other primary cohort members**
286 Both the *Actinobacterial* 'acIV' (C111) and *Comamonadaceae* families were more
287 abundant in winter when the lake was covered by ice (Fig. S2, KW $p<0.001$, KWmc
288 $p<0.001$). While 'acIV' was abundant and ubiquitous in the time-series (Fig. S1, Table S1),
289 and also prominent in the *"Ca*. Planktophila" time-series 1° cohort (Fig. 2a), no phylotypes
290 were detected in the mixed-cultures. *Comamonadaceae* was abundant but patchy in the
291 time-series and mixed cultures.
292
293 **DISCUSSION**
294 **Taxon relative abundance and environmental correlates**
295 If microbial abundance is directly linked to nutrient availability or other physico-chemical
296 parameters, then a strong correlation between geo-chemical parameters and relative
297 abundance is expected. The lack of strong correlation between OTUs or clades and
298 environmental parameters in Lake Erken indicates that during the 8 years of sample
299 collection, microbial community composition was not directly and strongly controlled, but
300 only influenced, by the measured abiotic environmental variables. While short-term studies
301 have shown significant strong trends and correlations (12), such correlations tend to be

302 eclipsed in studies spanning multiple years by shifts operating at longer frequencies and
303 biotic interactions (36–39). Overall, the strength and number (identified via consensus
304 network) of correlations between taxon relative abundance and the weakness of
305 environmental predictors for most of the taxa suggest that biological aspects (including
306 predation, which we did not study) are the greater determinants of microbial population
307 dynamics, composition and persistence in Lake Erken. It stands therefore that the
308 dominance of the three streamlined genera in the lake is attributable more to their
309 interactions with their community than selection by environment.
310

311 **Microbial networks**

312 Assortativity is a predicted product of ecotype or niche differentiation at species level and
313 environmental filtering at higher taxonomic levels (40). Ecotype divergence of strains
314 reduces competition between different members of a population. The detection of highest
315 assortativity in the time-series networks at phylum level supports environmental selection
316 as important in the lake habitat but not at taxonomic levels lower than phyla. It is unlikely
317 that environmental filtering is the main driver behind the success of these three genera in
318 Lake Erken. High assortativity in the mixed cultures at all levels except phylum, supports
319 that niche differentiation contributed to co-occurrence in the cultures. Neither result
320 however fully supports, nor refutes, ecotype theory as the explanation of coexistence or
321 dominance by "*Ca.* Planktophila", "*Ca.* Nanopelagicus", or "*Ca.* Fonsibacter" genera.

322 The strong correlation between "*Ca*. Fonsibacter" (and also *Comamonadaceae*) prevalence
323 in the time-series and mixed-cultures supports that its occurrence in the cultures may be
324 attributed to being present in the inocula and that the filtered lake water media was able to
325 meet an important proportion of its metabolic needs. Conversely, the lack of correlation in
326 either relative abundance or prevalence, for the *Nanopelagicales* genera (and also the
327 'acIV' family) in the time-series compared to mixed-cultures means that there is not support
328 for their detection being merely due to high-abundance in the inocula. We infer that the co-
329 presence of other phylotypes were bigger contributors to their growth. It should be noted
330 that negative correlations are not detectable in the mixed-cultures as the absence of a
331 phylotype could be stochastic rather than a result of competitive exclusion. Moreover,
332 competition that prevents growth in a culture cannot be identified.
333
334 **Actinobacterial order Nanopelagicales 'acI'**
335 *Nanopelagicales* in Erken were likely more competitive in colder, darker, and oxygen
336 depleted waters. While *Nanopelagicales* has been associated with oxygen saturated waters
337 in some locations (41) this lineage has been associated with colder deeper water at others
338 (30), though the variation in association could be due to the considerable genetic diversity
339 encompassed in this bacterial order (42). It is also probable that solar-driven formation of
340 oxygen radicals in surface waters inhibit the growth of *katG*$^{-}$ species of *Nanopelagicales*.

341 Of the seven *Nanopelagicales* phylotypes in the mixed-cultures, six were members of the
342 candidate genus "*Ca*. Nanopelagicus" suggesting that this genuss dependencies were more
343 easily met compared to the *"Ca.* Planktophila spp.*"* found in Lake Erken. However,
344 genomic evidence so far has shown the opposite as "Ca. Nanopelagicus spp." have more
345 numerous auxotrophies than *"Ca.* Planktophila spp." (23). It is therefore possible that the
346 difference in culturability seen in this experiment was due to reduced costs via gene loss in
347 combination with suitable metabolic partners being in the inoculum. In particular the lower
348 maintenance costs due to the loss of *katG* (25, 33) could increase fitness of the "*Ca*.
349 Nanopelagicus". The successful maintenance of *"Ca.* Planktophila*"* in culture pinpointed
350 the low functionality of some *Nanopelagicales* catalase-peroxidases as the culprit for the
351 difficulties in growing these bacteria. Peroxide is proposed to be an up-regulator of a DNA
352 repair (43) and in the *"Ca.* Planktophila*"* with the putative low functioning KatG type, the
353 gene is embedded in a region of DNA hosting other genes for repair of damage done to
354 DNA under oxidative stress. *P. limnetica* (acI-A3) genomes have this repair cassette but
355 their *katG* is located elsewhere. The most parsimonious explanation for *Nanopelagicales*
356 *katG* presence/absence/homology is that the now low functioning version located in a
357 repair cassette is in the process of being lost and that the *"Ca.* Planktophila spp.*"* with a
358 high functioning KatG obtained this version via a HGT event. While *P. vernalis* in Lake
359 Erken is likely $katG^{+}$, it is likely that *P. limnetica* and *N. limnes* from Lake Erken rely on
360 "*Ca*. Fonsibacter" in combination with either *Flavobacterium* or *Comamonadaceae* to

361 detoxify peroxide. This would fit within the community cooperation model. While *"Ca.*
362 Nanopelagicus" has more numerous auxotrophies compared to *"Ca.* Planktophila*"* (23)
363 the higher culturability of *"Ca*. Nanopelagicus" supports that it may be the cost to
364 habouring *katG*, whether high or low functioning, and as *katG* is larger than average (44),
365 and as long as a functional aquaporin exists (45), and cells don't grow too fast, then it
366 would be an unnecessary cost, as cells can rely on other community members for oxidative
367 stress reduction. Further the reduction in amino-acid synthesis pathways and therefore
368 greater reliance on exogenous sources of amino-acids in "*Ca*. Nanopelagicus" appears to
369 have paid off if their greater cultivability in these minimalist cultures is any indicator. This
370 fits within the community detritus model.
371
372 **α-Proteobacterial genus "Ca. Fonsibacter" (LD12)**
373 Earlier short-term studies found that "*Ca*. Fonsibacter" is highly abundant in Lake Erken
374 in the late summer-early autumn immediately after an algal bloom i.e. the 'clear water'
375 phase of the lake's cycle (12). However, we found support for higher long-term relative
376 abundances in winter and spring and an associated negative correlation with water
377 temperature. These differences in findings could be due to changes in abundance between
378 years of different ecotypes (with different temperature and nutrient preferences) that were
379 not detected via 16S surveys. The ecotype differentiation within "*Ca.* Fonsibacter" is
380 centered around the HVR1 which is thought to encode genes for defense, i.e. predator

381 escape via alteration of cellular membranes (3, 46). There appears to be limited intra-
382 population variability in anabolic or catabolic traits (14, 24, 47) and as such, ecotype
383 differentiation for overcoming auxotrophies is not supported. While no direct evidence
384 exists for "*Ca*. Fonsibacter", its similarity to the marine sister clade Pelagibacter suggests
385 that it has overcome auxotrophic limitations by scavenging metabolites and other
386 compounds produced by phototrophs (48). In Erken this would primarily be dying and
387 senescent *Stramenopiles* as reflected in the apparent negative correlation. Evidence points
388 to "*Ca*. Fonsibacter" relying on detritus to meet its dependencies and ecotype
389 differentiation for defense.
390 The combination of network and other correlation analyses with review of published
391 genomes supports that the most of *Nanopelagicales* have anabolic ecotype cooperation
392 with reliance on a detoxifying community cohort as main structuring factors. A few *"Ca.*
393 Planktophila*"* are putatively KatG+ but as they were not detected in Lake Erken it was not
394 possible to determine the most appropriate cooperation models in this study. "*Ca*.
395 Fonsibacter" may escape predation via ecotype differentiation, while broadly relying on
396 community detritus for anabolic deficits.
397
398 **MATERIALS AND METHODS**
399 **Time series water sampling and environmental parameters**

Time series water samples were obtained from dimictic Lake Erken (59.1510N, 18.1360E) central Sweden at monthly to weekly intervals spanning a total period from February 2007 to November 2015 (Table S6). Samples for genomic and chemical analysis were taken from the 20 m water column at one meter intervals from the deepest point of the lake (Fig. S7). Samples for the total water column were pooled during periods of mixing. Samples of the upper oxygenated water column (epilimnion), middle (metalimnion), and lower oxygen depleted (hypolimnion) were pooled into separate composite samples during summer stratification. Annual dates of mixing, stratification, ice-over, and ice-thaw varied, as did depth of stratigraphic layers (Table S6). Temperature and oxygen concentrations were analyzed every meter with a portable Oxi 340i oxygen meter equipped with a Cellox 325-20WTW probe and were used to determine stratification. Chemical analysis of pooled water samples was carried out by the Lake Erken Field Station ISO certified laboratory and included pH, conductivity, nitrate, nitrite, alkalinity, turbidity, suspended matter, phosphate, ammonium, total and particulate phosphorus and nitrogen, chlorophyll-a, suspended particulate organic matter, silicate, water-colour and absorbance at 420nm measurements. Water samples for genomic analysis were collected by gentle vacuum filtration onto 0.2 μm membrane filters (Supor-200 Membrane Disc Filters, 47mm; Pall Corporation, East Hills, NY, USA). Filters were individually stored in Eppendorfs and transported back to the University Uppsala laboratory where they were stored at -80 °C until further processing.

420

421 **DNA extraction and SSU rRNA gene amplicon preparation**

422 DNA was extracted from filters using MoBio Ultra-clean soil DNA extraction kits (MoBio,
423 Carlsbad) as per manufacturers' instructions. The V3-V4 region of the 16S rRNA gene was
424 amplified using the S-D-Bact-0341-b-S-17 (Bakt_341F: CCTACGGGNGGCWGCAG)
425 forward primer and S-D-Bact-0785-b-A-21 (805RN: GACTACNVGGGTATCTAATCC)
426 reverse primer (49). Template DNA was amplified in duplicate 20 µl reactions containing
427 1.0 U Q5 high fidelity DNA polymerase (NEB, UK), 0.25 µM primers, 200 µM dNTP mix,
428 and 0.4 µg BSA. The thermocycler program was an initial denaturation step at 95 °C for
429 30 s, followed by 20 cycles of dissociation at 95 °C for 10 s, annealing at 53 °C for 30 s
430 extension at 72 °C for 30 s, with a final extension of 2 min at 72 °C. Amplicons were pooled
431 and purified with Agencourt AMPure XP purification system (Beckman Coulter, Danvers,
432 MA, USA). Purified amplicons (2 µl) were amplified in second step 20 µl reactions to
433 introduce sample MIDS to each end of the amplicons. Reactions contained 1.0 U Q5 high
434 fidelity DNA polymerase (NEB, UK), 0.25 µM primers, 200 µM dNTP mix, and 0.4 µg
435 BSA. Thermocycler program was an initial denaturation step at 95 °C for 30 s, followed
436 by 15 cycles of dissociation at 95 °C for 10 s, annealing at 66 °C for 30 s, extension at 72
437 °C for 30 s, with a final extension of 2 min at 72 °C. Amplicons were again purified using
438 Agencourt AMPure XP purification system (Beckman Coulter, Danvers, MA, USA) then
439 quantified using PicoGreen (Invitrogen) before pooling at equi-molar amounts for each

440 run. Amplicons were sequenced at SciLifeLab SNP/SEQ service on the MiSeq Illumina
441 platform with 300bp paired-end libraries. MiSeq data was processed, including de-
442 multiplexing, by the sequencing centre using Illumina pipelines where all reads with more
443 than 8% mismatch to adapter-MID sequences were discarded.
444
445 **Water sampling, culture conditions, and DNA amplification for mixed-cultures**
446 Water as both a source of growth-media and culturable-cells was collected from the
447 epilimnion of Lake Erken on March 8th 2016. Water for growth-media was filter-sterilised
448 twice through Sterivex filter cartridges 0.22 µm (Millipore). Ten minutes exposure to UV
449 light further disinfected and disrupted cellular and viral nucleic acids. Ten microliters of
450 untreated lake water (10^6 cells/ml) were used as a source of cells for inoculation of mixed-
451 cultures. The untreated lake water was diluted in 1 L of the triple sterilized lake water
452 (approximately 10 cells/ml). Ninety-eight 1 ml cultures were placed in sterile 50 ml falcon
453 tubes to allow oxygenation and incubated for two months under 12-hour light-dark cycles
454 at 11 °C to replicate lake conditions at time of collection. The V3-V4 region of the 16S
455 rRNA gene was amplified using the S-D-Bact-0341-b-S-17 (Bakt_341F:
456 CCTACGGGNGGCWGCAG) forward primer and S-D-Bact-0785-b-A-21 (805RN:
457 GACTACNVGGGTATCTAATCC) reverse primer (49) One microliter of the mixed-
458 culture after one freeze-thaw cycle was used as template and Q5® High-Fidelity DNA
459 Polymerase. The PCR conditions consisted of an initial step at 98°C for 10 minutes.

460 Thereafter, 35 cycles of denaturation at 98°C for 10 s, annealing for 30 s and extension at
461 72°C for 30 s. Annealing temperature was 48°C which has been tested to get unbiased
462 products of non-mismatch and 3-mismatch isolates. Final extension was performed at 72°C
463 for 2 minutes. PCR products were purified with magnetic beads (Angecourt AMPure). A
464 second PCR was conducted for attaching standard illumina handles and index primers as
465 described above. Primers and MIDS detailed in Table S7. Prior to sequencing a portion of
466 amplification product was run on gel to check for amplification and used as proxy for
467 growth of culture. Only data from the 60 cultures with a visible band after the first PCR
468 were included in analyses.
469
470 **Amplicon bioinformatics**
471 Demultiplexed MiSeq time-series and mixed-culture data, and 454 Ti time-series data
472 obtained from Lake Erken 2008 (11), were pre-processed separately and joined once
473 chimeras had been removed. Pre-processing of MiSeq data started with SeqPrep v1.2
474 (https://github.com/jstjohn/SeqPrep.git) with default settings to join paired end
475 reads while discarding reads shorter than 30b and all un-paired reads. 454 and MiSeq
476 reads with ambiguous bases, low quality average, and low quality sections, were
477 discarded and trimmed with Qiime v1.9 (50) split_libraries and multiple split libraries
478 fastq respectively. Singletons and sequences identified as chimeric were removed

479 with Qiimes' identify_chimeric_seqs, filter_fasta using usearch v6.1 (51), with default
480 settings and option parameter 'intersection' to select out sequences identified by both
481 de-novo and reference-based methods. Both GreenGenes (v105 13-8) and SILVA
482 (v128) reference datasets were used (52, 53). Sequences from the two Miseq runs and
483 the 454 data from 2008 were then combined. Reads were clustered at 97% identity
484 into OTUs then assigned taxonomy based on SILVA v128 using QIIMES closed pick
485 OTU method (uses only template recognition and does not allow for de-novo
486 sequences) to allow for differences in DNA processing and sequencing across the
487 three datasets. Sequences from the negative control were removed from the entire
488 dataset. Samples with potential human-microbiome contamination (Table S8) were
489 identified and discarded from the dataset due to potential effect on community
490 metabolism. Sequences identified as eukaryotic (including fungal, mitochondrial and
491 chloroplast origin) or archaeal were deliberately retained despite the primers not
492 specifically being selected to amplify such sequences. as they were considered critical
493 in the mixed-culture analyses from a community metabolism perspective. Samples
494 were normalized using Qiimes single rarefraction to even depth of 1500 reads per
495 sample for the time-series and 500 reads for the mixed-culture samples. These
496 normalised OTU tables were then used for all downstream analyses.
497

498 **Statistical analyses of time-series environmental correlates and selected taxa**

499 Statistical analyses were done using R v3.3.1 (54) in the RStudio IDE v0.99.903 (55)
500 and graphs processed for publication in Inkscape v0.91. Environmental parameters
501 were removed if more than 40% of the values were missing then remaining twenty-
502 one tested for co-linearity with the usdm v1.1.18 package with vifstep followed by
503 vifcor functions and removed if VIF >10 or correlations were over 0.7 sequentially
504 starting with those with highest number of missing values. The remaining 14
505 environmental predictors were tested for statistical differences in connection to lake-
506 cycle with Kruskal-Wallis (kruskal.test of the base stats package) and graphed using
507 beanplot v1.2 (56) and scales v0.4.1. Statistical significance of differences between
508 individual paired cycle points were determined post-hoc with the Kruskal-Wallis
509 multiple comparison (KWmc) using thekruskal.mc test of the pgirmess package
510 v1.6.4 (57). Environmental parameter correlates with selected taxa relative
511 abundances were screened using Pearsons correlation with an alpha level of 0.001
512 and p-values estimated and corrected for multiple comparisons with Benjamini,
513 Hochberg, and Yekutieli (BH) false discovery rate using the psyche package v1.6.12
514 corr.test function and Pearsons moment r reported. Parameters with weak to
515 moderate correlation (r >|0.3|) were plotted in R to observe monotonicity.

516

517 **Network analyses**

518 The normalized time-series OTU table was further processed for network and correlation
519 analyses to remove rare OTUs that, while potentially important within the community, are
520 known to inhibit identification of ecological relationships between OTUs and also between
521 OTUs and their environment. Recommendations include OTU tables with 50% or less
522 sparcity and n_{eff} (inverse Simpsons) above an average of 10 and the use of multiple methods
523 (58). A range of filtering parameters based on frequency of OTU detection, from 2 samples
524 up to 46 (20% of samples) and total relative abundance ranging from 1 up to 150 (0.1%),
525 was used on the OTU table to remove type I errors in detection of co-occurrence networks
526 (58). The best filter combination was found with OTUs at a total relative abundance above
527 0.1% and occurring in at least 29 samples. All OTU tables were analysed in R (54) using
528 the vegan v2.4-1 package (59) diversity function to find the n_{eff} and the base stats v3.3.1
529 package to find min, max and mean of the n_{eff}. The reduced OTU table was analysed with
530 SparCC (60), Pearsons correlation coefficient, and SPIEC-EASI (61). SparCC analyses
531 were carried out in Rstudio (55) using both the rsparcc and spiec.easi packages with rsparcc
532 parameters: max.iter=100, th=0.1, exiter=10 and pseudo p-values generated with
533 spiec.easis sparccboot (100 bootstraps) and pval.sparccboot scripts (two-sided test).
534 Pearson correlations were performed using the corr.test of psyche package using the
535 optional p-values adjustment for multiple testing of Benjamini, Hochberg and Yekutieli
536 false discovery rate. SPIEC-EASI correlations were analysed using the spiec.easi package

537 with parameters: method="mb", sel.criterion="stars", lambda.min.ratio=1e-2,
538 nlambda=20, and 50 repetitions. Singletons were removed from the normalized mixed-
539 culture OTU table prior to network analysis. The reduced OTU table was analysed with
540 Pearsons correlation, SpiecEasi (61) and on the binary version of the table, Dice-Sørensen
541 index (62, 63).
542 Correlations between OTUs in the time-series and mixed cultures were retained if the
543 correlation co-efficient was greater than 0.3 absolute for Pearsons, SparCC, or Dice-
544 Sørensen, or if detected at all by SPIEC-EASI. Time-series and culture networks consisting
545 of OTU connections detected by at least two correlation methods were visualized in
546 Cytoscape. All further statistical analyses of the networks were performed in Cytoscape
547 except for assortativity analyses which were done in R using the igraph v1.1.2 package
548 nominal assortativity function. Sub-networks of the identified dominant genera ("*Ca.*
549 Planktophila", "*Ca*. Nanopelagicus", and "*Ca*. Fonsibacter") and their primary cohorts
550 were extracted from the two initial networks and visualized separately. Layout for the
551 networks were "edge-weighted spring-embedded" with edge weights calculated from the
552 average of the correlations. Some nodes (OTUs) were moved to the side of overlapping
553 nodes for clarity or away from an edge if it appeared to pass through an unconnected node.
554 Correlations between *Nanopelagicales* OTU and "*Ca*. Fonsibacter" OTU detection in the
555 time-series versus the mixed cultures were investigated by Pearson correlation of their
556 relative abundance (log av.ra), prevalence (percent samples or cultures detected in) and

557 relative abundance versus prevalence to determine if detection in culture was correlated to
558 occurrence in the lake.
559
560 **Metagenomic sequencing**
561 Samples for metagenomic sequencing were selected by highest relative abundance of
562 clades of interest according to amplicon data, greatest available DNA mass, and greatest
563 variation in environmental parameters. This was done to examine samples with the greatest
564 potential divergence of environmental conditions under which microbes of interest were
565 most abundant. These included samples taken during lake circulation and summer
566 stratification which separately including hypo, meta and epi-limnionic samples. Two
567 samples per category were selected from different years. Lower biomass during ice cover
568 meant there was insufficient DNA for metagenomic sequencing. DNA library for
569 metagenomic sequencing was prepped using the Illumina TruSeq Nano, 550bp NeoPrep
570 and sequenced with Illumina HiSeq 2500 High Output V4 PE 2x125bp at SciLifeLab
571 Stockholm.
572
573 **Metagenome bioinformatics**
574 HiSeq Illumina paired end reads were pre-processed first using SeqPrep v1.2 (64) to filter
575 out low quality reads using default settings resulting in removal of all reads less than 30b
576 in length, un-paired reads and initial clipping of adapter sequences. Next sequences were

577 cleaned using Trimmomatic v0.36 (65) with removal of any remaining Illumina adapters
578 allowing for 2 mismatches; followed by 4b sliding window trim with minimum average
579 quality of 15; then trimming of read ends with 'maxinfo' value less than 1. The eight
580 metagenomes were individually assembled with megahit (66) using iterative kmers from
581 21 to 121 at 10b intervals, minimum coverage of 2, and minimum length of 200.
582 Concatenated assemblies were cleaned of Illumina artifact low complexity (long
583 homopolymer) contigs with prinseq-lite v0.20.4 (67) removing contigs consisting of
584 greater than 80% one nucleotide.
585
586 **Investigation of katG gene and its product: catalase-peroxidase I**
587 Known impediments to axenic growth of freshwater isolates are multiple auxotrophies, and
588 slow growth and replication rates. The successful axenic cultivation of "*Ca*. Planktophila"
589 strains by addition of high functioning catalase to reduce oxidative stress, in addition to
590 catering for predicted auxotrophies, showed that there can be other considerations. To
591 investigate how both "*Ca*. Planktophila" and "*Ca*. Nanopelagicus" phylotypes grow in
592 Lake Erken and why "*Ca*. Nanopelagicus" grew more frequently in the minimalist mixed
593 cultures, the katG gene was investigated. First katG amino-acid sequences for species
594 identified as belonging to the Nanopelagicales and its primary cohort were retrieved from
595 IMG and GenBank (68, 69). Second sequences for whom the gene product has been
596 assayed and/or the 3D crystal structure is known were identified via Phyre2 and then

597 collected from RCSB PDB (rscb.org (70, 71)). Next these sequences were used as BLAST
598 (72) queries to retrieve closely related (> 80% aa identity) sequences from Lake Erken
599 metagenomes. The collected amino-acid sequences were aligned using MAFFT e-insi and
600 "leavegappy" options (73). Sequences with greater than 99% identity were identified using
601 CD-Hit and removed (74) while sequences from uniRef useful for constructing a guide tree
602 were then also added (75, 76). The alignment was evaluated for best amino-acid
603 substitution model using IQ-tree v.1.6.12 modelfinder (77). This alignment was then
604 analysed and used to construct a tree in RAXML v.0.9.0 (78) with WAG+IG+R4 model
605 and 1000 bootstraps and catalase-peroxidase from *Oryza rufipogon* set as outgroup. To
606 examine functionality secondary structure was predicted and aligned to known models then
607 tertiary structure predicted in Phyre^2 (71). Further the location of katG genes in
608 Nanopelagicales genomes and MAGs was examined in IMG using the gene-
609 neighbourhood function. Images were downloaded with screen shot and processed in
610 Inkscape.
611
612 **Data availability**
613 Time-series MiSeq sequences were deposited in ENA
614 (https://www.ebi.ac.uk/ena/browser/view/PRJEB20109) under accession numbers:
615 ERS1884053 to ERS1884254, sample details including MIDS are also in Table S6. Mixed-
616 culture sequences are available from ENA PRJEB35605

(https://www.ebi.ac.uk/ena/browser/view/PRJEB35605) with details also in Table S7. 454 data used in the time-series is available at SRA (https://trace.ncbi.nlm.nih.gov/Traces/sra/) under accession SRR097418 (11). HiSeq metagenome sequences are available at ENA project PRJEB37497 under accession numbers: ERS4415328 to ERS4415335 and the assembled metagenomes are available at IMG (https://img.jgi.doe.gov/cgi-bin/m/main.cgi) with TOIDs: 3300020141, 3300020151, 3300020159, 3300020160, 3300020161, 3300020162, 3300020172, & 3300020205 details in Table S9. Rscripts used for correlation and network analyses are available from https://github.com/rmondav/publications.

ACKNOWLEDGEMENTS

RM, sequencing of the metagenomes, and publication supported by grants from the Malméns Stiftelsen. SG was supported by VR (grant 2012-4592), Olsson-Borgh, Knut and Alice Wallenberg Foundation (grant KAW 2013.0091), SciLifeLab Fellowship, and Kungl. Vetenskapsakademiens stiftelser (BS2017-0044). SB was supported by grants from the Swedish Research Council (VR) and the Swedish Research Council Formas. The sequencing of the Lake Erken 16S rRNA time-series dataset and chemical data was supported by the Swedish Infrastructure for Ecosystem Science (SITES) with technical assistance from Dr. Omneya Ahmed, Pilar López Hernández, Helena Enderskog, Erika Bridell and Kristiina Mustonen. Sequencing of the mixed cultures was supported by a collaborative (grant from the K&A2012-4592), Olsson-Borgh, Knut and Alice Wallenberg

foundation (PI. Lars Tranvik). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. We gratefully acknowledge the computing resources provided by SNIC through Uppsala Multidisciplinary Centre for Advanced Computational Science (UPPMAX) under UPPNEX projects 2015047, 2016272, & 2017147 and sequencing infrastructure support from the SciLifeLab National Genomics Infrastructure.

AUTHOR CONTRIBUTIONS

SB, SL, and RM contributed to the genomic sampling design, organization, and selection. SLG and MB designed and implemented the mixed-culture experiments. RM designed and performed the bioinformatic analyses. RM and SLG synthesized the conceptual framework which was drafted by RM in consultation with SLG, with editorial contributions from all authors.

REFERENCES

1. **Tipton L**, **Darcy JL**, **Hynson NA**. 2019. A developing symbiosis: Enabling cross-talk between ecologists and microbiome scientists. Front. Microbiol. **10**:1–10.
2. **Koeppel A**, **Perry EB**, **Sikorski J**, **Krizanc D**, **Warner A**, **Ward DM**, **Rooney AP**, **Brambilla E**, **Connor N**, **Ratcliff RM**, **Nevo E**, **Cohan FM**. 2008. Identifying the fundamental units of bacterial diversity: A paradigm shift to incorporate

657 ecology into bacterial systematics. Proc. Natl. Acad. Sci. U. S. A. **105**:2504–2509.

658 3. **Rodriguez-Valera F**, **Martin-Cuadrado A-B**, **Rodriguez-Brito B**, **Pašić L**,
659 **Thingstad TF**, **Rohwer F**, **Mira A**. 2009. Explaining microbial population
660 genomics through phage predation. Nat. Rev. Microbiol. **7**:828–836.

661 4. **Fernandez VI**, **Yawata Y**, **Stocker R**. 2019. A Foraging Mandala for Aquatic
662 Microorganisms. ISME J. **13**:563–575.

663 5. **Eriksson C**, **Forsberg C**. 1992. Nutrient Interactions and Phytoplankton
664 Growth during the Spring Bloom Period in Lake Erken, Sweden. Int. Rev. der
665 gesamten Hydrobiol. und Hydrogr.

666 6. **Bell RT**, **Stensdotter U**, **Pettersson K**, **Istanovics V**, **Pierson DC**. 1998.
667 Microbial dynamics and phosphorus turnover in Lake ErkenAdvances in
668 Limnology 51: Lake Erken - 50 Years of Limnological Research.

669 7. **Pettersson K**, **Grust K**, **Weyhenmeyer G**, **Blenckner T**. 2003. Seasonality of
670 chlorophyll and nutrients in Lake Erken - Effects of weather conditions.
671 Hydrobiologia.

672 8. **Haglund A-L**, **Törnblom E**, **Boström B**, **Tranvik L**. 2002. Large differences in
673 the fraction of active bacteria in plankton, sediments, and biofilm. Microb. Ecol.
674 **43**:232–41.

675 9. **Lymer D**, **Lindström ES**. 2010. Changing phosphorus concentration and

676 subsequent prophage induction alter composition of a freshwater viral
677 assemblage. Freshw. Biol. **55**:1984–1996.
678 10. **Burgmer T**, **Reiss J**, **Wickham S**, **Hillebrand H**. 2010. Effects of snail grazers
679 and light on the benthic microbial food web in periphyton communities. Aquat.
680 Microb. Ecol. **61**:163–178.
681 11. **Eiler A**, **Heinrich F**, **Bertilsson S**. 2012. Coherent dynamics and association
682 networks among lake bacterioplankton taxa. ISME J. **6**:330–42.
683 12. **Heinrich F**, **Eiler A**, **Bertilsson S**. 2013. Seasonality and environmental control
684 of freshwater SAR11 (LD12) in a temperate lake (Lake Erken, Sweden). Aquat.
685 Microb. Ecol. **70**:33–44.
686 13. **Eiler A**, **Bertilsson S**. 2004. Composition of freshwater bacterial communities
687 associated with cyanobacterial blooms in four Swedish lakes. Environ.
688 Microbiol. **6**:1228–1243.
689 14. **Henson MW**, **Lanclos VC**, **Faircloth BC**, **Thrash JC**. 2018. Cultivation and
690 genomics of the first freshwater SAR11 (LD12) isolate. ISME J. **12**:1846–1860.
691 15. **Newton RJ**, **Jones SE**, **Eiler A**, **McMahon KD**, **Bertilsson S**. 2011. A guide to the
692 natural history of freshwater lake bacteria.Microbiology and molecular biology
693 reviews : MMBR.
694 16. **Warnecke F**, **Amann R**, **Pernthaler J**. 2004. Actinobacterial 16S rRNA genes

695 from freshwater habitats cluster in four distinct lineages. Environ. Microbiol.

696 17. **Warnecke F**, **Sommaruga R**, **Sekar R**, **Hofer JS**, **Pernthaler J**. 2005.
697 Abundances, identity, and growth state of actinobacteria in mountain lakes of
698 different UV transparency. Appl. Environ. Microbiol.

699 18. **Glockner FO**, **Zaichikov E**, **Belkova N**, **Denissova L**, **Pernthaler J**, **Pernthaler**
700 **A**, **Amann R**. 2000. Comparative 16S rRNA analysis of lake bacterioplankton
701 reveals globally distributed phylogenetic clusters including an abundant group
702 of actinobacteria. Appl. Environ. Microbiol.

703 19. **Li J**, **Zhang J**, **Liu L**, **Fan Y**, **Li L**, **Yang Y**, **Lu Z**, **Zhang X**. 2015. Annual periodicity
704 in planktonic bacterial and archaeal community composition of eutrophic Lake
705 Taihu. Sci. Rep. **5**:1–14.

706 20. **Comte J**, **Lovejoy C**, **Crevecoeur S**, **Vincent WF**. 2016. Co-occurrence patterns
707 in aquatic bacterial communities across changing permafrost landscapes.
708 Biogeosciences **13**:175–190.

709 21. **Garcia SL**, **Buck M**, **Hamilton JJ**, **Wurzbacher C**, **Grossart H-P**, **McMahon KD**,
710 **Eiler A**. 2018. Model Communities Hint at Promiscuous Metabolic Linkages
711 between Ubiquitous Free-Living Freshwater Bacteria. mSphere **3**:1–8.

712 22. **Eckert EM**, **Baumgartner M**, **Huber IM**, **Pernthaler J**. 2013. Grazing resistant
713 freshwater bacteria profit from chitin and cell-wall-derived organic carbon.

714 Environ. Microbiol. **15**:2019–2030.

715 23. **Neuenschwander SM**, **Ghai R**, **Pernthaler J**, **Salcher MM**. 2017.
716 Microdiversification in genome-streamlined ubiquitous freshwater
717 Actinobacteria. ISME J. **12**:185–198.

718 24. **Eiler A**, **Mondav R**, **Sinclair L**, **Fernandez-Vidal L**, **Scofield DDG**, **Scwientek**
719 **P**, **Martinez-Garcia M**, **Torrents D**, **McMahon KDKD**, **Andersson SGESGE**,
720 **Stepanauskas R**, **Woyke T**, **Bertilsson S**, **Schwientek P**, **Martinez-Garcia M**,
721 **Torrents D**, **McMahon KDKD**, **Andersson SGESGE**, **Stepanauskas R**, **Woyke**
722 **T**, **Bertilsson S**. 2015. Tuning fresh: radiation through rewiring of central
723 metabolism in streamlined bacteria. ISME J. **10**:1–13.

724 25. **Kim S**, **Kang I**, **Seo J-HH**, **Cho J-CC**. 2019. Culturing the ubiquitous freshwater
725 actinobacterial acI lineage by supplying a biochemical 'helper' catalase. ISME J.
726 **112**:343640.

727 26. **Rodriguez-R L**, **Tsementzi D**, **Luo C**, **Konstantinidis K**. 2019. Iterative
728 Subtractive Binning of Freshwater Chronoseries Metagenomes Identifies of
729 over Four Hundred Novel Species and their Ecologic Preferences. bioRxiv
730 826941.

731 27. **Giovannoni SJ**, **Cameron Thrash J**, **Temperton B**. 2014. Implications of
732 streamlining theory for microbial ecology. ISME J. Nature Publishing Group.

733 28. **Tripp HJ**. 2013. The unique metabolism of SAR11 aquatic bacteria. J. Microbiol.
734 Seoul Korea **51**:147–53.
735 29. **Pernthaler J**, **Posch T**, **Šimek K**, **Vrba J**, **Pernthaler A**, **Glöckner FO**, **Nübel U**,
736 **Psenner R**, **Amann R**. 2001. Predator-Specific Enrichment of Actinobacteria
737 from a Cosmopolitan Freshwater Clade in Mixed Continuous Culture. Appl.
738 Environ. Microbiol.
739 30. **Salcher MM**. 2014. Same same but different: Ecological niche partitioning of
740 planktonic freshwater prokaryotes. J. Limnol. **73**:74–87.
741 31. **Dadon-Pilosof A**, **Conley KR**, **Jacobi Y**, **Haber M**, **Lombard F**, **Sutherland KR**,
742 **Steindler L**, **Tikochinski Y**, **Richter M**, **Oliver Glamp F**, **Suzuki MT**, **West NJ**,
743 **Genin A**, **Yahel G**. 2017. Surface properties of SAR11 bacteria facilitate grazing
744 avoidance. Nat. Microbiol. 1.
745 32. **Våge S**, **Storesund JE**, **Thingstad TF**. 2013. SAR11 viruses and defensive host
746 strains. Nature **499**:E3-4.
747 33. **Morris JJ**, **Lenski RE**, **Zinser ER**. 2012. The Black Queen Hypothesis: Evolution
748 of Dependencies through Adaptive Gene Loss. MBio.
749 34. **Pacheco AR**, **Moel M**, **Segrè D**. 2019. Costless metabolic secretions as drivers
750 of interspecies interactions in microbial ecosystems. Nat. Commun. **10**:103.
751 35. **Pande S**, **Kost C**. 2017. Bacterial unculurability and the formation of

752 intercellular metabolic networks. Trends Microbiol. **xx**:1–13.

753 36. **Hughes BB**, **Beas-Luna R**, **Barner AK**, **Brewitt K**, **Brumbaugh DR**, **Cerny-**
754 **Chipman EB**, **Close SL**, **Coblentz KE**, **de Nesnera KL**, **Drobnitch ST**, **Figurski**
755 **JD**, **Focht B**, **Friedman M**, **Freiwald J**, **Heady KK**, **Heady WN**, **Hettinger A**,
756 **Johnson A**, **Karr KA**, **Mahoney B**, **Moritsch MM**, **Osterback A-MK**, **Reimer J**,
757 **Robinson J**, **Rohrer T**, **Rose JM**, **Sabal M**, **Segui LM**, **Shen C**, **Sullivan J**,
758 **Zuercher R**, **Raimondi PT**, **Menge BA**, **Grorud-Colvert K**, **Novak M**, **Carr MH**.
759 2017. Long-Term Studies Contribute Disproportionately to Ecology and Policy.
760 Bioscience **67**:271–281.

761 37. **Brown JH**, **Whitham TG**, **Morgan Ernest SK**, **Gehring CA**. 2001. Complex
762 species interactions and the dynamics of ecological systems: Long-term
763 experiments. Science (80-.). American Association for the Advancement of
764 Science.

765 38. **Ducklow HW**, **Doney SC**, **Steinberg DK**. 2009. Contributions of Long-Term
766 Research and Time-Series Observations to Marine Ecology and
767 Biogeochemistry. Ann. Rev. Mar. Sci. **1**:279–302.

768 39. **Franklin JF**. 1989. Importance and Justification of Long-Term Studies in
769 Ecology, p. 3–19. *In* Long-Term Studies in Ecology. Springer New York.

770 40. **Philippot L**, **Andersson SGE**, **Battin TJ**, **Prosser JI**, **Schimel JP**, **Whitman WB**,

771 **Hallin S**. 2010. The ecological coherence of high bacterial taxonomic ranks. Nat.
772 Rev. Microbiol. **8**:523–9.
773 41. **Garcia SL**, **Salka I**, **Grossart HP**, **Warnecke F**. 2013. Depth-discrete profiles of
774 bacterial communities reveal pronounced spatio-temporal dynamics related to
775 lake stratification. Environ. Microbiol. Rep. **5**:549–555.
776 42. **Garcia SL**, **Stevens SLR**, **Crary B**, **Martinez-Garcia M**, **Stepanauskas R**,
777 **Woyke T**, **Tringe SG**, **Andersson SGE**, **Bertilsson S**, **Malmstrom RR**,
778 **McMahon KD**. 2017. Contrasting patterns of genome-level diversity across
779 distinct co-occurring bacterial populations. ISME J.
780 43. **D'Autréaux B**, **Toledano MB**. 2007. ROS as signalling molecules: Mechanisms
781 that generate specificity in ROS homeostasis. Nat. Rev. Mol. Cell Biol. Nature
782 Publishing Group.
783 44. **Koonin E V**, **Wolf YI**. 2008. Genomics of bacteria and archaea: the emerging
784 dynamic view of the prokaryotic world. Nucleic Acids Res.2008/10/25.
785 **36**:6688–6719.
786 45. **Bienert GP**, **Chaumont F**. 2014. Aquaporin-facilitated transmembrane
787 diffusion of hydrogen peroxide. Biochim. Biophys. Acta - Gen. Subj. **1840**:1596–
788 1604.
789 46. **Zaremba-Niedzwiedzka K**, **Viklund J**, **Zhao W**, **Ast J**, **Sczyrba A**, **Woyke T**,

790 **McMahon K**, **Bertilsson S**, **Stepanauskas R**, **Andersson SGE**. 2013. Single-cell
791 genomics reveal low recombination frequencies in freshwater bacteria of the
792 SAR11 clade. Genome Biol. **14**:R130.
793 47. **Tsementzi D**, **Rodriguez-R LM**, **Ruiz-Perez CA**, **Meziti A**, **Hatt JK**,
794 **Konstantinidis KT**. 2019. Ecogenomic characterization of widespread, closely-
795 related SAR11 clades of the freshwater genus "Candidatus Fonsibacter" and
796 proposal of Ca. Fonsibacter lacus sp. nov. Syst. Appl. Microbiol. **42**:495–505.
797 48. **Moore ER**, **Davie-Martin CL**, **Giovannoni SJ**, **Halsey KH**. 2019. Pelagibacter
798 metabolism of diatom-derived volatile organic compounds imposes an
799 energetic tax on photosynthetic carbon fixation. Environ. Microbiol. **00**:4–6.
800 49. **Herlemann DP**, **Labrenz M**, **Jürgens K**, **Bertilsson S**, **Waniek JJ**, **Andersson**
801 **AF**. 2011. Transitions in bacterial communities along the 2000 km salinity
802 gradient of the Baltic Sea. ISME J. **5**:1571–9.
803 50. **Caporaso JG**, **Kuczynski J**, **Stombaugh J**, **Bittinger K**, **Bushman FD**, **Costello**
804 **EK**, **Fierer N**, **Peña AG**, **Goodrich JK**, **Gordon JI**, **Huttley GA**, **Kelley ST**,
805 **Knights D**, **Koenig JE**, **Ley RE**, **Lozupone CA**, **McDonald D**, **Muegge BD**,
806 **Pirrung M**, **Reeder J**, **Sevinsky JR**, **Turnbaugh PJ**, **Walters WA**, **Widmann J**,
807 **Yatsunenko T**, **Zaneveld J**, **Knight R**. 2010. QIIME allows analysis of high-
808 throughput community sequencing data. Nat. Methods **7**:335–336.

809 51. **Edgar RC**. 2010. Search and clustering orders of magnitude faster than BLAST.
810 Bioinformatics **26**:2460–2461.
811 52. **McDonald D**, **Price MN**, **Goodrich J**, **Nawrocki EP**, **DeSantis TZ**, **Probst A**,
812 **Andersen GL**, **Knight R**, **Hugenholtz P**. 2012. An improved Greengenes
813 taxonomy with explicit ranks for ecological and evolutionary analyses of
814 bacteria and archaea. ISME J. **6**:610–8.
815 53. **Yilmaz P**, **Parfrey LW**, **Yarza P**, **Gerken J**, **Pruesse E**, **Quast C**, **Schweer T**,
816 **Peplies J**, **Ludwig W**, **Glöckner FO**. 2014. The SILVA and "all-species Living
817 Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res. **42**:643–648.
818 54. **R-Core-team**. 2011. R: A Language and Environment for Statistical Computing.
819 2.14.1. R Foundation for Statistical Computing, Vienna, Austria.
820 55. **RStudio**. 2012. RStudio: Integrated development environment for R. 0.97.248.
821 Boston, MA.
822 56. **Kampstra P**. 2008. Beanplot: A Boxplot Alternative for Visual Comparison of
823 Distributions. J. Stat. Softw. **28**:1–9.
824 57. **Giraudoux P**. 2012. pgirmess: Data analysis in ecology. R package version 1.5.6.
825 58. **Weiss S**, **Van Treuren W**, **Lozupone C**, **Faust K**, **Friedman J**, **Deng Y**, **Xia LC**,
826 **Xu ZZ**, **Ursell L**, **Alm EJ**, **Birmingham A**, **Cram JA**, **Fuhrman JA**, **Raes J**, **Sun F**,
827 **Zhou J**, **Knight R**. 2016. Correlation detection strategies in microbial data sets

828 vary widely in sensitivity and precision. ISME J. **10**:1669–1681.

829 59. **Oksanen J**, **Blanchet FG**, **Kindt R**, **Legendre P**, **Minchin PR**, **O'Hara B**,
830 **Simpson GL**, **Solymos P**, **Stevens MHH**, **Wagner H**. 2013. vegan: Community
831 ecology package. R package version 2.0-7.

832 60. **Friedman J**, **Alm EJ**. 2012. Inferring correlation networks from genomic survey
833 data. PLoS Comput. Biol. **8**:e1002687.

834 61. **Kurtz ZD**, **Müller CL**, **Miraldi ER**, **Littman DR**, **Blaser MJ**, **Bonneau RA**. 2015.
835 Sparse and Compositionally Robust Inference of Microbial Ecological Networks.
836 PLoS Comput. Biol. **11**:e1004226.

837 62. **Sørensen T**. 1948. A method of establishing groups of equal amplitude in plant
838 sociology based on similarity of species and its application to analyses of the
839 vegetation on Danish commons. Biol. Skr. **5**:1–34.

840 63. **Dice LR**. 1945. Measures of the Amount of Ecologic Association Between
841 Species. Ecology **26**:297–302.

842 64. 2014. SeqPrep. 1.2.

843 65. **Bolger AM**, **Lohse M**, **Usadel B**. 2014. Trimmomatic: A flexible trimmer for
844 Illumina sequence data. Bioinformatics **30**:2114–2120.

845 66. **Li D**, **Liu CM**, **Luo R**, **Sadakane K**, **Lam TW**. 2014. MEGAHIT: An ultra-fast
846 single-node solution for large and complex metagenomics assembly via

847 succinct de Bruijn graph. Bioinformatics **31**:1674–1676.

848 67. **Schmieder R**, **Edwards R**. 2011. Quality control and preprocessing of
849 metagenomic datasets. Bioinformatics **27**:863–864.

850 68. **Markowitz VM**, **Chen I-M a**, **Palaniappan K**, **Chu K**, **Szeto E**, **Grechkin Y**,
851 **Ratner A**, **Jacob B**, **Huang J**, **Williams P**, **Huntemann M**, **Anderson I**,
852 **Mavromatis K**, **Ivanova NN**, **Kyrpides NC**. 2012. IMG: the Integrated Microbial
853 Genomes database and comparative analysis system. Nucleic Acids Res.
854 **40**:D115-22.

855 69. **Sayers EW**, **Barrett T**, **Benson D a**, **Bolton E**, **Bryant SH**, **Canese K**,
856 **Chetvernin V**, **Church DM**, **DiCuccio M**, **Federhen S**, **Feolo M**, **Fingerman IM**,
857 **Geer LY**, **Helmberg W**, **Kapustin Y**, **Landsman D**, **Lipman DJ**, **Lu Z**, **Madden**
858 **TL**, **Madej T**, **Maglott DR**, **Marchler-Bauer A**, **Miller V**, **Mizrachi I**, **Ostell J**,
859 **Panchenko A**, **Phan L**, **Pruitt KD**, **Schuler GD**, **Sequeira E**, **Sherry ST**,
860 **Shumway M**, **Sirotkin K**, **Slotta D**, **Souvorov A**, **Starchenko G**, **Tatusova T a**,
861 **Wagner L**, **Wang Y**, **Wilbur WJ**, **Yaschenko E**, **Ye J**. 2011. Database resources
862 of the National Center for Biotechnology Information. Nucleic Acids Res.
863 **39**:D38-51.

864 70. **Berman HM**, **Westbrook J**, **Feng Z**, **Gilliland G**, **Bhat TN**, **Weissig H**,
865 **Shindyalov IN**, **Bourne PE**. 2000. The Protein Data BankNucleic Acids

866 Research.

867 71. **Kelley LA**, **Mezulis S**, **Yates CM**, **Wass MN**, **Sternberg MJE**. 2015. The Phyre2
868 web portal for protein modeling, prediction and analysis. Nat. Protoc. **10**:845–
869 858.

870 72. **Camacho C**, **Coulouris G**, **Avagyan V**, **Ma N**, **Papadopoulos J**, **Bealer K**,
871 **Madden TL**. 2009. BLAST+: architecture and applications. BMC Bioinformatics
872 **10**:421.

873 73. **Katoh K**, **Kuma KI**, **Toh H**, **Miyata T**. 2005. MAFFT version 5: Improvement in
874 accuracy of multiple sequence alignment. Nucleic Acids Res. **33**:511–518.

875 74. **Niu B**, **Fu L**, **Sun S**, **Li W**. 2010. Artificial and natural duplicates in
876 pyrosequencing reads of metagenomic data. BMC Bioinformatics **11**:187.

877 75. **Suzek BE**, **Huang H**, **McGarvey P**, **Mazumder R**, **Wu CH**. 2007. UniRef:
878 comprehensive and non-redundant UniProt reference clusters. Bioinformatics
879 **23**:1282–8.

880 76. **Magrane M**, **Consortium U**. 2011. UniProt Knowledgebase: a hub of integrated
881 protein data. Database (Oxford). **2011**:bar009.

882 77. **Kalyaanamoorthy S**, **Minh BQ**, **Wong TKF**, **Von Haeseler A**, **Jermiin LS**. 2017.
883 ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat.
884 Methods **14**:587–589.

885 78. **Kozlov AM**, **Darriba D**, **Flouri T**, **Morel B**, **Stamatakis A**. 2019. RAxML-NG: a
886 fast, scalable and user-friendly tool for maximum likelihood phylogenetic
887 inference. Bioinformatics.
888
889