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 of them 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 and the cooperation may be reciprocal, expensive and non-reciprocal (Black Queen hypothesis), or costless (byproduct benefits). 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 *Planktophila* and "*Ca.* Fonsibacter" show no correlation to each other 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 *Planktophila* is dependent upon) and its most likely providers in the lake. Further we found that while ecotype and community cooperation combined may explain *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 **IMPORTANCE**

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

49

50 **INTRODUCTION**

51 The stable coexistence of species is described by a continuum of inter-specific symbiotic
52 interactions from parasitism through neutralism to mutualism (1). The extended
53 coexistence of microbial strains is described by intra-specific relationship models limited
54 to niche theory (ecotype differentiation) (2) and predator control of dominant strains (3).
55 Combining these ecological theories of co-existence with phylogenetic information can
56 produce a simplified model with a new holistic perspective for understanding aquatic
57 microbial communities. In aquatic microbial communities, especially for non-particle-
58 associated bacterioplankton, physical structures and proximity as seen in soil, host-
59 associated microbiome, and biofilm communities (4) are absent. As cooperation models

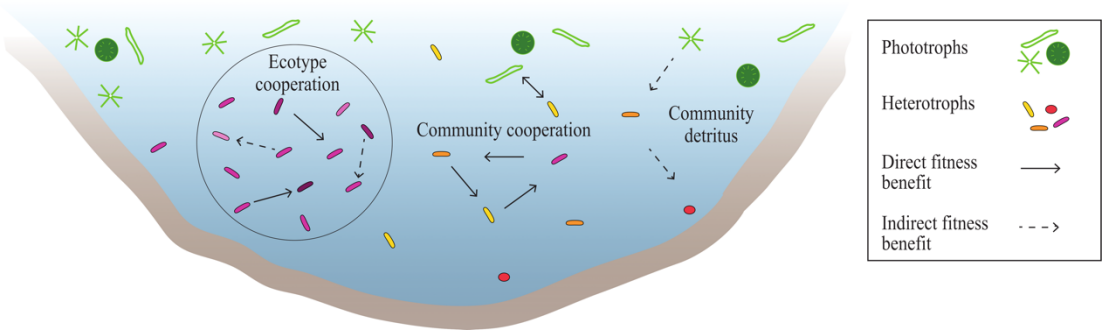
60 have been developed from the study of such tightly interwoven assemblages their models
61 may not be the most suitable. Absent also is the habitat isolation which limits dispersal as
62 seen in e.g. soil ecosystems. In aquatic systems interactions between nutrients and biota
63 can occur from micro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

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

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

escape from viral predation (23, 32). A large N_e where the divergence is focused in immunogenic or predator-escape mechanisms (defense specialist) will help a species survive population sweeps caused by predators, though it does not explain the intra nor inter-specific diversity or the seeming lack of competition between these three genera.

We combined the continuum of symbiotic interactions with phylogenetic information to produce a simplified model with a new holistic perspective for understanding aquatic microbial communities. (Fig. 1). Ecotype cooperation model describes how anabolic variability within a species population i.e. strain level variation, enables all the strains working together to produce nutrients required for cellular function and replication. It also allows for recombination and swapping of metabolic modules across strains. The community cooperation model also has a basis in anabolic complementarity but places the metabolic variation within a symbiotic group of different species whereby altruistic members provide expensive (Black Queen Hypothesis) or costless common goods (33, 34). The community detrital model is similar to community cooperation but focuses on uptake and catabolism of nutrients released by cellular death mediated by toxicity, starvation, or predation. Aquatic habitats in particular, due to transport and diffusion, are understood to have a unique niche for detritivores that may select for streamlined microbes (35).



Here we examine an 8-year time-series along with 26 minimalist mixed-cultures and selected lake metagenomes, to identify putative supporting microbes (cohort) and whether that support is best described by ecotype, community, and/or detrital models. We began with OTU network modelling (to define a candidate cohort that might provide for the target three genera dependencies), examined the phylogenetic relatedness of each cohort (to determine if a signal for ecotype or community existed), reviewed the metabolic diversity of the target group (to establish its dependencies and diversity), and investigated potential providers of the common-goods catalase-peroxidase (KatG) (a known dependency of some *Nanopelagicales*) within the metagenomes. Our proposed framework focusses on neutral to positive interactions as the laboratory experimental design does not permit detection of negative interactions even though timeseries can expose potential negative interactions.

RESULTS

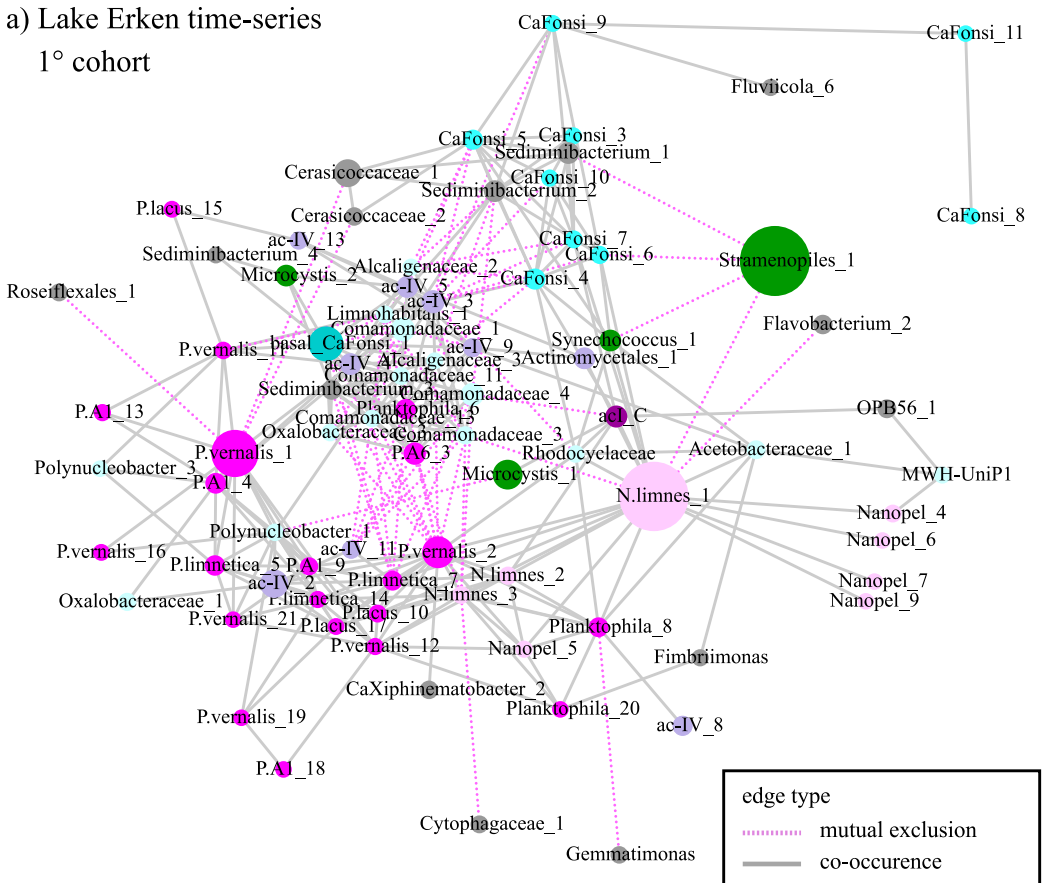
Taxon abundance and environmental correlates

151 The most abundant clades detected in the SSU rRNA gene 8-year time-series were the
152 *Actinobacterial* order *Nanopelagicales* at 29% average relative abundance (av.ra), the
153 eukaryote phytoplankton genus *Stramenopiles* 12% av.r.a, the *Actinobacterial* clade 'acIV'
154 (C111) 11% av.ra, the *β-proteobacterial* family *Comamonadaceae* 5% av.ra, the *α-*
155 *Proteobacterial* candidate genus "*Ca.* Fonsibacter" 4% av.ra, and the *Verrucomicrobial*
156 family *Cerasicoccaceae* 3% av.ra (Fig. S1, Table S1). There were few significant
157 correlations or differences (Pearsons correlation or Kruskal-Wallis (KW) with post-hoc
158 testing to identify which pairs were significantly different (KWmc)) between dominant
159 clade abundance and a broad range of environmental predictors such as lake-cycle/season
160 or physico-chemical conditions of the water (Fig. S2, Table S2). The three significant linear
161 correlations detected had only moderate strength (r >= 0.4) with cloud or wedge-shaped
162 distributions (Fig. S3).
163
164 **Microbial networks**
165 The OTU network from both the time-series and mixed-cultures (Fig. S4) shows
166 modularity and taxonomic assortativity (grouping of related organisms or under-
167 dispersion) at all taxonomic levels with greater assortativity in the mixed-cultures (Table
168 S3). Assortativity increased with taxonomic level peaking at phylum level in the timeseries,
169 while assortativity was lowest at phylum level in the cultures. First neighbours (the primary
170 cohort) of OTUs in networks provide evidence (but not proof) for preferential association

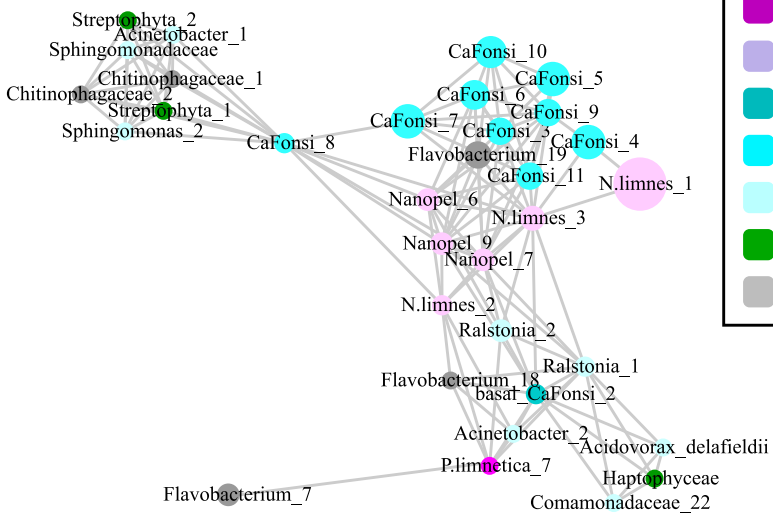
171 between phylotypes, or when negatively correlated may indicate mutual exclusion. The
172 primary cohorts of "*Ca.* Planktophila, "*Ca*. Nanopelagicus", and "*Ca*. Fonsibacter" (Fig.
173 2) are overlapping and include phylotypes from each other's cohorts. Primary cohorts from
174 the time-series and mixed-cultures are similar, though the correlation type and strength
175 vary. There was a moderate correlation between relative abundance of OTUs in the
176 timeseries (Fig. 2a) and relative abundance in cultures (Pearsons $r = 0.58$, $p_{adj} = 0$, Table
177 S4) but no other correlations between relative abundance or prevalence (percentage of
178 samples an OTU was detected in) were observed.
179
180 **Actinobacterial order Nanopelagicales (acI)**
181 *Nanopelagicales* in Erken had higher relative abundances in the hypolimnion, (KW
182 $p<0.001$, KWmc $p<0.001$) and were negatively correlated with % O_2 (Pearsons $r = -$
183 0.4, $p<0.001$). *Nanopelagicales* were the most abundant clade in Lake Erken and were
184 represented by 31 OTUs in the lake time-series network with nine "*Ca*.
185 Nanopelagicus", 21 *"Ca.* Planktophila*"*, and one acI-c (Fig. 2a, Fig. S4). A subset of six
186 "*Ca*. Nanopelagicus" and one *"Ca.* Planktophila*"* phylotype were detected in the mixed-
187 culture based network (Fig. 2b, Fig. S4).
188 The time-series primary (1°) cohort (i.e. the cohort immediately surrounding) of *"Ca.*
189 Planktophila*"* included: other *Actinobacteria*, *Armatimonadetes*, *Chloroflexi*,

190 *Gemmatimonadetes*, *α*- and *β-Proteobacteria*, and *Verrucomicrobia*, while in the
191 mixed-cultures only *Actinobacteria*, *Bacteroidetes*, and *α*- and *β-Proteobacteria* co-
192 existed. A pairwise comparison of the phylogenetic distance and correlation
193 (averaged edge value from the correlation network) was conducted to see if there was
194 a relationship between OTU genetic relatedness and co-occurrence; to distinguish
195 between ecotype and directed reciprocation models. Within the *"Ca.* Planktophila*"* 1°
196 cohort, there was no support for a correlation between phylogenetic distance and co-
197 occurrence (Fig. 3) and *"Ca.* Planktophila*"* was the only genera tested that had
198 negative intra-genus correlations. *"Ca.* Planktophila*"* had more intra-phyla (i.e. non-
199 *Nanopegicales Actinobacteria*) correlations than the other genera tested. Unlike the
200 other genera, *"Ca.* Planktophila*"* had no correlations (positive or negative) with
201 phytoplankton in its 1° network (Fig. 3). Additionally, *"Ca.* Planktophila*"* were the
202 dominant Actinobacterial genus in the time-series with "*Ca*. P. vernalis*"* being most
203 abundant (Fig. 2a). Conversely, in the mixed-cultures, only one *Planktophila* phylotype
204 was detected, "*Ca*. P. limnetica*"* (Fig. 2b).
205 *"Ca.* Planktophila spp." were successfully maintained in axenic culture via the addition
206 of bovine catalase and therefore the presence, absence, and functionality of the
207 relevant *Nanopelagicales* catalase-peroxidase (KatG) is debated. We investigated the
208 presence of the *katG* gene, its gene neighbourhood, and the homology and topology of

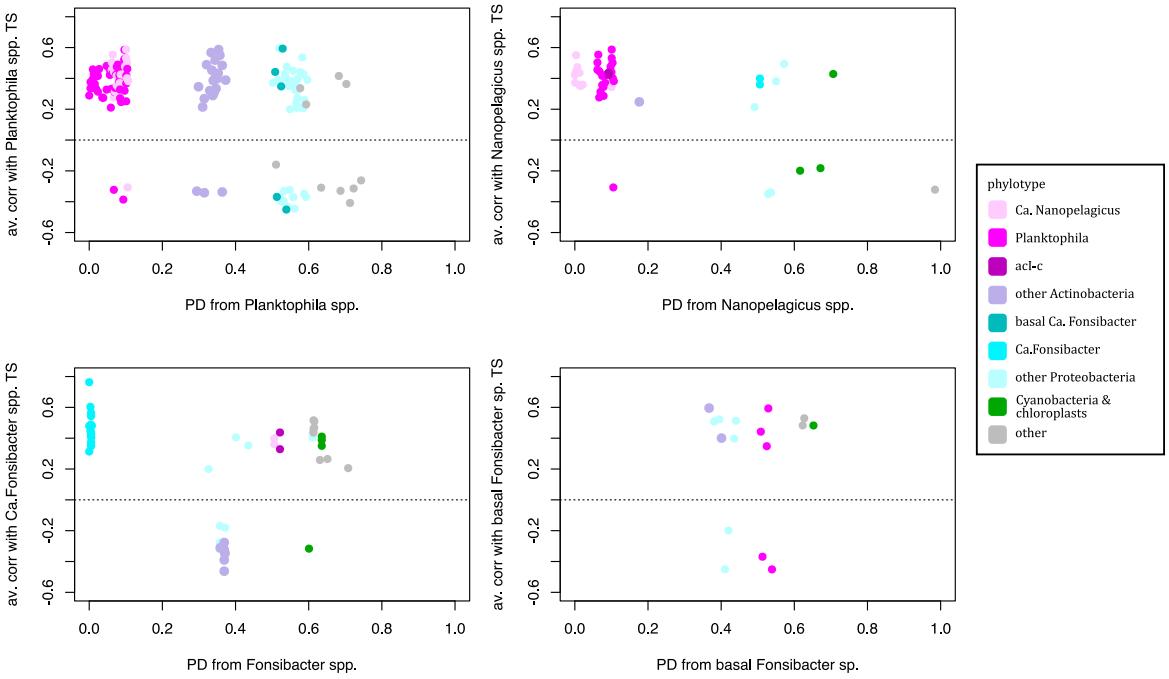
a) Lake Erken time-series
1° cohort



b) Mixed-cultures from lake water
1° cohort



210 its translated product, KatG, to establish if there is evidence to support its activity level or
211 effect on host abundance. "*Ca*. P. vernalis" (acI-A7) KatG had highest homology to the high
212 functioning *Synechoccocus* version (Fig. S5) and the gene is found near a tRNA (Fig. S6).
213 "*Ca*. P. limnetica"s KatG had highest identity with the assayed low activity KatG from "*Ca*.
214 P. rubra" (IMCC 25003) and was located within a putative DNA repair region (Fig. S5 &
215 S6). When modelled these KatGs were most similar to *Burkholderia pseudomaleii*s with a
216 predicted β-sheet around one of the active site residues instead of an α-helix, and are 10 to
217 20 amino-acid residues longer than known high functioning versions i.e. *Burkholderia* type
218 A (Fig. S5). All known active site residues involved in heme interaction or enzyme
219 switching were however present. "*Ca*. Planktophila limnetica", "*Ca*. P. vernalis", "*Ca*. P.
220 lacus", *and* "*Ca*. N. limnes" all have the repair cassette without *katG*. Metagenomic *katG*
221 sequences were recovered matching "*Ca*. P. dulcis", "*Ca*. P. limnetica", "*Ca*. P.
222 versatilis", and "*Ca*. P. vernalis".
223 "*Ca*. Nanopelagicus" 1° cohort in the time-series included other *Actinobacteria*,
224 *Bacteroidetes*, α- and β-*Proteobacteria*, and the phytoplanktonic *Synechococcus* and
225 *Stramenopiles* (identified by chloroplast sequences), while in the mixed-cultures
226 there was a subset of these phyla with the addition of one γ-*proteobacterial* phylotype
227 (Fig 2).
228



To calculate whether the presence of an OTU in the mixed-cultures could be attributed to the probability of it being in the inoculum, the total and average relative abundances, and the number of samples an OTU was detected in (prevalence) was compared across the two both datasets with Pearsons correlation. There was no correlation between "*Ca*. Nanopelagicus" OTU relative abundance or prevalence in the lake time-series as compared to the mixed-cultures (Table S4). "*Ca*. Nanopelagicus" was overall more prevalent in mixed cultures than in the lake time-series. "*Ca*. Nanopelagicus" had strong correlation with "Ca. Fonsibacter" in both datasets. There

238 were between 3 and 19 other phylotypes (non-*Nanopelgicales,* non-Fonsibacter)
239 detected in cultures with "*Ca*. Nanopelagicus" (Table S5).
240 The most abundant "*Ca*. Nanopelagicus" phylotypes in both datasets were assigned to the
241 *katG*$^{-}$ "*Ca*. Nanopelagicus limnes". Several metagenome-recovered putative "*Ca*.
242 Nanopelagicus" *katG* genes were identified in IMG databases and while these KatGs had
243 greatest homology with *Burkholderia* fold type B, they were not located within the repair
244 cassette and there was no synteny between them. No "*Ca*. Nanopelagicus" katG sequences
245 were recovered from our metagenomes.
246
247 **α-Proteobacterial genus "Ca. Fonsibacter" (LD12)**
248 "*Ca*. Fonsibacter spp." had higher relative abundance from winter to spring and moderate
249 negative correlation with water temperature ($r = -0.4$, $p<0.001$, Table S2). "*Ca*.
250 Fonsibacter" was represented by 10 OTUs in each of the time-series and mixed-culture
251 networks, with nine OTUs common to both and the alternating OTUs being the least
252 abundant and least prevalent in both environments. The "*Ca*. Fonsibacter" 1° cohort from
253 the time-series includes other "*Ca*. Fonsibacter", *Bacteroidetes*, *α*- and *β-Proteobacteria*,
254 *Actinobacteria* including "*Ca*. Nanopelagicus" and acI-c, and phytoplankton
255 *Synechococcus* and *Stramenopiles*. The "*Ca*. Fonsibacter" 1° cohort from mixed-culture
256 experiments also included other "*Ca*. Fonsibacter" phylotypes, "*Ca*. Nanopelagicus",

257 *Bacteroidetes*, other *α*- and *β-Proteobacteria* and macro-algal eukaryotes of the
258 *Streptophyta* order. There was a strong correlation detected between "*Ca*. Fonsibacter"
259 OTU prevalence in the lake time-series compared to the prevalence of the respective OTU
260 in the mixed-cultures (corr = 0.82, n=11, p adj = 0, Table S4). Based on OTU identity,
261 phylogenetic branch placement, and network connections, two "*Ca*. Fonsibacter"
262 phylotypes were designated as basal and analysed separately for 1° cohort. Unlike the core
263 "*Ca*. Fonsibacter" phylotypes, these basal phylotypes had no negative correlations with
264 phytoplankton and no correlations with other "*Ca*. Fonsibacter" phylotypes (Fig. 3).
265
266 **Other primary cohort members**
267 Both the *Actinobacterial* 'acIV' (C111) and *Comamonadaceae* families were more
268 abundant in winter when the lake was covered by ice (Fig. S2, KW $p<0.001$, KWmc
269 $p<0.001$). While 'acIV' was abundant and ubiquitous in the time-series (Fig. S1, Table S1),
270 and also prominent in the *"Ca*. Planktophila" time-series 1° cohort (Fig. 2a), no phylotypes
271 were detected in the mixed-cultures. *Comamonadaceae* was abundant but patchy in the
272 time-series and mixed cultures.
273
274 **DISCUSSION**
275 **Taxon relative abundance and environmental correlates**

276 If microbial abundance is directly linked to nutrient availability or other physico-chemical
277 parameters, then a strong correlation between geo-chemical parameters and relative
278 abundance is expected. The lack of strong correlation between OTUs or clades and
279 environmental parameters in Lake Erken indicates that during the 8 years of sample
280 collection, microbial community composition was not directly and strongly controlled, but
281 only influenced, by the measured abiotic environmental variables. While short-term studies
282 have shown significant strong trends and correlations (12), such correlations tend to be
283 eclipsed in studies spanning multiple years by shifts operating at longer frequencies and
284 biotic interactions (36–39). Overall, the strength and number (identified via consensus
285 network) of correlations between taxon relative abundance and the weakness of
286 environmental predictors for most of the taxa suggest that biological aspects (including
287 predation, which we did not study) are the greater determinants of microbial population
288 dynamics, composition and persistence in Lake Erken. It stands therefore that the
289 dominance of the three streamlined genera in the lake is attributable more to their
290 interactions with their community than selection by environment.
291
292 **Microbial networks**
293 Assortativity is a predicted product of ecotype or niche differentiation at species level and
294 environmental filtering at higher taxonomic levels (40). Ecotype divergence of strains
295 reduces competition between different members of a population. The detection of highest

assortativity in the time-series networks at phylum level supports environmental selection as important in the lake habitat but not at taxonomic levels lower than phyla. It is unlikely that environmental filtering is the main driver behind the success of these three genera in Lake Erken. High assortativity in the mixed cultures at all levels except phylum, supports that niche differentiation contributed to co-occurrence in the cultures. Neither result however fully supports, nor refutes, ecotype theory as the explanation of coexistence or dominance by "*Ca.* Planktophila", "*Ca.* Nanopelagicus", or "*Ca.* Fonsibacter" genera.

The strong correlation between "*Ca*. Fonsibacter" (and also *Comamonadaceae*) prevalence in the time-series and mixed-cultures supports that its occurrence in the cultures may be attributed to being present in the inocula and that the filtered lake water media was able to meet an important proportion of its metabolic needs. Conversely, the lack of correlation in either relative abundance or prevalence, for the *Nanopelagicales* genera (and also the 'acIV' family) in the time-series compared to mixed-cultures means that there is not support for their detection being merely due to high-abundance in the inocula. We infer that the co-presence of other phylotypes were bigger contributors to their growth. It should be noted that negative correlations are not detectable in the mixed-cultures as the absence of a phylotype could be stochastic rather than a result of competitive exclusion. Moreover, competition that prevents growth in a culture cannot be identified.

316 **Actinobacterial order Nanopelagicales 'acI'**

317 *Nanopelagicales* in Erken were likely more competitive in colder, darker, and oxygen
318 depleted waters. While *Nanopelagicales* has been associated with oxygen saturated waters
319 in some locations (41) this lineage has been associated with colder deeper water at others
320 (30), though the variation in association could be due to the considerable genetic diversity
321 encompassed in this bacterial order (42). It is also probable that solar-driven formation of
322 oxygen radicals in surface waters inhibit the growth of *katG*$^{-}$ species of *Nanopelagicales*.
323 Of the seven *Nanopelagicales* phylotypes in the mixed-cultures, six were members of the
324 candidate genus "*Ca*. Nanopelagicus" suggesting that this genuss dependencies were more
325 easily met compared to the *"Ca.* Planktophila spp.*"* found in Lake Erken. However,
326 genomic evidence so far has shown the opposite as "Ca. Nanopelagicus spp." have more
327 numerous auxotrophies than *"Ca.* Planktophila spp." (23). It is therefore possible that the
328 difference in culturability seen in this experiment was due to reduced costs via gene loss in
329 combination with suitable metabolic partners being in the inoculum. In particular the lower
330 maintenance costs due to the loss of *katG* (25, 33) could increase fitness of the "*Ca*.
331 Nanopelagicus". The successful maintenance of *"Ca.* Planktophila*"* in culture pinpointed
332 the low functionality of some *Nanopelagicales* catalase-peroxidases as the culprit for the
333 difficulties in growing these bacteria. Peroxide is proposed to be an up-regulator of a DNA
334 repair (43) and in the *"Ca.* Planktophila*"* with the putative low functioning KatG type, the
335 gene is embedded in a region of DNA hosting other genes for repair of damage done to

336 DNA under oxidative stress. *P. limnetica* (acI-A3) genomes have this repair cassette but
337 their *katG* is located elsewhere. The most parsimonious explanation for *Nanopelagicales*
338 *katG* presence/absence/homology is that the now low functioning version located in a
339 repair cassette is in the process of being lost and that the *"Ca.* Planktophila spp.*"* with a
340 high functioning KatG obtained this version via a HGT event. While *P. vernalis* in Lake
341 Erken is likely *katG*$^{+}$, it is likely that *P. limnetica* and *N. limnes* from Lake Erken rely on
342 "*Ca*. Fonsibacter" in combination with either *Flavobacterium* or *Comamonadaceae* to
343 detoxify peroxide. This would fit within the community cooperation model. While *"Ca.*
344 Nanopelagicus" has more numerous auxotrophies compared to *"Ca.* Planktophila*"* (23)
345 the higher culturability of *"Ca*. Nanopelagicus" supports that it may be the cost to
346 habouring *katG*, whether high or low functioning, and as *katG* is larger than average (44),
347 and as long as a functional aquaporin exists (45), and cells don't grow too fast, then it
348 would be an unnecessary cost, as cells can rely on other community members for oxidative
349 stress reduction. Further the reduction in amino-acid synthesis pathways and therefore
350 greater reliance on exogenous sources of amino-acids in "*Ca*. Nanopelagicus" appears to
351 have paid off if their greater cultivability in these minimalist cultures is any indicator. This
352 fits within the community detritus model.
353
354 **α-Proteobacterial genus "Ca. Fonsibacter" (LD12)**

355 Earlier short-term studies found that "*Ca*. Fonsibacter" is highly abundant in Lake Erken
356 in the late summer-early autumn immediately after an algal bloom i.e. the 'clear water'
357 phase of the lake's cycle (12). However, we found support for higher long-term relative
358 abundances in winter and spring and an associated negative correlation with water
359 temperature. These differences in findings could be due to changes in abundance between
360 years of different ecotypes (with different temperature and nutrient preferences) that were
361 not detected via 16S surveys. The ecotype differentiation within "*Ca.* Fonsibacter" is
362 centered around the HVR1 which is thought to encode genes for defense, i.e. predator
363 escape via alteration of cellular membranes (3, 46). There appears to be limited intra-
364 population variability in anabolic or catabolic traits (14, 24, 47) and as such, ecotype
365 differentiation for overcoming auxotrophies is not supported. While no direct evidence
366 exists for "*Ca*. Fonsibacter", its similarity to the marine sister clade Pelagibacter suggests
367 that it has overcome auxotrophic limitations by scavenging metabolites and other
368 compounds produced by phototrophs (48). In Erken this would primarily be dying and
369 senescent *Stramenopiles* as reflected in the apparent negative correlation. Evidence points
370 to "*Ca*. Fonsibacter" relying on detritus to meet its dependencies and ecotype
371 differentiation for defense.
372 The combination of network and other correlation analyses with review of published
373 genomes supports that the most of *Nanopelagicales* have anabolic ecotype cooperation
374 with reliance on a detoxifying community cohort as main structuring factors. A few *"Ca.*

375 Planktophila*"* are putatively KatG+ but as they were not detected in Lake Erken it was not
376 possible to determine the most appropriate cooperation models in this study. "*Ca*.
377 Fonsibacter" may escape predation via ecotype differentiation, while broadly relying on
378 community detritus for anabolic deficits.
379
380 **MATERIALS AND METHODS**
381 **Time series water sampling and environmental parameters**
382 Time series water samples were obtained from dimictic Lake Erken (59.1510N, 18.1360E)
383 central Sweden at monthly to weekly intervals spanning a total period from February 2007
384 to November 2015 (Table S6). Samples for genomic and chemical analysis were taken from
385 the 20 m water column at one meter intervals from the deepest point of the lake (Fig. S7).
386 Samples for the total water column were pooled during periods of mixing. Samples of the
387 upper oxygenated water column (epilimnion), middle (metalimnion), and lower oxygen
388 depleted (hypolimnion) were pooled into separate composite samples during summer
389 stratification. Annual dates of mixing, stratification, ice-over, and ice-thaw varied, as did
390 depth of stratigraphic layers (Table S6). Temperature and oxygen concentrations were
391 analyzed every meter with a portable Oxi 340i oxygen meter equipped with a Cellox 325-
392 20WTW probe and were used to determine stratification. Chemical analysis of pooled
393 water samples was carried out by the Lake Erken Field Station ISO certified laboratory and
394 included pH, conductivity, nitrate, nitrite, alkalinity, turbidity, suspended matter,

395 phosphate, ammonium, total and particulate phosphorus and nitrogen, chlorophyll-a,
396 suspended particulate organic matter, silicate, water-colour and absorbance at 420nm
397 measurements. Water samples for genomic analysis were collected by gentle vacuum
398 filtration onto 0.2 µm membrane filters (Supor-200 Membrane Disc Filters, 47mm; Pall
399 Corporation, East Hills, NY, USA). Filters were individually stored in Eppendorfs and
400 transported back to the University Uppsala laboratory where they were stored at -80 °C
401 until further processing.
402
403 **DNA extraction and SSU rRNA gene amplicon preparation**
404 DNA was extracted from filters using MoBio Ultra-clean soil DNA extraction kits (MoBio,
405 Carlsbad) as per manufacturers' instructions. The V3-V4 region of the 16S rRNA gene was
406 amplified using the S-D-Bact-0341-b-S-17 (Bakt_341F: CCTACGGGNGGCWGCAG)
407 forward primer and S-D-Bact-0785-b-A-21 (805RN: GACTACNVGGGTATCTAATCC)
408 reverse primer (49). Template DNA was amplified in duplicate 20 µl reactions containing
409 1.0 U Q5 high fidelity DNA polymerase (NEB, UK), 0.25 µM primers, 200 µM dNTP mix,
410 and 0.4 µg BSA. The thermocycler program was an initial denaturation step at 95 °C for
411 30 s, followed by 20 cycles of dissociation at 95 °C for 10 s, annealing at 53 °C for 30 s
412 extension at 72 °C for 30 s, with a final extension of 2 min at 72 °C. Amplicons were pooled
413 and purified with Agencourt AMPure XP purification system (Beckman Coulter, Danvers,
414 MA, USA). Purified amplicons (2 µl) were amplified in second step 20 µl reactions to

415 introduce sample MIDS to each end of the amplicons. Reactions contained 1.0 U Q5 high
416 fidelity DNA polymerase (NEB, UK), 0.25 µM primers, 200 µM dNTP mix, and 0.4 µg
417 BSA. Thermocycler program was an initial denaturation step at 95 °C for 30 s, followed
418 by 15 cycles of dissociation at 95 °C for 10 s, annealing at 66 °C for 30 s, extension at 72
419 °C for 30 s, with a final extension of 2 min at 72 °C. Amplicons were again purified using
420 Agencourt AMPure XP purification system (Beckman Coulter, Danvers, MA, USA) then
421 quantified using PicoGreen (Invitrogen) before pooling at equi-molar amounts for each
422 run. Amplicons were sequenced at SciLifeLab SNP/SEQ service on the MiSeq Illumina
423 platform with 300bp paired-end libraries. MiSeq data was processed, including de-
424 multiplexing, by the sequencing centre using Illumina pipelines where all reads with more
425 than 8% mismatch to adapter-MID sequences were discarded.
426
427 **Water sampling, culture conditions, and DNA amplification for mixed-cultures**
428 Water as both a source of growth-media and culturable-cells was collected from the
429 epilimnion of Lake Erken on March 8th 2016. Water for growth-media was filter-sterilised
430 twice through Sterivex filter cartridges 0.22 µm (Millipore). Ten minutes exposure to UV
431 light further disinfected and disrupted cellular and viral nucleic acids. Ten microliters of
432 untreated lake water (10^6 cells/ml) were used as a source of cells for inoculation of mixed-
433 cultures. The untreated lake water was diluted in 1 L of the triple sterilized lake water
434 (approximately 10 cells/ml). Ninety-eight 1 ml cultures were placed in sterile 50 ml falcon

435 tubes to allow oxygenation and incubated for two months under 12-hour light-dark cycles
436 at 11 °C to replicate lake conditions at time of collection. The V3-V4 region of the 16S
437 rRNA gene was amplified using the S-D-Bact-0341-b-S-17 (Bakt_341F:
438 CCTACGGGNGGCWGCAG) forward primer and S-D-Bact-0785-b-A-21 (805RN:
439 GACTACNVGGGTATCTAATCC) reverse primer (49) One microliter of the mixed-
440 culture after one freeze-thaw cycle was used as template and Q5® High-Fidelity DNA
441 Polymerase. The PCR conditions consisted of an initial step at 98°C for 10 minutes.
442 Thereafter, 35 cycles of denaturation at 98°C for 10 s, annealing for 30 s and extension at
443 72°C for 30 s. Annealing temperature was 48°C which has been tested to get unbiased
444 products of non-mismatch and 3-mismatch isolates. Final extension was performed at 72°C
445 for 2 minutes. PCR products were purified with magnetic beads (Angecourt AMPure). A
446 second PCR was conducted for attaching standard illumina handles and index primers as
447 described above. Primers and MIDS detailed in Table S7. Prior to sequencing a portion of
448 amplification product was run on gel to check for amplification and used as proxy for
449 growth of culture. Only data from the 60 cultures with a visible band after the first PCR
450 were included in analyses.
451
452 **Amplicon bioinformatics**
453 Demultiplexed MiSeq time-series and mixed-culture data, and 454 Ti time-series data
454 obtained from Lake Erken 2008 (11), were pre-processed separately and joined once

455 chimeras had been removed. Pre-processing of MiSeq data started with SeqPrep v1.2
456 (https://github.com/jstjohn/SeqPrep.git) with default settings to join paired end
457 reads while discarding reads shorter than 30b and all un-paired reads. 454 and MiSeq
458 reads with ambiguous bases, low quality average, and low quality sections, were
459 discarded and trimmed with Qiime v1.9 (50) split_libraries and multiple split libraries
460 fastq respectively. Singletons and sequences identified as chimeric were removed
461 with Qiimes' identify_chimeric_seqs, filter_fasta using usearch v6.1 (51), with default
462 settings and option parameter 'intersection' to select out sequences identified by both
463 de-novo and reference-based methods. Both GreenGenes (v105 13-8) and SILVA
464 (v128) reference datasets were used (52, 53). Sequences from the two Miseq runs and
465 the 454 data from 2008 were then combined. Reads were clustered at 97% identity
466 into OTUs then assigned taxonomy based on SILVA v128 using QIIMES closed pick
467 OTU method (uses only template recognition and does not allow for de-novo
468 sequences) to allow for differences in DNA processing and sequencing across the
469 three datasets. Sequences from the negative control were removed from the entire
470 dataset. Samples with potential human-microbiome contamination (Table S8) were
471 identified and discarded from the dataset due to potential effect on community
472 metabolism. Sequences identified as eukaryotic (including fungal, mitochondrial and
473 chloroplast origin) or archaeal were deliberately retained despite the primers not

474 specifically being selected to amplify such sequences. as they were considered critical
475 in the mixed-culture analyses from a community metabolism perspective. Samples
476 were normalized using Qiimes single rarefraction to even depth of 1500 reads per
477 sample for the time-series and 500 reads for the mixed-culture samples. These
478 normalised OTU tables were then used for all downstream analyses.
479
480 **Statistical analyses of time-series environmental correlates and selected taxa**
481 Statistical analyses were done using R v3.3.1 (54) in the RStudio IDE v0.99.903 (55)
482 and graphs processed for publication in Inkscape v0.91. Environmental parameters
483 were removed if more than 40% of the values were missing then remaining twenty-
484 one tested for co-linearity with the usdm v1.1.18 package with vifstep followed by
485 vifcor functions and removed if VIF >10 or correlations were over 0.7 sequentially
486 starting with those with highest number of missing values. The remaining 14
487 environmental predictors were tested for statistical differences in connection to lake-
488 cycle with Kruskal-Wallis (kruskal.test of the base stats package) and graphed using
489 beanplot v1.2 (56) and scales v0.4.1. Statistical significance of differences between
490 individual paired cycle points were determined post-hoc with the Kruskal-Wallis
491 multiple comparison (KWmc) using thekruskal.mc test of the pgirmess package
492 v1.6.4 (57). Environmental parameter correlates with selected taxa relative

493 abundances were screened using Pearsons correlation with an alpha level of 0.001
494 and p-values estimated and corrected for multiple comparisons with Benjamini,
495 Hochberg, and Yekutieli (BH) false discovery rate using the psyche package v1.6.12
496 corr.test function and Pearsons moment r reported. Parameters with weak to
497 moderate correlation (r >|0.3|) were plotted in R to observe monotonicity.
498
499 **Network analyses**
500 The normalized time-series OTU table was further processed for network and correlation
501 analyses to remove rare OTUs that, while potentially important within the community, are
502 known to inhibit identification of ecological relationships between OTUs and also between
503 OTUs and their environment. Recommendations include OTU tables with 50% or less
504 sparcity and n_{eff} (inverse Simpsons) above an average of 10 and the use of multiple methods
505 (58). A range of filtering parameters based on frequency of OTU detection, from 2 samples
506 up to 46 (20% of samples) and total relative abundance ranging from 1 up to 150 (0.1%),
507 was used on the OTU table to remove type I errors in detection of co-occurrence networks
508 (58). The best filter combination was found with OTUs at a total relative abundance above
509 0.1% and occurring in at least 29 samples. All OTU tables were analysed in R (54) using
510 the vegan v2.4-1 package (59) diversity function to find the n_{eff} and the base stats v3.3.1
511 package to find min, max and mean of the n_{eff}. The reduced OTU table was analysed with
512 SparCC (60), Pearsons correlation coefficient, and SPIEC-EASI (61). SparCC analyses

513 were carried out in Rstudio (55) using both the rsparcc and spiec.easi packages with rsparcc
514 parameters: max.iter=100, th=0.1, exiter=10 and pseudo p-values generated with
515 spiec.easis sparccboot (100 bootstraps) and pval.sparccboot scripts (two-sided test).
516 Pearson correlations were performed using the corr.test of psyche package using the
517 optional p-values adjustment for multiple testing of Benjamini, Hochberg and Yekutieli
518 false discovery rate. SPIEC-EASI correlations were analysed using the spiec.easi package
519 with parameters: method="mb", sel.criterion="stars", lambda.min.ratio=1e-2,
520 nlambda=20, and 50 repetitions. Singletons were removed from the normalized mixed-
521 culture OTU table prior to network analysis. The reduced OTU table was analysed with
522 Pearsons correlation, SpiecEasi (61) and on the binary version of the table, Dice-Sørensen
523 index (62, 63).
524 Correlations between OTUs in the time-series and mixed cultures were retained if the
525 correlation co-efficient was greater than 0.3 absolute for Pearsons, SparCC, or Dice-
526 Sørensen, or if detected at all by SPIEC-EASI. Time-series and culture networks consisting
527 of OTU connections detected by at least two correlation methods were visualized in
528 Cytoscape. All further statistical analyses of the networks were performed in Cytoscape
529 except for assortativity analyses which were done in R using the igraph v1.1.2 package
530 nominal assortativity function. Sub-networks of the identified dominant genera ("*Ca.*
531 Planktophila", "*Ca*. Nanopelagicus", and "*Ca*. Fonsibacter") and their primary cohorts
532 were extracted from the two initial networks and visualized separately. Layout for the

533 networks were "edge-weighted spring-embedded" with edge weights calculated from the
534 average of the correlations. Correlations between *Nanopelagicales* OTU and "*Ca*.
535 Fonsibacter" OTU detection in the time-series versus the mixed cultures were investigated
536 by Pearson correlation of their relative abundance (log av.ra), prevalence (percent samples
537 or cultures detected in) and relative abundance versus prevalence to determine if detection
538 in culture was correlated to occurrence in the lake.
539
540 **Metagenomic sequencing**
541 Samples for metagenomic sequencing were selected by highest relative abundance of
542 clades of interest according to amplicon data, greatest available DNA mass, and greatest
543 variation in environmental parameters. This was done to examine samples with the greatest
544 potential divergence of environmental conditions under which microbes of interest were
545 most abundant. These included samples taken during lake circulation and summer
546 stratification which separately including hypo, meta and epi-limnionic samples. Two
547 samples per category were selected from different years. Lower biomass during ice cover
548 meant there was insufficient DNA for metagenomic sequencing. DNA library for
549 metagenomic sequencing was prepped using the Illumina TruSeq Nano, 550bp NeoPrep
550 and sequenced with Illumina HiSeq 2500 High Output V4 PE 2x125bp at SciLifeLab
551 Stockholm.
552

553 **Metagenome bioinformatics**

554 HiSeq Illumina paired end reads were pre-processed first using SeqPrep v1.2 (64) to filter
555 out low quality reads using default settings resulting in removal of all reads less than 30b
556 in length, un-paired reads and initial clipping of adapter sequences. Next sequences were
557 cleaned using Trimmomatic v0.36 (65) with removal of any remaining Illumina adapters
558 allowing for 2 mismatches; followed by 4b sliding window trim with minimum average
559 quality of 15; then trimming of read ends with 'maxinfo' value less than 1. The eight
560 metagenomes were individually assembled with megahit (66) using iterative kmers from
561 21 to 121 at 10b intervals, minimum coverage of 2, and minimum length of 200.
562 Concatenated assemblies were cleaned of Illumina artifact low complexity (long
563 homopolymer) contigs with prinseq-lite v0.20.4 (67) removing contigs consisting of
564 greater than 80% one nucleotide.

565

566 **Investigation of katG gene and its product: catalase-peroxidase I**

567 Known impediments to axenic growth of freshwater isolates are multiple auxotrophies, and
568 slow growth and replication rates. The successful axenic cultivation of "*Ca*. Planktophila"
569 strains by addition of high functioning catalase to reduce oxidative stress, in addition to
570 catering for predicted auxotrophies, showed that there can be other considerations. To
571 investigate how both "*Ca*. Planktophila" and "*Ca*. Nanopelagicus" phylotypes grow in
572 Lake Erken and why "*Ca*. Nanopelagicus" grew more frequently in the minimalist mixed

573 cultures, the katG gene was investigated. First katG amino-acid sequences for species
574 identified as belonging to the Nanopelagicales and its primary cohort were retrieved from
575 IMG and GenBank (68, 69). Second sequences for whom the gene product has been
576 assayed and/or the 3D crystal structure is known were identified via Phyre2 and then
577 collected from RCSB PDB (rscb.org (70, 71)). Next these sequences were used as BLAST
578 (72) queries to retrieve closely related (> 80% aa identity) sequences from Lake Erken
579 metagenomes. The collected amino-acid sequences were aligned using MAFFT e-insi and
580 "leavegappy" options (73). Sequences with greater than 99% identity were identified using
581 CD-Hit and removed (74) while sequences from uniRef useful for constructing a guide tree
582 were then also added (75, 76). The alignment was evaluated for best amino-acid
583 substitution model using IQ-tree v.1.6.12 modelfinder (77). This alignment was then
584 analysed and used to construct a tree in RAXML v.0.9.0 (78) with WAG+IG+R4 model
585 and 1000 bootstraps and catalase-peroxidase from *Oryza rufipogon* set as outgroup. To
586 examine functionality secondary structure was predicted and aligned to known models then
587 tertiary structure predicted in Phyre^2 (71). Further the location of katG genes in
588 Nanopelagicales genomes and MAGs was examined in IMG using the gene-
589 neighbourhood function. Images were downloaded with screen shot and processed in
590 Inkscape.
591
592 **Data availability**

593 Time-series MiSeq sequences were deposited in ENA
594 (https://www.ebi.ac.uk/ena/browser/view/PRJEB20109) under accession numbers:
595 ERS1884053 to ERS1884254, sample details including MIDS are also in Table S6. Mixed-
596 culture sequences are available from ENA PRJEB35605
597 (https://www.ebi.ac.uk/ena/browser/view/PRJEB35605) with details also in Table S7. 454
598 data used in the time-series is available at SRA (https://trace.ncbi.nlm.nih.gov/Traces/sra/)
599 under accession SRR097418 (11). HiSeq metagenome sequences are available at ENA
600 project PRJEB37497 under accession numbers: ERS4415328 to ERS4415335 and the
601 assembled metagenomes are available at IMG (https://img.jgi.doe.gov/cgi-bin/m/main.cgi)
602 with TOIDs: 3300020141, 3300020151, 3300020159, 3300020160, 3300020161,
603 3300020162, 3300020172, & 3300020205 details in Table S9. Rscripts used for correlation
604 and network analyses are available from https://github.com/rmondav/publications.
605

606 **ACKNOWLEDGEMENTS**

607 RM, sequencing of the metagenomes, and publication supported by grants from the
608 Malméns Stiftelsen. SG was supported by VR (grant 2012-4592), Olsson-Borgh, Knut and
609 Alice Wallenberg Foundation (grant KAW 2013.0091), SciLifeLab Fellowship, and Kungl.
610 Vetenskapsakademiens stiftelser (BS2017-0044). SB was supported by grants from the
611 Swedish Research Council (VR) and the Swedish Research Council Formas. The
612 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-

633 talk between ecologists and microbiome scientists. Front. Microbiol. **10**:1–10.

634 2. **Koeppel A**, **Perry EB**, **Sikorski J**, **Krizanc D**, **Warner A**, **Ward DM**, **Rooney AP**,
635 **Brambilla E**, **Connor N**, **Ratcliff RM**, **Nevo E**, **Cohan FM**. 2008. Identifying the
636 fundamental units of bacterial diversity: A paradigm shift to incorporate
637 ecology into bacterial systematics. Proc. Natl. Acad. Sci. U. S. A. **105**:2504–2509.

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

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

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

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

649 7. **Pettersson K**, **Grust K**, **Weyhenmeyer G**, **Blenckner T**. 2003. Seasonality of
650 chlorophyll and nutrients in Lake Erken - Effects of weather
651 conditionsHydrobiologia.

652 8. **Haglund A-L**, **Törnblom E**, **Boström B**, **Tranvik L**. 2002. Large differences in
653 the fraction of active bacteria in plankton, sediments, and biofilm. Microb. Ecol.
654 **43**:232–41.
655 9. **Lymer D**, **Lindström ES**. 2010. Changing phosphorus concentration and
656 subsequent prophage induction alter composition of a freshwater viral
657 assemblage. Freshw. Biol. **55**:1984–1996.
658 10. **Burgmer T**, **Reiss J**, **Wickham S**, **Hillebrand H**. 2010. Effects of snail grazers
659 and light on the benthic microbial food web in periphyton communities. Aquat.
660 Microb. Ecol. **61**:163–178.
661 11. **Eiler A**, **Heinrich F**, **Bertilsson S**. 2012. Coherent dynamics and association
662 networks among lake bacterioplankton taxa. ISME J. **6**:330–42.
663 12. **Heinrich F**, **Eiler A**, **Bertilsson S**. 2013. Seasonality and environmental control
664 of freshwater SAR11 (LD12) in a temperate lake (Lake Erken, Sweden). Aquat.
665 Microb. Ecol. **70**:33–44.
666 13. **Eiler A**, **Bertilsson S**. 2004. Composition of freshwater bacterial communities
667 associated with cyanobacterial blooms in four Swedish lakes. Environ.
668 Microbiol. **6**:1228–1243.
669 14. **Henson MW**, **Lanclos VC**, **Faircloth BC**, **Thrash JC**. 2018. Cultivation and
670 genomics of the first freshwater SAR11 (LD12) isolate. ISME J. **12**:1846–1860.

671 15. **Newton RJ**, **Jones SE**, **Eiler A**, **McMahon KD**, **Bertilsson S**. 2011. A guide to the
672 natural history of freshwater lake bacteria.Microbiology and molecular biology
673 reviews : MMBR.
674 16. **Warnecke F**, **Amann R**, **Pernthaler J**. 2004. Actinobacterial 16S rRNA genes
675 from freshwater habitats cluster in four distinct lineages. Environ. Microbiol.
676 17. **Warnecke F**, **Sommaruga R**, **Sekar R**, **Hofer JS**, **Pernthaler J**. 2005.
677 Abundances, identity, and growth state of actinobacteria in mountain lakes of
678 different UV transparency. Appl. Environ. Microbiol.
679 18. **Glockner FO**, **Zaichikov E**, **Belkova N**, **Denissova L**, **Pernthaler J**, **Pernthaler**
680 **A**, **Amann R**. 2000. Comparative 16S rRNA analysis of lake bacterioplankton
681 reveals globally distributed phylogenetic clusters including an abundant group
682 of actinobacteria. Appl. Environ. Microbiol.
683 19. **Li J**, **Zhang J**, **Liu L**, **Fan Y**, **Li L**, **Yang Y**, **Lu Z**, **Zhang X**. 2015. Annual periodicity
684 in planktonic bacterial and archaeal community composition of eutrophic Lake
685 Taihu. Sci. Rep. **5**:1–14.
686 20. **Comte J**, **Lovejoy C**, **Crevecoeur S**, **Vincent WF**. 2016. Co-occurrence patterns
687 in aquatic bacterial communities across changing permafrost landscapes.
688 Biogeosciences **13**:175–190.
689 21. **Garcia SL**, **Buck M**, **Hamilton JJ**, **Wurzbacher C**, **Grossart H-P**, **McMahon KD**,

690 **Eiler A**. 2018. Model Communities Hint at Promiscuous Metabolic Linkages
691 between Ubiquitous Free-Living Freshwater Bacteria. mSphere **3**:1–8.
692 22. **Eckert EM**, **Baumgartner M**, **Huber IM**, **Pernthaler J**. 2013. Grazing resistant
693 freshwater bacteria profit from chitin and cell-wall-derived organic carbon.
694 Environ. Microbiol. **15**:2019–2030.
695 23. **Neuenschwander SM**, **Ghai R**, **Pernthaler J**, **Salcher MM**. 2017.
696 Microdiversification in genome-streamlined ubiquitous freshwater
697 Actinobacteria. ISME J. **12**:185–198.
698 24. **Eiler A**, **Mondav R**, **Sinclair L**, **Fernandez-Vidal L**, **Scofield DDG**, **Scwientek**
699 **P**, **Martinez-Garcia M**, **Torrents D**, **McMahon KDKD**, **Andersson SGESGE**,
700 **Stepanauskas R**, **Woyke T**, **Bertilsson S**, **Schwientek P**, **Martinez-Garcia M**,
701 **Torrents D**, **McMahon KDKD**, **Andersson SGESGE**, **Stepanauskas R**, **Woyke**
702 **T**, **Bertilsson S**. 2015. Tuning fresh: radiation through rewiring of central
703 metabolism in streamlined bacteria. ISME J. **10**:1–13.
704 25. **Kim S**, **Kang I**, **Seo J-HH**, **Cho J-CC**. 2019. Culturing the ubiquitous freshwater
705 actinobacterial acI lineage by supplying a biochemical 'helper' catalase. ISME J.
706 **112**:343640.
707 26. **Rodriguez-R L**, **Tsementzi D**, **Luo C**, **Konstantinidis K**. 2019. Iterative
708 Subtractive Binning of Freshwater Chronoseries Metagenomes Identifies of

709 over Four Hundred Novel Species and their Ecologic Preferences. bioRxiv
710 826941.
711 27. **Giovannoni SJ**, **Cameron Thrash J**, **Temperton B**. 2014. Implications of
712 streamlining theory for microbial ecology. ISME J. Nature Publishing Group.
713 28. **Tripp HJ**. 2013. The unique metabolism of SAR11 aquatic bacteria. J. Microbiol.
714 Seoul Korea **51**:147–53.
715 29. **Pernthaler J**, **Posch T**, **Šimek K**, **Vrba J**, **Pernthaler A**, **Glöckner FO**, **Nübel U**,
716 **Psenner R**, **Amann R**. 2001. Predator-Specific Enrichment of Actinobacteria
717 from a Cosmopolitan Freshwater Clade in Mixed Continuous Culture. Appl.
718 Environ. Microbiol.
719 30. **Salcher MM**. 2014. Same same but different: Ecological niche partitioning of
720 planktonic freshwater prokaryotes. J. Limnol. **73**:74–87.
721 31. **Dadon-Pilosof A**, **Conley KR**, **Jacobi Y**, **Haber M**, **Lombard F**, **Sutherland KR**,
722 **Steindler L**, **Tikochinski Y**, **Richter M**, **Oliver Glamp F**, **Suzuki MT**, **West NJ**,
723 **Genin A**, **Yahel G**. 2017. Surface properties of SAR11 bacteria facilitate grazing
724 avoidance. Nat. Microbiol. 1.
725 32. **Våge S**, **Storesund JE**, **Thingstad TF**. 2013. SAR11 viruses and defensive host
726 strains. Nature **499**:E3-4.
727 33. **Morris JJ**, **Lenski RE**, **Zinser ER**. 2012. The Black Queen Hypothesis: Evolution

728 of Dependencies through Adaptive Gene Loss. MBio.

729 34. **Pacheco AR**, **Moel M**, **Segrè D**. 2019. Costless metabolic secretions as drivers
730 of interspecies interactions in microbial ecosystems. Nat. Commun. **10**:103.

731 35. **Pande S**, **Kost C**. 2017. Bacterial unculurability and the formation of
732 intercellular metabolic networks. Trends Microbiol. **xx**:1–13.

733 36. **Hughes BB**, **Beas-Luna R**, **Barner AK**, **Brewitt K**, **Brumbaugh DR**, **Cerny-**
734 **Chipman EB**, **Close SL**, **Coblentz KE**, **de Nesnera KL**, **Drobnitch ST**, **Figurski**
735 **JD**, **Focht B**, **Friedman M**, **Freiwald J**, **Heady KK**, **Heady WN**, **Hettinger A**,
736 **Johnson A**, **Karr KA**, **Mahoney B**, **Moritsch MM**, **Osterback A-MK**, **Reimer J**,
737 **Robinson J**, **Rohrer T**, **Rose JM**, **Sabal M**, **Segui LM**, **Shen C**, **Sullivan J**,
738 **Zuercher R**, **Raimondi PT**, **Menge BA**, **Grorud-Colvert K**, **Novak M**, **Carr MH**.
739 2017. Long-Term Studies Contribute Disproportionately to Ecology and Policy.
740 Bioscience **67**:271–281.

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

745 38. **Ducklow HW**, **Doney SC**, **Steinberg DK**. 2009. Contributions of Long-Term
746 Research and Time-Series Observations to Marine Ecology and

Biogeochemistry. Ann. Rev. Mar. Sci. **1**:279–302.

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

40. **Philippot L**, **Andersson SGE**, **Battin TJ**, **Prosser JI**, **Schimel JP**, **Whitman WB**, **Hallin S**. 2010. The ecological coherence of high bacterial taxonomic ranks. Nat. Rev. Microbiol. **8**:523–9.

41. **Garcia SL**, **Salka I**, **Grossart HP**, **Warnecke F**. 2013. Depth-discrete profiles of bacterial communities reveal pronounced spatio-temporal dynamics related to lake stratification. Environ. Microbiol. Rep. **5**:549–555.

42. **Garcia SL**, **Stevens SLR**, **Crary B**, **Martinez-Garcia M**, **Stepanauskas R**, **Woyke T**, **Tringe SG**, **Andersson SGE**, **Bertilsson S**, **Malmstrom RR**, **McMahon KD**. 2017. Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations. ISME J.

43. **D'Autréaux B**, **Toledano MB**. 2007. ROS as signalling molecules: Mechanisms that generate specificity in ROS homeostasis. Nat. Rev. Mol. Cell Biol. Nature Publishing Group.

44. **Koonin E V**, **Wolf YI**. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res.2008/10/25. **36**:6688–6719.

766 45. **Bienert GP**, **Chaumont F**. 2014. Aquaporin-facilitated transmembrane
767 diffusion of hydrogen peroxide. Biochim. Biophys. Acta - Gen. Subj. **1840**:1596–
768 1604.
769 46. **Zaremba-Niedzwiedzka K**, **Viklund J**, **Zhao W**, **Ast J**, **Sczyrba A**, **Woyke T**,
770 **McMahon K**, **Bertilsson S**, **Stepanauskas R**, **Andersson SGE**. 2013. Single-cell
771 genomics reveal low recombination frequencies in freshwater bacteria of the
772 SAR11 clade. Genome Biol. **14**:R130.
773 47. **Tsementzi D**, **Rodriguez-R LM**, **Ruiz-Perez CA**, **Meziti A**, **Hatt JK**,
774 **Konstantinidis KT**. 2019. Ecogenomic characterization of widespread, closely-
775 related SAR11 clades of the freshwater genus "Candidatus Fonsibacter" and
776 proposal of Ca. Fonsibacter lacus sp. nov. Syst. Appl. Microbiol. **42**:495–505.
777 48. **Moore ER**, **Davie-Martin CL**, **Giovannoni SJ**, **Halsey KH**. 2019. Pelagibacter
778 metabolism of diatom-derived volatile organic compounds imposes an
779 energetic tax on photosynthetic carbon fixation. Environ. Microbiol. **00**:4–6.
780 49. **Herlemann DP**, **Labrenz M**, **Jürgens K**, **Bertilsson S**, **Waniek JJ**, **Andersson**
781 **AF**. 2011. Transitions in bacterial communities along the 2000 km salinity
782 gradient of the Baltic Sea. ISME J. **5**:1571–9.
783 50. **Caporaso JG**, **Kuczynski J**, **Stombaugh J**, **Bittinger K**, **Bushman FD**, **Costello**
784 **EK**, **Fierer N**, **Peña AG**, **Goodrich JK**, **Gordon JI**, **Huttley GA**, **Kelley ST**,

785 **Knights D**, **Koenig JE**, **Ley RE**, **Lozupone CA**, **McDonald D**, **Muegge BD**,
786 **Pirrung M**, **Reeder J**, **Sevinsky JR**, **Turnbaugh PJ**, **Walters WA**, **Widmann J**,
787 **Yatsunenko T**, **Zaneveld J**, **Knight R**. 2010. QIIME allows analysis of high-
788 throughput community sequencing data. Nat. Methods **7**:335–336.
789 51. **Edgar RC**. 2010. Search and clustering orders of magnitude faster than BLAST.
790 Bioinformatics **26**:2460–2461.
791 52. **McDonald D**, **Price MN**, **Goodrich J**, **Nawrocki EP**, **DeSantis TZ**, **Probst A**,
792 **Andersen GL**, **Knight R**, **Hugenholtz P**. 2012. An improved Greengenes
793 taxonomy with explicit ranks for ecological and evolutionary analyses of
794 bacteria and archaea. ISME J. **6**:610–8.
795 53. **Yilmaz P**, **Parfrey LW**, **Yarza P**, **Gerken J**, **Pruesse E**, **Quast C**, **Schweer T**,
796 **Peplies J**, **Ludwig W**, **Glöckner FO**. 2014. The SILVA and "all-species Living
797 Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res. **42**:643–648.
798 54. **R-Core-team**. 2011. R: A Language and Environment for Statistical Computing.
799 2.14.1. R Foundation for Statistical Computing, Vienna, Austria.
800 55. **RStudio**. 2012. RStudio: Integrated development environment for R. 0.97.248.
801 Boston, MA.
802 56. **Kampstra P**. 2008. Beanplot: A Boxplot Alternative for Visual Comparison of
803 Distributions. J. Stat. Softw. **28**:1–9.

804 57. **Giraudoux P**. 2012. pgirmess: Data analysis in ecology. R package version 1.5.6.

805 58. **Weiss S**, **Van Treuren W**, **Lozupone C**, **Faust K**, **Friedman J**, **Deng Y**, **Xia LC**,
806 **Xu ZZ**, **Ursell L**, **Alm EJ**, **Birmingham A**, **Cram JA**, **Fuhrman JA**, **Raes J**, **Sun F**,
807 **Zhou J**, **Knight R**. 2016. Correlation detection strategies in microbial data sets
808 vary widely in sensitivity and precision. ISME J. **10**:1669–1681.

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

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

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

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

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

822 64. 2014. SeqPrep. 1.2.

823 65. **Bolger AM**, **Lohse M**, **Usadel B**. 2014. Trimmomatic: A flexible trimmer for
824 Illumina sequence data. Bioinformatics **30**:2114–2120.
825 66. **Li D**, **Liu CM**, **Luo R**, **Sadakane K**, **Lam TW**. 2014. MEGAHIT: An ultra-fast
826 single-node solution for large and complex metagenomics assembly via
827 succinct de Bruijn graph. Bioinformatics **31**:1674–1676.
828 67. **Schmieder R**, **Edwards R**. 2011. Quality control and preprocessing of
829 metagenomic datasets. Bioinformatics **27**:863–864.
830 68. **Markowitz VM**, **Chen I-M a**, **Palaniappan K**, **Chu K**, **Szeto E**, **Grechkin Y**,
831 **Ratner A**, **Jacob B**, **Huang J**, **Williams P**, **Huntemann M**, **Anderson I**,
832 **Mavromatis K**, **Ivanova NN**, **Kyrpides NC**. 2012. IMG: the Integrated Microbial
833 Genomes database and comparative analysis system. Nucleic Acids Res.
834 **40**:D115-22.
835 69. **Sayers EW**, **Barrett T**, **Benson D a**, **Bolton E**, **Bryant SH**, **Canese K**,
836 **Chetvernin V**, **Church DM**, **DiCuccio M**, **Federhen S**, **Feolo M**, **Fingerman IM**,
837 **Geer LY**, **Helmberg W**, **Kapustin Y**, **Landsman D**, **Lipman DJ**, **Lu Z**, **Madden**
838 **TL**, **Madej T**, **Maglott DR**, **Marchler-Bauer A**, **Miller V**, **Mizrachi I**, **Ostell J**,
839 **Panchenko A**, **Phan L**, **Pruitt KD**, **Schuler GD**, **Sequeira E**, **Sherry ST**,
840 **Shumway M**, **Sirotkin K**, **Slotta D**, **Souvorov A**, **Starchenko G**, **Tatusova T a**,
841 **Wagner L**, **Wang Y**, **Wilbur WJ**, **Yaschenko E**, **Ye J**. 2011. Database resources

842 of the National Center for Biotechnology Information. Nucleic Acids Res.
843 **39**:D38-51.
844 70. **Berman HM**, **Westbrook J**, **Feng Z**, **Gilliland G**, **Bhat TN**, **Weissig H**,
845 **Shindyalov IN**, **Bourne PE**. 2000. The Protein Data BankNucleic Acids
846 Research.
847 71. **Kelley LA**, **Mezulis S**, **Yates CM**, **Wass MN**, **Sternberg MJE**. 2015. The Phyre2
848 web portal for protein modeling, prediction and analysis. Nat. Protoc. **10**:845–
849 858.
850 72. **Camacho C**, **Coulouris G**, **Avagyan V**, **Ma N**, **Papadopoulos J**, **Bealer K**,
851 **Madden TL**. 2009. BLAST+: architecture and applications. BMC Bioinformatics
852 **10**:421.
853 73. **Katoh K**, **Kuma KI**, **Toh H**, **Miyata T**. 2005. MAFFT version 5: Improvement in
854 accuracy of multiple sequence alignment. Nucleic Acids Res. **33**:511–518.
855 74. **Niu B**, **Fu L**, **Sun S**, **Li W**. 2010. Artificial and natural duplicates in
856 pyrosequencing reads of metagenomic data. BMC Bioinformatics **11**:187.
857 75. **Suzek BE**, **Huang H**, **McGarvey P**, **Mazumder R**, **Wu CH**. 2007. UniRef:
858 comprehensive and non-redundant UniProt reference clusters. Bioinformatics
859 **23**:1282–8.
860 76. **Magrane M**, **Consortium U**. 2011. UniProt Knowledgebase: a hub of integrated

protein data. Database (Oxford). **2011**:bar009.

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

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

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

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

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

880 5. **Eriksson C**, **Forsberg C**. 1992. Nutrient Interactions and Phytoplankton
881 Growth during the Spring Bloom Period in Lake Erken, Sweden. Int. Rev. der
882 gesamten Hydrobiol. und Hydrogr.
883 6. **Bell RT**, **Stensdotter U**, **Pettersson K**, **Istanovics V**, **Pierson DC**. 1998.
884 Microbial dynamics and phosphorus turnover in Lake ErkenAdvances in
885 Limnology 51: Lake Erken - 50 Years of Limnological Research.
886 7. **Pettersson K**, **Grust K**, **Weyhenmeyer G**, **Blenckner T**. 2003. Seasonality of
887 chlorophyll and nutrients in Lake Erken - Effects of weather
888 conditionsHydrobiologia.
889 8. **Haglund A-L**, **Törnblom E**, **Boström B**, **Tranvik L**. 2002. Large differences in
890 the fraction of active bacteria in plankton, sediments, and biofilm. Microb. Ecol.
891 **43**:232–41.
892 9. **Lymer D**, **Lindström ES**. 2010. Changing phosphorus concentration and
893 subsequent prophage induction alter composition of a freshwater viral
894 assemblage. Freshw. Biol. **55**:1984–1996.
895 10. **Burgmer T**, **Reiss J**, **Wickham S**, **Hillebrand H**. 2010. Effects of snail grazers
896 and light on the benthic microbial food web in periphyton communities. Aquat.
897 Microb. Ecol. **61**:163–178.
898 11. **Eiler A**, **Heinrich F**, **Bertilsson S**. 2012. Coherent dynamics and association

899 networks among lake bacterioplankton taxa. ISME J. **6**:330–42.

900 12. **Heinrich F**, **Eiler A**, **Bertilsson S**. 2013. Seasonality and environmental control
901 of freshwater SAR11 (LD12) in a temperate lake (Lake Erken, Sweden). Aquat.
902 Microb. Ecol. **70**:33–44.

903 13. **Eiler A**, **Bertilsson S**. 2004. Composition of freshwater bacterial communities
904 associated with cyanobacterial blooms in four Swedish lakes. Environ.
905 Microbiol. **6**:1228–1243.

906 14. **Henson MW**, **Lanclos VC**, **Faircloth BC**, **Thrash JC**. 2018. Cultivation and
907 genomics of the first freshwater SAR11 (LD12) isolate. ISME J. **12**:1846–1860.

908 15. **Newton RJ**, **Jones SE**, **Eiler A**, **McMahon KD**, **Bertilsson S**. 2011. A guide to the
909 natural history of freshwater lake bacteria.Microbiology and molecular biology
910 reviews : MMBR.

911 16. **Warnecke F**, **Amann R**, **Pernthaler J**. 2004. Actinobacterial 16S rRNA genes
912 from freshwater habitats cluster in four distinct lineages. Environ. Microbiol.

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

916 18. **Glockner FO**, **Zaichikov E**, **Belkova N**, **Denissova L**, **Pernthaler J**, **Pernthaler**
917 **A**, **Amann R**. 2000. Comparative 16S rRNA analysis of lake bacterioplankton

918 reveals globally distributed phylogenetic clusters including an abundant group
919 of actinobacteria. Appl. Environ. Microbiol.
920 19. **Li J**, **Zhang J**, **Liu L**, **Fan Y**, **Li L**, **Yang Y**, **Lu Z**, **Zhang X**. 2015. Annual periodicity
921 in planktonic bacterial and archaeal community composition of eutrophic Lake
922 Taihu. Sci. Rep. **5**:1–14.
923 20. **Comte J**, **Lovejoy C**, **Crevecoeur S**, **Vincent WF**. 2016. Co-occurrence patterns
924 in aquatic bacterial communities across changing permafrost landscapes.
925 Biogeosciences **13**:175–190.
926 21. **Garcia SL**, **Buck M**, **Hamilton JJ**, **Wurzbacher C**, **Grossart H-P**, **McMahon KD**,
927 **Eiler A**. 2018. Model Communities Hint at Promiscuous Metabolic Linkages
928 between Ubiquitous Free-Living Freshwater Bacteria. mSphere **3**:1–8.
929 22. **Eckert EM**, **Baumgartner M**, **Huber IM**, **Pernthaler J**. 2013. Grazing resistant
930 freshwater bacteria profit from chitin and cell-wall-derived organic carbon.
931 Environ. Microbiol. **15**:2019–2030.
932 23. **Neuenschwander SM**, **Ghai R**, **Pernthaler J**, **Salcher MM**. 2017.
933 Microdiversification in genome-streamlined ubiquitous freshwater
934 Actinobacteria. ISME J. **12**:185–198.
935 24. **Eiler A**, **Mondav R**, **Sinclair L**, **Fernandez-Vidal L**, **Scofield DDG**, **Scwientek**
936 **P**, **Martinez-Garcia M**, **Torrents D**, **McMahon KDKD**, **Andersson SGESGE**,

Stepanauskas R, **Woyke T**, **Bertilsson S**, **Schwientek P**, **Martinez-Garcia M**, **Torrents D**, **McMahon KDKD**, **Andersson SGESGE**, **Stepanauskas R**, **Woyke T**, **Bertilsson S**. 2015. Tuning fresh: radiation through rewiring of central metabolism in streamlined bacteria. ISME J. **10**:1–13.

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

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

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

28. **Tripp HJ**. 2013. The unique metabolism of SAR11 aquatic bacteria. J. Microbiol. Seoul Korea **51**:147–53.

29. **Pernthaler J**, **Posch T**, **Šimek K**, **Vrba J**, **Pernthaler A**, **Glöckner FO**, **Nübel U**, **Psenner R**, **Amann R**. 2001. Predator-Specific Enrichment of Actinobacteria from a Cosmopolitan Freshwater Clade in Mixed Continuous Culture. Appl. Environ. Microbiol.

956 30. **Salcher MM**. 2014. Same same but different: Ecological niche partitioning of
957 planktonic freshwater prokaryotes. J. Limnol. **73**:74–87.
958 31. **Dadon-Pilosof A**, **Conley KR**, **Jacobi Y**, **Haber M**, **Lombard F**, **Sutherland KR**,
959 **Steindler L**, **Tikochinski Y**, **Richter M**, **Oliver Glamp F**, **Suzuki MT**, **West NJ**,
960 **Genin A**, **Yahel G**. 2017. Surface properties of SAR11 bacteria facilitate grazing
961 avoidance. Nat. Microbiol. 1.
962 32. **Våge S**, **Storesund JE**, **Thingstad TF**. 2013. SAR11 viruses and defensive host
963 strains. Nature **499**:E3-4.
964 33. **Morris JJ**, **Lenski RE**, **Zinser ER**. 2012. The Black Queen Hypothesis: Evolution
965 of Dependencies through Adaptive Gene Loss. MBio.
966 34. **Pacheco AR**, **Moel M**, **Segrè D**. 2019. Costless metabolic secretions as drivers
967 of interspecies interactions in microbial ecosystems. Nat. Commun. **10**:103.
968 35. **Pande S**, **Kost C**. 2017. Bacterial unculurability and the formation of
969 intercellular metabolic networks. Trends Microbiol. **xx**:1–13.
970 36. **Hughes BB**, **Beas-Luna R**, **Barner AK**, **Brewitt K**, **Brumbaugh DR**, **Cerny-**
971 **Chipman EB**, **Close SL**, **Coblentz KE**, **de Nesnera KL**, **Drobnitch ST**, **Figurski**
972 **JD**, **Focht B**, **Friedman M**, **Freiwald J**, **Heady KK**, **Heady WN**, **Hettinger A**,
973 **Johnson A**, **Karr KA**, **Mahoney B**, **Moritsch MM**, **Osterback A-MK**, **Reimer J**,
974 **Robinson J**, **Rohrer T**, **Rose JM**, **Sabal M**, **Segui LM**, **Shen C**, **Sullivan J**,

975 **Zuercher R**, **Raimondi PT**, **Menge BA**, **Grorud-Colvert K**, **Novak M**, **Carr MH**.
976 2017. Long-Term Studies Contribute Disproportionately to Ecology and Policy.
977 Bioscience **67**:271–281.
978 37. **Brown JH**, **Whitham TG**, **Morgan Ernest SK**, **Gehring CA**. 2001. Complex
979 species interactions and the dynamics of ecological systems: Long-term
980 experiments. Science (80-.). American Association for the Advancement of
981 Science.
982 38. **Ducklow HW**, **Doney SC**, **Steinberg DK**. 2009. Contributions of Long-Term
983 Research and Time-Series Observations to Marine Ecology and
984 Biogeochemistry. Ann. Rev. Mar. Sci. **1**:279–302.
985 39. **Franklin JF**. 1989. Importance and Justification of Long-Term Studies in
986 Ecology, p. 3–19. *In* Long-Term Studies in Ecology. Springer New York.
987 40. **Philippot L**, **Andersson SGE**, **Battin TJ**, **Prosser JI**, **Schimel JP**, **Whitman WB**,
988 **Hallin S**. 2010. The ecological coherence of high bacterial taxonomic ranks. Nat.
989 Rev. Microbiol. **8**:523–9.
990 41. **Garcia SL**, **Salka I**, **Grossart HP**, **Warnecke F**. 2013. Depth-discrete profiles of
991 bacterial communities reveal pronounced spatio-temporal dynamics related to
992 lake stratification. Environ. Microbiol. Rep. **5**:549–555.
993 42. **Garcia SL**, **Stevens SLR**, **Crary B**, **Martinez-Garcia M**, **Stepanauskas R**,

994 **Woyke T**, **Tringe SG**, **Andersson SGE**, **Bertilsson S**, **Malmstrom RR**,
995 **McMahon KD**. 2017. Contrasting patterns of genome-level diversity across
996 distinct co-occurring bacterial populations. ISME J.
997 43. **D'Autréaux B**, **Toledano MB**. 2007. ROS as signalling molecules: Mechanisms
998 that generate specificity in ROS homeostasis. Nat. Rev. Mol. Cell Biol. Nature
999 Publishing Group.
1000 44. **Koonin E V**, **Wolf YI**. 2008. Genomics of bacteria and archaea: the emerging
1001 dynamic view of the prokaryotic world. Nucleic Acids Res.2008/10/25.
1002 **36**:6688–6719.
1003 45. **Bienert GP**, **Chaumont F**. 2014. Aquaporin-facilitated transmembrane
1004 diffusion of hydrogen peroxide. Biochim. Biophys. Acta - Gen. Subj. **1840**:1596–
1005 1604.
1006 46. **Zaremba-Niedzwiedzka K**, **Viklund J**, **Zhao W**, **Ast J**, **Sczyrba A**, **Woyke T**,
1007 **McMahon K**, **Bertilsson S**, **Stepanauskas R**, **Andersson SGE**. 2013. Single-cell
1008 genomics reveal low recombination frequencies in freshwater bacteria of the
1009 SAR11 clade. Genome Biol. **14**:R130.
1010 47. **Tsementzi D**, **Rodriguez-R LM**, **Ruiz-Perez CA**, **Meziti A**, **Hatt JK**,
1011 **Konstantinidis KT**. 2019. Ecogenomic characterization of widespread, closely-
1012 related SAR11 clades of the freshwater genus "Candidatus Fonsibacter" and

1013 proposal of Ca. Fonsibacter lacus sp. nov. Syst. Appl. Microbiol. **42**:495–505.

1014 48. **Moore ER**, **Davie-Martin CL**, **Giovannoni SJ**, **Halsey KH**. 2019. Pelagibacter
1015 metabolism of diatom-derived volatile organic compounds imposes an
1016 energetic tax on photosynthetic carbon fixation. Environ. Microbiol. **00**:4–6.

1017 49. **Herlemann DP**, **Labrenz M**, **Jürgens K**, **Bertilsson S**, **Waniek JJ**, **Andersson**
1018 **AF**. 2011. Transitions in bacterial communities along the 2000 km salinity
1019 gradient of the Baltic Sea. ISME J. **5**:1571–9.

1020 50. **Caporaso JG**, **Kuczynski J**, **Stombaugh J**, **Bittinger K**, **Bushman FD**, **Costello**
1021 **EK**, **Fierer N**, **Peña AG**, **Goodrich JK**, **Gordon JI**, **Huttley GA**, **Kelley ST**,
1022 **Knights D**, **Koenig JE**, **Ley RE**, **Lozupone CA**, **McDonald D**, **Muegge BD**,
1023 **Pirrung M**, **Reeder J**, **Sevinsky JR**, **Turnbaugh PJ**, **Walters WA**, **Widmann J**,
1024 **Yatsunenko T**, **Zaneveld J**, **Knight R**. 2010. QIIME allows analysis of high-
1025 throughput community sequencing data. Nat. Methods **7**:335–336.

1026 51. **Edgar RC**. 2010. Search and clustering orders of magnitude faster than BLAST.
1027 Bioinformatics **26**:2460–2461.

1028 52. **McDonald D**, **Price MN**, **Goodrich J**, **Nawrocki EP**, **DeSantis TZ**, **Probst A**,
1029 **Andersen GL**, **Knight R**, **Hugenholtz P**. 2012. An improved Greengenes
1030 taxonomy with explicit ranks for ecological and evolutionary analyses of
1031 bacteria and archaea. ISME J. **6**:610–8.

1032 53. **Yilmaz P**, **Parfrey LW**, **Yarza P**, **Gerken J**, **Pruesse E**, **Quast C**, **Schweer T**,
1033 **Peplies J**, **Ludwig W**, **Glöckner FO**. 2014. The SILVA and "all-species Living
1034 Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res. **42**:643–648.
1035 54. **R-Core-team**. 2011. R: A Language and Environment for Statistical Computing.
1036 2.14.1. R Foundation for Statistical Computing, Vienna, Austria.
1037 55. **RStudio**. 2012. RStudio: Integrated development environment for R. 0.97.248.
1038 Boston, MA.
1039 56. **Kampstra P**. 2008. Beanplot: A Boxplot Alternative for Visual Comparison of
1040 Distributions. J. Stat. Softw. **28**:1–9.
1041 57. **Giraudoux P**. 2012. pgirmess: Data analysis in ecology. R package version 1.5.6.
1042 58. **Weiss S**, **Van Treuren W**, **Lozupone C**, **Faust K**, **Friedman J**, **Deng Y**, **Xia LC**,
1043 **Xu ZZ**, **Ursell L**, **Alm EJ**, **Birmingham A**, **Cram JA**, **Fuhrman JA**, **Raes J**, **Sun F**,
1044 **Zhou J**, **Knight R**. 2016. Correlation detection strategies in microbial data sets
1045 vary widely in sensitivity and precision. ISME J. **10**:1669–1681.
1046 59. **Oksanen J**, **Blanchet FG**, **Kindt R**, **Legendre P**, **Minchin PR**, **O'Hara B**,
1047 **Simpson GL**, **Solymos P**, **Stevens MHH**, **Wagner H**. 2013. vegan: Community
1048 ecology package. R package version 2.0-7.
1049 60. **Friedman J**, **Alm EJ**. 2012. Inferring correlation networks from genomic survey
1050 data. PLoS Comput. Biol. **8**:e1002687.

1051 61. **Kurtz ZD**, **Müller CL**, **Miraldi ER**, **Littman DR**, **Blaser MJ**, **Bonneau RA**. 2015.
1052 Sparse and Compositionally Robust Inference of Microbial Ecological Networks.
1053 PLoS Comput. Biol. **11**:e1004226.
1054 62. **Sørensen T**. 1948. A method of establishing groups of equal amplitude in plant
1055 sociology based on similarity of species and its application to analyses of the
1056 vegetation on Danish commons. Biol. Skr. **5**:1–34.
1057 63. **Dice LR**. 1945. Measures of the Amount of Ecologic Association Between
1058 Species. Ecology **26**:297–302.
1059 64. 2014. SeqPrep. 1.2.
1060 65. **Bolger AM**, **Lohse M**, **Usadel B**. 2014. Trimmomatic: A flexible trimmer for
1061 Illumina sequence data. Bioinformatics **30**:2114–2120.
1062 66. **Li D**, **Liu CM**, **Luo R**, **Sadakane K**, **Lam TW**. 2014. MEGAHIT: An ultra-fast
1063 single-node solution for large and complex metagenomics assembly via
1064 succinct de Bruijn graph. Bioinformatics **31**:1674–1676.
1065 67. **Schmieder R**, **Edwards R**. 2011. Quality control and preprocessing of
1066 metagenomic datasets. Bioinformatics **27**:863–864.
1067 68. **Markowitz VM**, **Chen I-M a**, **Palaniappan K**, **Chu K**, **Szeto E**, **Grechkin Y**,
1068 **Ratner A**, **Jacob B**, **Huang J**, **Williams P**, **Huntemann M**, **Anderson I**,
1069 **Mavromatis K**, **Ivanova NN**, **Kyrpides NC**. 2012. IMG: the Integrated Microbial

1070 Genomes database and comparative analysis system. Nucleic Acids Res.
1071 **40**:D115-22.
1072 69. **Sayers EW**, **Barrett T**, **Benson D a**, **Bolton E**, **Bryant SH**, **Canese K**,
1073 **Chetvernin V**, **Church DM**, **DiCuccio M**, **Federhen S**, **Feolo M**, **Fingerman IM**,
1074 **Geer LY**, **Helmberg W**, **Kapustin Y**, **Landsman D**, **Lipman DJ**, **Lu Z**, **Madden**
1075 **TL**, **Madej T**, **Maglott DR**, **Marchler-Bauer A**, **Miller V**, **Mizrachi I**, **Ostell J**,
1076 **Panchenko A**, **Phan L**, **Pruitt KD**, **Schuler GD**, **Sequeira E**, **Sherry ST**,
1077 **Shumway M**, **Sirotkin K**, **Slotta D**, **Souvorov A**, **Starchenko G**, **Tatusova T a**,
1078 **Wagner L**, **Wang Y**, **Wilbur WJ**, **Yaschenko E**, **Ye J**. 2011. Database resources
1079 of the National Center for Biotechnology Information. Nucleic Acids Res.
1080 **39**:D38-51.
1081 70. **Berman HM**, **Westbrook J**, **Feng Z**, **Gilliland G**, **Bhat TN**, **Weissig H**,
1082 **Shindyalov IN**, **Bourne PE**. 2000. The Protein Data BankNucleic Acids
1083 Research.
1084 71. **Kelley LA**, **Mezulis S**, **Yates CM**, **Wass MN**, **Sternberg MJE**. 2015. The Phyre2
1085 web portal for protein modeling, prediction and analysis. Nat. Protoc. **10**:845–
1086 858.
1087 72. **Camacho C**, **Coulouris G**, **Avagyan V**, **Ma N**, **Papadopoulos J**, **Bealer K**,
1088 **Madden TL**. 2009. BLAST+: architecture and applications. BMC Bioinformatics

1089 **10**:421.

1090 73. **Katoh K**, **Kuma KI**, **Toh H**, **Miyata T**. 2005. MAFFT version 5: Improvement in
1091 accuracy of multiple sequence alignment. Nucleic Acids Res. **33**:511–518.

1092 74. **Niu B**, **Fu L**, **Sun S**, **Li W**. 2010. Artificial and natural duplicates in
1093 pyrosequencing reads of metagenomic data. BMC Bioinformatics **11**:187.

1094 75. **Suzek BE**, **Huang H**, **McGarvey P**, **Mazumder R**, **Wu CH**. 2007. UniRef:
1095 comprehensive and non-redundant UniProt reference clusters. Bioinformatics
1096 **23**:1282–8.

1097 76. **Magrane M**, **Consortium U**. 2011. UniProt Knowledgebase: a hub of integrated
1098 protein data. Database (Oxford). **2011**:bar009.

1099 77. **Kalyaanamoorthy S**, **Minh BQ**, **Wong TKF**, **Von Haeseler A**, **Jermiin LS**. 2017.
1100 ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat.
1101 Methods **14**:587–589.

1102 78. **Kozlov AM**, **Darriba D**, **Flouri T**, **Morel B**, **Stamatakis A**. 2019. RAxML-NG: a
1103 fast, scalable and user-friendly tool for maximum likelihood phylogenetic
1104 inference. Bioinformatics.

1105

1106 **FIGURE LEGENDS**

1107 Fig. 1. Cooperation models shown in a lake setting. Ecotype Cooperation Model describes
1108 anabolic variability as the main contributor to co-existence while allowing for
1109 recombination and swapping of metabolic modules across strains. Community Cooperation
1110 Model also encompasses anabolic complementarity but between rather than within species
1111 and includes common goods ranging from expensive to costless. The Community Detrital
1112 Model places complementarity both between and within species but describes catabolic
1113 variation.
1114
1115 Fig. 2. OTU networks of the primary cohorts from amplicons of a) time-series aand b)
1116 mixed-cultures. Nodes sized by average relative abundance in time-series and by
1117 percentage of cultures detected in for mixed-cultures. Pink dotted lines indicate mutual
1118 exclusion of OTUs while grey solid lines are co-occurrence.
1119
1120 Fig. 3. The relationship between OTU phylogenetic distance and correlation of the primary
1121 cohort phylotypes to selected genera: *"Ca. Planktophila"* upper-left, "*Ca*. Nanopelagicus"
1122 upper-right, "*Ca*. Fonsibacter" lower-left, and basal "*Ca*. Fonsibacter "lower-right.