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

Running title: Cooperation models for freshwater streamlined bacteria

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

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

[b] National Bioinformatics Infrastructure Sweden, Sweden

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

[d] Swedish University of Agricultural Sciences, Uppsala, 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 *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.

IMPORTANCE

This study examines evolutionary and ecological relationships of three of the most ubiquitous and abundant freshwater bacterial genera. Due to high abundance, these genera might have a significant influence on nutrient cycling in freshwaters worldwide and this study adds a layer of understanding to how seemingly competing clades of bacteria can co-exist by having different cooperation strategies. Our synthesis ties together network and ecological theory with empirical evidence and lays out a framework for how the functioning of populations within complex microbial communities can be studied.

INTRODUCTION

The stable coexistence of species is described by a continuum of inter-specific symbiotic interactions from parasitism through neutralism to mutualism the (1). The extended coexistence of microbial strains is described by intra-specific relationship models limited to niche theory (ecotype differentiation) (2) and predator control of dominant strains (3). Combining these ecological theories of co-existence with phylogenetic information can produce a simplified model with a new holistic perspective for understanding aquatic microbial communities. In aquatic microbial communities, especially for bacterioplankton, physical structures and proximity as seen in soil, microbiome, and biofilm communities (4) is absent. As cooperation models have been developed from the study of such tightly interwoven assemblages their models may not be the most suitable. Absent also is the

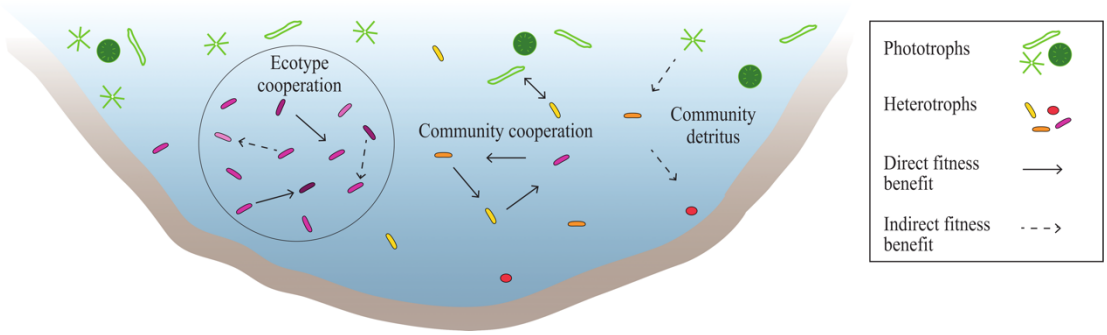
60 habitat isolation which limits dispersal as seen in e.g. soil ecosystems. In aquatic systems
61 interactions between nutrients and biota can occur from micro up to macro-scale distances
62 involving stochastic effectors from Brownian motion up to whole system circulation
63 currents. Direct but distant interactions are therefore difficult to identify based purely on
64 proximity as common goods and toxins in the environment can come from multiple sources
65 and be transported long distances.

66 Lake Erken has been extensively studied for several decades, and ample background
67 information on nutrient cycling, planktonic communities and biogeochemical processes
68 combined with well-established infrastructure makes this ecological observatory an
69 excellent choice for the study of competing microbial assemblage processes (5–7). Several
70 studies have investigated the microbial component of this lake ecosystem (8–13), revealing
71 that the most abundant microbes in Lake Erken belonged to the LD12 (now described as
72 the candidate genus "*Ca.* Fonsibacter" (14)) and acI (ACK-M1, hgcI, but now described as
73 the candidate order *Nanopelagicales*) clades (11, 12, 15–20). "*Ca.* Fonsibacter" is a non-
74 marine genus within the ubiquitous and abundant *Pelagibacteraceae* family which includes
75 the marine *Pelagibacter* genus (SAR11). *Nanopelagicales* is a freshwater order with two
76 described genera, *"Ca.* Planktophila*"* and "*Ca*. Nanopelagicus". Both "*Ca.* Fonsibacter"
77 and the *Nanopelagicales* have small cell size, reduced genomes (1.16 to 1.48 Mbp),
78 multiple auxotrophies, a requirement for reduced sulfur compounds, have rhodopsins, and

79 have been recalcitrant to maintenance in axenic laboratory culture (14, 21–23). "*Ca.*
80 Fonsibacter" and "*Ca*. Nanopelagicus" genera both have additional amino-acid
81 auxotrophies (23, 24). Recently a single strain of "*Ca.* Fonsibacter" and two strains of the
82 *Nanopelagicales* genus *"Ca.* Planktophila*"* (acI-A and acI-A4) were successfully
83 maintained in axenic cultures by customized liquid media in the case of "*Ca.* Fonsibacter"
84 (14) and the addition of an active enzyme to filtered lake-water media for *"Ca.*
85 Planktophila*"* (25). The difference in culture methods point towards a difference in lifestyle
86 strategies and therefore evolution and ecology in these two abundant and ubiquitous
87 freshwater bacterioplankton clades.
88 The ubiquity of the small-genome bacterial lineages, such as *Nanopelagicales* and "*Ca*.
89 Fonsibacter" in freshwaters, is thought to be a consequence of their streamlined genomes
90 and small cell size (26). On one hand, small cell size increases surface area to volume
91 ratio, and improves diffusion, therefore decreasing time for foraging for non-motile
92 organisms (4). On the other hand, reduced genomes have lost extraneous metabolic
93 functions while having very low gene redundancy, reduced regulatory components, and
94 high coding density (27). This combined streamlining and shrinkage reduces the energetic
95 and material cost of genome, proteome, and cellular maintenance. Because their small
96 genomes encode fewer metabolic functions than average, it has been proposed they require
97 a large effective population (large N_e) within which ecotypes diverge to provide the
98 genomic flexibility usually seen in individual cells of bacteria with larger genomes. A large

99 N_e is also required for purifying selection to be efficient at reducing and maintaining
100 reduced genome size. Such streamlined microbes were first noted in low nutrient aquatic
101 environments and streamlining has since then been suggested as being of selective
102 advantage under reduced nitrogen availability (28). *Nanopelagicales* and "*Ca*.
103 Fonsibacter" clades are dominant and successful in non-marine environments such as Lake
104 Erken, which far from being nutrient limited is classified as mesotrophic on a
105 eutrophication trajectory. Further, as Lake Erken is ice and snow covered in winter
106 (temperature and light limiting) and stratified in summer (thermo and oxy-clines), it is not
107 a stable environment where organisms with reduced regulatory capacity are expected to
108 thrive. It is unlikely that dominance in such productive lakes is made possible purely by
109 streamlining, as the subsequent loss of genomic flexibility required to respond to a dynamic
110 and variable habitat such as Lake Erken would counteract the benefits. Furthermore, large
111 populations attract predators though both clades may escape some grazing predation due
112 to their small size (29, 30). "*Ca*. Fonsibacter" may additionally have similar membrane
113 properties to its sister marine lineage SAR11(31) to further elude grazing. All three genera
114 encode cell membrane modification genes in hypervariable regions predicted to assist
115 escape from viral predation (23, 32). A large N_e where the divergence is focused in
116 immunogenic or predator-escape mechanisms (defense specialist) will help a species
117 survive population sweeps caused by predators, though it does not explain the intra nor
118 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 detritavores 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

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

145 **RESULTS**

146 **Taxon abundance and environmental correlates**

147 The most abundant clades detected in the SSU rRNA gene 8-year time-series were the
148 *Actinobacterial* order *Nanopelagicales* at 29% average relative abundance (av.ra), the
149 eukaryote phytoplankton genus *Stramenopiles* 12% av.r.a, the *Actinobacterial* clade 'acIV'
150 (C111) 11% av.ra, the *β-proteobacterial* family *Comamonadaceae* 5% av.ra, the *α-*
151 *Proteobacterial* candidate genus "*Ca.* Fonsibacter" 4% av.ra, and the *Verrucomicrobial*
152 family *Cerasicoccaceae* 3% av.ra (Fig. S1, Table S1). There were few significant
153 correlations or differences (Pearsons correlation or Kruskal-Wallis (KW) with post-hoc
154 testing to identify which pairs were significantly different (KWmc)) between dominant

155 clade abundance and a broad range of environmental predictors such as lake-cycle/season
156 or physico-chemical conditions of the water (Fig. S2, Table S2). The three significant linear
157 correlations detected had only moderate strength (r >= 0.4) with cloud or wedge-shaped
158 distributions (Fig. S3).
159

160 **Microbial networks**

161 The OTU network from both the time-series and mixed-cultures (Fig. S4) shows
162 modularity and taxonomic assortativity (grouping of related organisms or under-
163 dispersion) at all taxonomic levels with greater assortativity in the mixed-cultures (Table
164 S3). Assortativity increased with taxonomic level peaking at phylum level in the timeseries,
165 while assortativity was lowest at phylum level in the cultures. First neighbours (the primary
166 cohort) of OTUs in networks provide evidence (but not proof) for preferential association
167 between phylotypes, or when negatively correlated may indicate mutual exclusion. The
168 primary cohorts of "*Ca.* Planktophila, "*Ca*. Nanopelagicus", and "*Ca*. Fonsibacter" (Fig.
169 2) are overlapping and include phylotypes from each other's cohorts. Primary cohorts from
170 the time-series and mixed-cultures are similar, though the correlation type and strength
171 vary. There was a moderate correlation between relative abundance of OTUs in the
172 timeseries and relative abundance in cultures (Pearsons r = 0.58, p_{adj} = 0, Table S4) but no
173 other correlations between relative abundance or prevalence (percentage of samples an
174 OTU was detected in) were observed.

175

176 **Actinobacterial order Nanopelagicales (acI)**

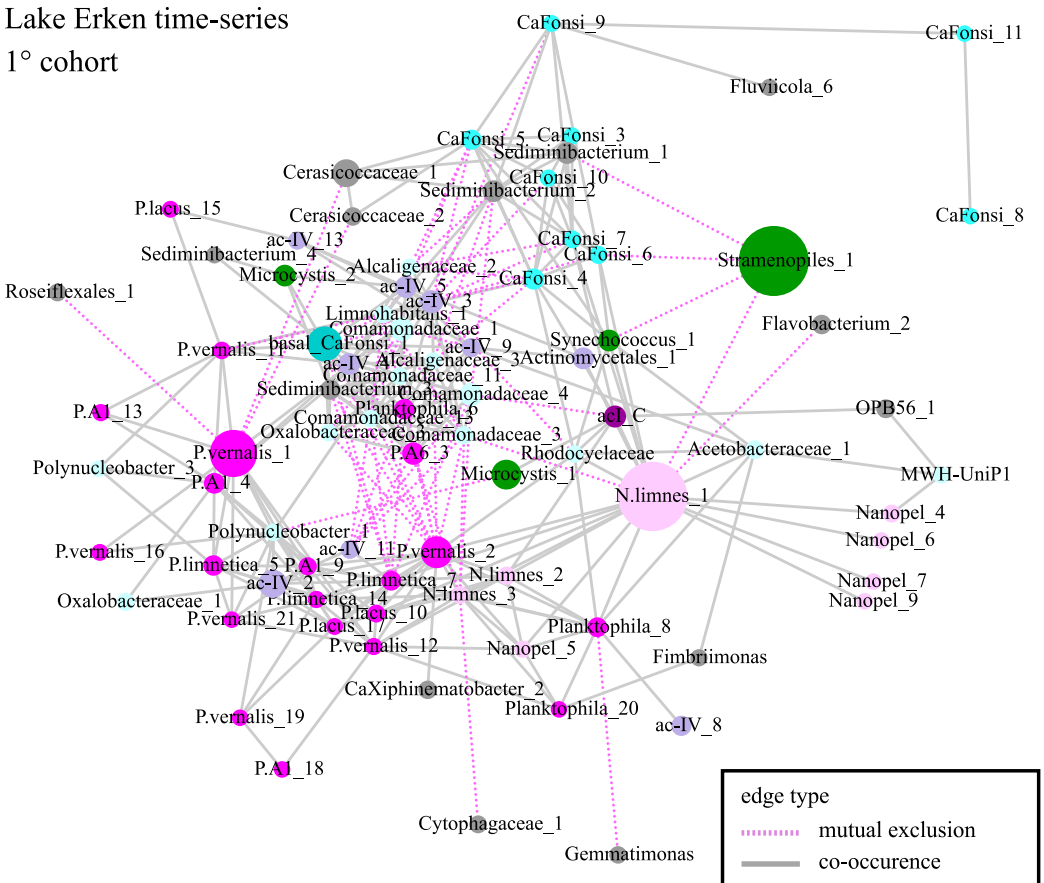
177 *Nanopelagicales* in Erken had higher relative abundances in the hypolimnion, (KW
178 $p<0.001$, KWmc $p<0.001$) and were negatively correlated with % O_2 (Pearsons $r = -0.4$,
179 $p<0.001$). *Nanopelagicales* were the most abundant clade in Lake Erken and were
180 represented by 31 OTUs in the lake time-series network with nine "*Ca*. Nanopelagicus",
181 21 *"Ca.* Planktophila*"*, and one acI-c (Fig. 2, Fig. S4). A subset of six "*Ca*. Nanopelagicus"
182 and one *"Ca.* Planktophila*"* phylotype were detected in the mixed-culture based network
183 (Fig. 2, Fig. S4).

184 The time-series primary (1°) cohort (i.e. the cohort immediately surrounding) of *"Ca.*
185 Planktophila*"* included: other *Actinobacteria*, *Armatimonadetes*, *Chloroflexi*,
186 *Gemmatimonadetes*, *α*- and *β-Proteobacteria*, and *Verrucomicrobia*, while in the mixed-
187 cultures only *Actinobacteria*, *Bacteroidetes*, and *α*- and *β-Proteobacteria* co-existed. A
188 pairwise comparison of the phylogenetic distance and correlation (averaged edge value
189 from the correlation network) was conducted to see if there was a relationship between
190 OTU genetic relatedness and co-occurrence; to distinguish between ecotype and directed
191 reciprocation models. Within the *"Ca.* Planktophila*"* 1° cohort, there was no support for a
192 correlation between phylogenetic distance and co-occurrence (Fig. 3) and *"Ca.*
193 Planktophila*"* was the only genera tested that had negative intra-genus correlations. *"Ca.*

194 Planktophila*"* had more intra-phyla (i.e. non-*Nanopegicales Actinobacteria*) correlations
195 than the other genera tested. Unlike the other genera, *"Ca.* Planktophila*"* had no
196 correlations (positive or negative) with phytoplankton in its 1° network (Fig. 3).
197 Additionally, *"Ca.* Planktophila*"* were the dominant Actinobacterial genus in the time-
198 series with "*Ca*. P. vernalis*"* being most abundant (Fig. 2). Conversely, in the mixed-
199 cultures, only one *Planktophila* phylotype was detected, "*Ca*. P. limnetica*"* (Fig. 2).
200 *"Ca.* Planktophila spp." were successfully maintained in axenic culture via the addition of
201 bovine catalase and therefore the presence, absence, and functionality of the relevant
202 *Nanopelagicales* catalase-peroxidase (KatG) is debated. We investigated the presence of
203 *katG*, its gene, its gene neighbourhood, and the homology and topology of its translated
204 product, KatG, to establish if there is evidence to support its activity level or effect on host
205 abundance. "*Ca*. P. vernalis*"* (acI-A7) KatG had highest homology to the high functioning
206 *Synechoccocus* version (Fig. S5) and the gene is found near a tRNA (Fig. S6). "*Ca*. P.
207 limnetica*"*s KatG had highest identity with the assayed low activity KatG from "*Ca*. P.
208 rubra*"* (IMCC 25003) and was located within a putative DNA repair region (Fig. S5 & S6).
209 When modelled these KatGs were most similar to *Burkholderia pseudomaleii*s with a
210 predicted β-sheet around one of the active site residues instead of an α-helix, and are 10 to
211 20 amino-acid residues longer than known high functioning versions i.e. *Burkholderia*
212 type A (Fig. S5). All known active site residues involved in heme interaction or enzyme
213 switching were however present. *"Ca*. Planktophila limnetica*", "Ca*. P. vernalis*",*

Lake Erken time-series

1° cohort



edge type

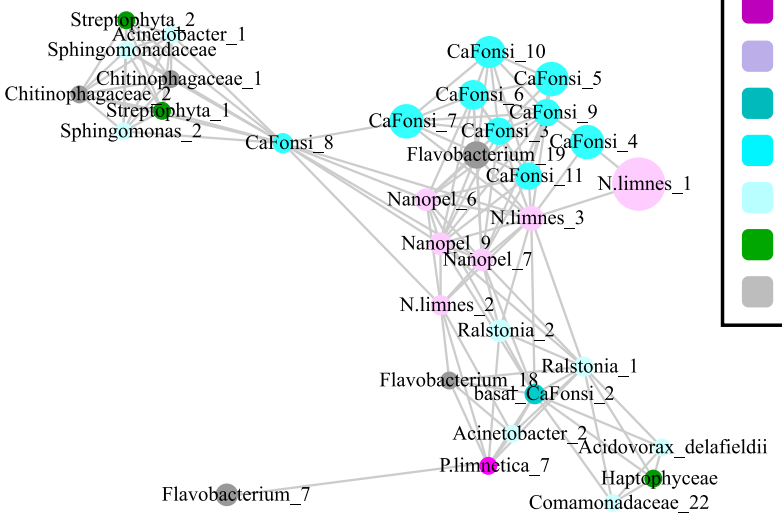
- mutual exclusion
- co-occurence

phylotype

- Ca. Nanopelagicus
- Planktophila
- acI-c
- other Actinobacteria
- basal Ca. Fonsibacter
- Ca.Fonsibacter
- other Proteobacteria
- Cyanobacteria & chloroplasts
- other

Mixed-cultures from lake water

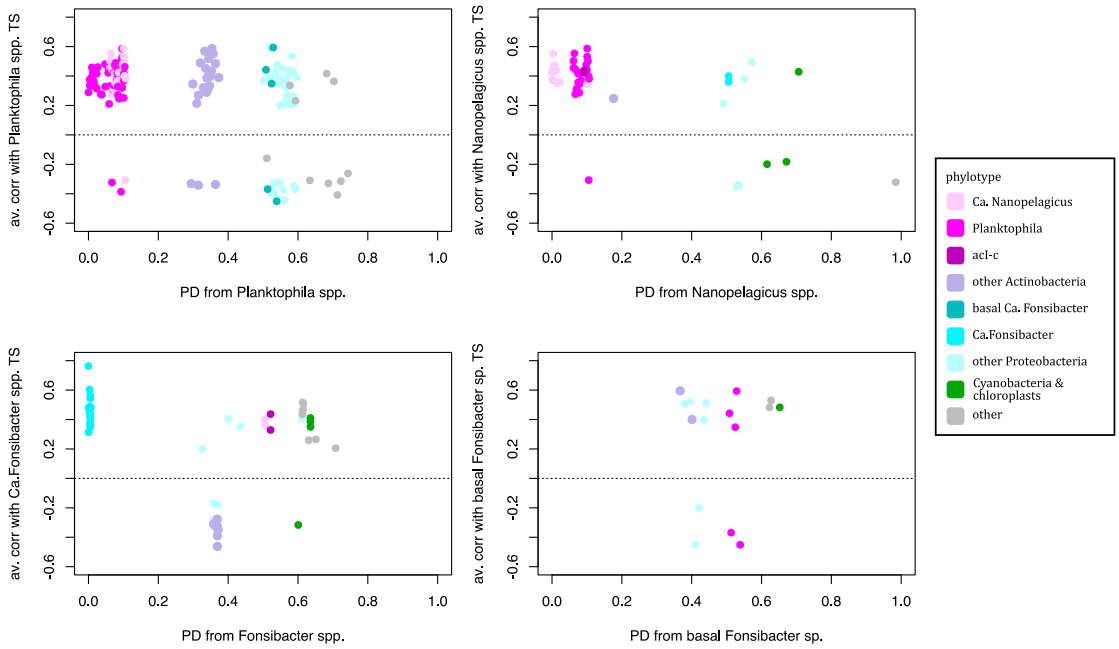
1° cohort



215 *"Ca*. P. lacus*", and "Ca*. N. limnes*"* all have the repair cassette without *katG*. Metagenomic
216 *katG* sequences were recovered matching *"Ca*. P. dulcis*", "Ca*. P. limnetica*", "Ca*. P.
217 versatilis*"*, and *"Ca*. P. vernalis*"*.
218 *"Ca*. Nanopelagicus" 1° cohort in the time-series included other *Actinobacteria*,
219 *Bacteroidetes*, *α*- and *β-Proteobacteria*, and the phytoplanktonic *Synechococcus* and
220 *Stramenopiles* (identified by chloroplast sequences), while in the mixed-cultures there
221 was a subset of these phyla with the addition of one *γ-proteobacterial* phylotype (Fig 2).
222 To calculate whether the presence of an OTU in the mixed-cultures could be attributed to
223 the probability of it being in the inoculum, the total and average relative abundances, and
224 the number of samples an OTU was detected in (prevalence) was compared across the two
225 both datasets with Pearsons correlation. There was no correlation between "*Ca*.
226 Nanopelagicus" OTU relative abundance or prevalence in the lake time-series as
227 compared to the mixed-cultures (Table S4). "*Ca*. Nanopelagicus" was overall more
228 prevalent in mixed cultures than in the lake time-series. "*Ca*. Nanopelagicus" had
229 strong correlation with "Ca. Fonsibacter" in both datasets. There were between 3 and
230 19 other phylotypes (non-*Nanopelgicales*, non-Fonsibacter) detected in cultures with
231 "*Ca*. Nanopelagicus" (Table S5).
232 The most abundant "*Ca*. Nanopelagicus" phylotypes in both datasets were assigned to the
233 *katG*$^{-}$ "*Ca*. Nanopelagicus limnes". Several metagenome-recovered putative "*Ca*.

234 Nanopelagicus" *katG* genes were identified in IMG databases and while these KatGs had
235 greatest homology with *Burkholderia* fold type B, they were not located within the repair
236 cassette and there was no synteny between them. No "*Ca*. Nanopelagicus" katG sequences
237 were recovered from our metagenomes.
238
239 **α-Proteobacterial genus "Ca. Fonsibacter" (LD12)**
240 "*Ca*. Fonsibacter spp." had higher relative abundance from winter to spring and moderate
241 negative correlation with water temperature (r = -0.4, p<0.001, Table S2). "*Ca*.
242 Fonsibacter" was represented by 10 OTUs in each of the time-series and mixed-culture
243 networks, with nine OTUs common to both and the alternating OTUs being the least
244 abundant and least prevalent in both environments. The "*Ca*. Fonsibacter" 1° cohort from
245 the time-series includes other "*Ca*. Fonsibacter", *Bacteroidetes*, *α*- and *β*-*Proteobacteria*,
246 *Actinobacteria* including "*Ca*. Nanopelagicus" and acI-c, and phytoplankton
247 *Synechococcus* and *Stramenopiles*. The "*Ca*. Fonsibacter" 1° cohort from mixed-culture
248 experiments also included other "*Ca*. Fonsibacter" phylotypes, "*Ca*. Nanopelagicus",
249 *Bacteroidetes*, other *α*- and *β*-*Proteobacteria* and macro-algal eukaryotes of the
250 *Streptophyta* order. There was a strong correlation detected between "*Ca*. Fonsibacter"
251 OTU prevalence in the lake time-series compared to the prevalence of the respective OTU
252 in the mixed-cultures (corr = 0.82, n=11, p adj = 0, Table S4). Based on OTU identity,

253 phylogenetic branch placement, and network connections, two "*Ca*. Fonsibacter"
254 phylotypes were designated as basal and analysed separately for 1° cohort. Unlike the core
255 "*Ca*. Fonsibacter" phylotypes, these basal phylotypes had no negative correlations with
256 phytoplankton and no correlations with other "*Ca*. Fonsibacter" phylotypes (Fig. 3).
257



258

259 **Other primary cohort members**

260 Both the *Actinobacterial* 'acIV' (C111) and *Comamonadaceae* families were more
261 abundant in winter when the lake was covered by ice (Fig. S2, KW $p<0.001$, KWmc
262 $p<0.001$).
263

264 **DISCUSSION**

265 **Taxon relative abundance and environmental correlates**

266 If microbial abundance is directly linked to nutrient availability or other physico-chemical
267 parameters, then a strong correlation between geo-chemical parameters and relative
268 abundance is expected. The lack of strong correlation between OTUs or clades and
269 environmental parameters in Lake Erken indicates that during the 8 years of sample
270 collection, microbial community composition was not directly and strongly controlled, but
271 only influenced, by the measured abiotic environmental variables. While short-term studies
272 have shown significant strong trends and correlations (12), such correlations tend to be
273 eclipsed in studies spanning multiple years by shifts operating at longer frequencies and
274 biotic interactions (36–39). Overall, the strength and number (identified via consensus
275 network) of correlations between taxon relative abundance and the weakness of
276 environmental predictors for most of the taxa suggest that biological aspects (including
277 predation, which we did not study) are the greater determinants of microbial population
278 dynamics, composition and persistence in Lake Erken. It stands therefore that the
279 dominance of the three streamlined genera in the lake is attributable more to their
280 interactions with their community than selection by environment.

281

282 **Microbial networks**

283 Assortativity is a predicted product of ecotype or niche differentiation at species level and
284 environmental filtering at higher taxonomic levels (40). Ecotype divergence of strains
285 reduces competition between different members of a population. The detection of highest
286 assortativity in the time-series networks at phylum level supports environmental selection
287 as important in the lake habitat but not at taxonomic levels lower than phyla. It is unlikely
288 that environmental filtering is the main driver behind the success of these three genera in
289 Lake Erken. High assortativity in the mixed cultures at all levels except phylum, supports
290 that niche differentiation contributed to co-occurrence in the cultures. Neither result
291 however fully supports, nor refutes, ecotype theory as the explanation of coexistence or
292 dominance by "*Ca.* Planktophila", "*Ca.* Nanopelagicus", or "*Ca.* Fonsibacter" genera.
293
294 The strong correlation between "*Ca*. Fonsibacter" prevalence in the time-series and mixed-
295 cultures supports that its detection in the cultures may be attributed to being present in the
296 inocula and the filtered lake water media was able to meet an important proportion of its
297 metabolic needs. Conversely, the lack of correlation in either relative abundance or
298 prevalence, for the *Nanopelagicales* genera in the time-series compared to mixed-cultures
299 means that there is not support for their detection being merely due to presence in the
300 inocula. We infer that the co-presence of other phylotypes were bigger contributors to their
301 growth. It should be noted that negative correlations are not detectable in the mixed-

302 cultures as the absence of a phylotype could be stochastic rather than a result of competitive
303 exclusion. Moreover, competition that prevents growth in a culture cannot be identified.
304
305 **Actinobacterial order Nanopelagicales 'acI'**
306 *Nanopelagicales* in Erken were likely more competitive in colder darker oxygen depleted
307 waters. While *Nanopelagicales* has been associated with oxygenated saturated waters in
308 some locations (41) it has been associated with colder deeper water at others (30), though
309 the variation in association could be due to the considerable genetic diversity encompassed
310 in this bacterial order (42). It is also probable that formation of oxygen radicals in surface
311 waters by sunlight inhibit the growth of *katG*$^{-}$ species of *Nanopelagicales*.
312 Of the seven *Nanopelagicales* phylotypes in the mixed-cultures, six were members of the
313 candidate genus "*Ca*. Nanopelagicus" suggesting that this genus dependencies were more
314 easily met compared to the *"Ca.* Planktophila spp.*"* found in Lake Erken. However,
315 genomic evidence so far has shown the opposite as "Ca. Nanopelagicus spp." have more
316 numerous auxotrophies than *"Ca.* Planktophila spp." (23). It is therefore possible that the
317 difference in culturability seen in this experiment was due to reduced costs via gene loss in
318 combination with suitable metabolic partners being in the inoculum. In particular the lower
319 maintenance costs due to the loss of *katG* (25, 33) could increase fitness of the "*Ca*.
320 Nanopelagicus". The successful maintenance of *"Ca.* Planktophila*"* in culture pinpointed
321 the low functionality of some *Nanopelagicales* catalase-peroxidases as the culprit for the

322 difficulties in growing these bacteria. Peroxide is proposed to be an up-regulator of a DNA
323 repair (43) and in the *"Ca.* Planktophila*"* with the putative low functioning KatG type, the
324 gene is embedded in a region of DNA with other genes for repair of damage done to DNA
325 under oxidative stress. *P. limnetica* (acI-A3) genomes have this repair cassette but their
326 *katG* is located elsewhere. The most parsimonious explanation for *Nanopelagicales katG*
327 presence/absence/homology is that the now low functioning version located in a repair
328 cassette is in the process of being lost and that the *"Ca.* Planktophila spp.*"* with a high
329 functioning KatG obtained this version via a HGT event. While *P. vernalis* in Lake Erken
330 is likely $katG^{+}$, it is likely that *P. limnetica* and *N. limnes* from Lake Erken rely on "*Ca*.
331 Fonsibacter" in combination with either *Flavobacterium* or *Comamonadaceae* to detoxify
332 their peroxide. This would fit within the community cooperation model. While *"Ca*.
333 Nanopelagicus" has more numerous auxotrophies compared to *"Ca.* Planktophila*"* (23)
334 the higher culturability of *"Ca*. Nanopelagicus" supports that it may be the cost to
335 habouring *katG*, whether high or low functioning, and as *katG* is larger than average (44),
336 and as long as a functional aquaporin exists (45), and cells don't grow too fast, then it
337 would be an unnecessary cost, as cells can rely on other community members for oxidative
338 stress reduction. Further the reduction in amino-acid synthesis pathways and therefore
339 greater reliance on exogenous sources of amino-acids in "*Ca*. Nanopelagicus" appears to
340 have paid off if their greater cultivability in these minimalist cultures is any indicator. This
341 fits within the community detritus model.

342

343 **α-Proteobacterial genus "Ca. Fonsibacter" (LD12)**

344 Earlier short-term studies have suggested that "*Ca*. Fonsibacter" is highly abundant in Lake
345 Erken in the late summer-early autumn immediately after an algal bloom i.e. the 'clear
346 water' phase of the lake's cycle (7). However, we found support for higher long-term
347 relative abundances in winter and spring and an associated negative correlation with water
348 temperature. These differences in findings could be due to changes in abundance of
349 different ecotypes that are not detectable via 16S surveys. The ecotype differentiation
350 within "*Ca.* Fonsibacter" is centered around the HVR1 which is thought to encode genes
351 for defense, i.e. predator escape via alteration of cellular membranes (3, 46). There appears
352 to be limited intra-population variability in anabolic or catabolic traits (14, 24, 47) and as
353 such, ecotype differentiation for overcoming auxotrophies is not supported. While no direct
354 evidence exists for "*Ca*. Fonsibacter", its similarity to the marine sister clade Pelagibacter
355 suggests that it has overcome auxotrophic limitations by scavenging metabolites and other
356 compounds produced by phototrophs (48). In Erken this would primarily be dying and
357 senescent *Stramenopiles* as reflected in the apparent negative correlation. Evidence points
358 to "*Ca*. Fonsibacter" relying on detritus to meet its dependencies and ecotype
359 differentiation for defense.

360 The combination of network and other correlation analyses with review of published
361 genomes supports that the most of *Nanopelagicales* have anabolic ecotype cooperation

362 with reliance on a detoxifying community cohort as main structuring factors. A few *"Ca.*
363 Planktophila*"* are putatively KatG+ but as they were not detected in Lake Erken it was not
364 possible to determine the most appropriate cooperation models in this study. "*Ca*.
365 Fonsibacter" may escape predation via ecotype differentiation, while broadly relying on
366 community detritus for anabolic deficits.
367

368 **MATERIALS AND METHODS**

369 **Time series water sampling and environmental parameters**

370 Time series water samples were obtained from dimictic Lake Erken (59.1510N, 18.1360E)
371 central Sweden at monthly to weekly intervals spanning a total period from February 2007
372 to November 2015 (Table S6). Samples for genomic and chemical analysis were taken from
373 the 20 m water column at one meter intervals from the deepest point of the lake (Fig. S7).
374 Samples for the total water column were pooled during periods of mixing. Samples of the
375 upper oxygenated water column (epilimnion), middle (metalimnion), and lower oxygen
376 depleted (hypolimnion) were pooled into separate composite samples during summer
377 stratification. Annual dates of mixing, stratification, ice-over, and ice-thaw varied, as did
378 depth of stratigraphic layers (Table S6). Temperature and oxygen concentrations were
379 analyzed every meter with a portable Oxi 340i oxygen meter equipped with a Cellox 325-
380 20WTW probe and were used to determine stratification. Chemical analysis of pooled
381 water samples was carried out by the Lake Erken Field Station ISO certified laboratory and

382 included pH, conductivity, nitrate, nitrite, alkalinity, turbidity, suspended matter,
383 phosphate, ammonium, total and particulate phosphorus and nitrogen, chlorophyll-a,
384 suspended particulate organic matter, silicate, water-colour and absorbance at 420nm
385 measurements. Water samples for genomic analysis were collected by gentle vacuum
386 filtration onto 0.2 µm membrane filters (Supor-200 Membrane Disc Filters, 47mm; Pall
387 Corporation, East Hills, NY, USA). Filters were individually stored in Eppendorfs and
388 transported back to the University Uppsala laboratory where they were stored at -80 °C
389 until further processing.
390
391 **DNA extraction and SSU rRNA gene amplicon preparation**
392 DNA was extracted from filters using MoBio Ultra-clean soil DNA extraction kits (MoBio,
393 Carlsbad) as per manufacturers' instructions. The V3-V4 region of the 16S rRNA gene was
394 amplified using the S-D-Bact-0341-b-S-17 (Bakt_341F: CCTACGGGNGGCWGCAG)
395 forward primer and S-D-Bact-0785-b-A-21 (805RN: GACTACNVGGGTATCTAATCC)
396 reverse primer (49). Template DNA was amplified in duplicate 20 µl reactions containing
397 1.0 U Q5 high fidelity DNA polymerase (NEB, UK), 0.25 µM primers, 200 µM dNTP mix,
398 and 0.4 µg BSA. The thermocycler program was an initial denaturation step at 95 °C for
399 30 s, followed by 20 cycles of dissociation at 95 °C for 10 s, annealing at 53 °C for 30 s
400 extension at 72 °C for 30 s, with a final extension of 2 min at 72 °C. Amplicons were pooled
401 and purified with Agencourt AMPure XP purification system (Beckman Coulter, Danvers,

402 MA, USA). Purified amplicons (2 µl) were amplified in second step 20 µl reactions to
403 introduce sample MIDS to each end of the amplicons. Reactions contained 1.0 U Q5 high
404 fidelity DNA polymerase (NEB, UK), 0.25 µM primers, 200 µM dNTP mix, and 0.4 µg
405 BSA. Thermocycler program was an initial denaturation step at 95 °C for 30 s, followed
406 by 15 cycles of dissociation at 95 °C for 10 s, annealing at 66 °C for 30 s, extension at 72
407 °C for 30 s, with a final extension of 2 min at 72 °C. Amplicons were again purified using
408 Agencourt AMPure XP purification system (Beckman Coulter, Danvers, MA, USA) then
409 quantified using PicoGreen (Invitrogen) before pooling at equi-molar amounts for each
410 run. Amplicons were sequenced at SciLifeLab SNP/SEQ service on the MiSeq Illumina
411 platform with 300bp paired-end libraries. MiSeq data was processed, including de-
412 multiplexing, by the sequencing centre using Illumina pipelines where all reads with more
413 than 8% mismatch to adapter-MID sequences were discarded.
414
415 **Water sampling, culture conditions, and DNA amplification for mixed-cultures**
416 Water as both a source of growth-media and culturable-cells was collected from the
417 epilimnion of Lake Erken on March 8th 2016. Water for growth-media was filter-sterilised
418 twice through Sterivex filter cartridges 0.22 µm (Millipore). Ten minutes exposure to UV
419 light further disinfected and disrupted cellular and viral nucleic acids. Ten microliters of
420 untreated lake water (10^6 cells/ml) were used as a source of cells for inoculation of mixed-
421 cultures. The untreated lake water was diluted in 1 L of the triple sterilized lake water

422 (approximately 10 cells/ml). Ninety-eight 1 ml cultures were placed in sterile 50 ml falcon
423 tubes to allow oxygenation and incubated for two months under 12-hour light-dark cycles
424 at 11 °C to replicate lake conditions at time of collection. The V3-V4 region of the 16S
425 rRNA gene was amplified using the S-D-Bact-0341-b-S-17 (Bakt_341F:
426 CCTACGGGNGGCWGCAG) forward primer and S-D-Bact-0785-b-A-21 (805RN:
427 GACTACNVGGGTATCTAATCC) reverse primer (49) One microliter of the mixed-
428 culture after one freeze-thaw cycle was used as template and Q5® High-Fidelity DNA
429 Polymerase. The PCR conditions consisted of an initial step at 98°C for 10 minutes.
430 Thereafter, 35 cycles of denaturation at 98°C for 10 s, annealing for 30 s and extension at
431 72°C for 30 s. Annealing temperature was 48°C which has been tested to get unbiased
432 products of non-mismatch and 3-mismatch isolates. Final extension was performed at 72°C
433 for 2 minutes. PCR products were purified with magnetic beads (Angecourt AMPure). A
434 second PCR was conducted for attaching standard illumina handles and index primers as
435 described above. Primers and MIDS detailed in Table S7. Prior to sequencing a portion of
436 amplification product was run on gel to check for amplification and used as proxy for
437 growth of culture. Only data from the 60 cultures with a visible band after the first PCR
438 were included in analyses.
439
440 **Amplicon bioinformatics**

441 Demultiplexed MiSeq time-series and mixed-culture data, and 454 Ti time-series data
442 obtained from Lake Erken 2008 (11), were pre-processed separately and joined once
443 chimeras had been removed. Pre-processing of MiSeq data started with SeqPrep v1.2
444 (https://github.com/jstjohn/SeqPrep.git) with default settings to join paired end reads while
445 discarding reads shorter than 30b and all un-paired reads. 454 and MiSeq reads with
446 ambiguous bases, low quality average, and low quality sections, were discarded and
447 trimmed with Qiime v1.9 (50) split_libraries and multiple split libraries fastq respectively.
448 Singletons and sequences identified as chimeric were removed with Qiimes'
449 identify_chimeric_seqs, filter_fasta using usearch v6.1 (51), with default settings and
450 option parameter 'intersection' to select out sequences identified by both de-novo and
451 reference-based methods. Both GreenGenes (v105 13-8) and SILVA (v128) reference
452 datasets were used (52, 53). Sequences from the two Miseq runs and the 454 data from
453 2008 were then combined. Reads were clustered at 97% identity into OTUs then assigned
454 taxonomy based on SILVA v128 using QIIMES closed pick OTU method (uses only
455 template recognition and does not allow for de-novo sequences) to allow for differences in
456 DNA processing and sequencing across the three datasets. Sequences from the negative
457 control were removed from the entire dataset. Samples with potential human-microbiome
458 contamination (Table S8) were identified and discarded from the dataset due to potential
459 effect on community metabolism. Sequences identified as eukaryotic (including fungal,
460 mitochondrial and chloroplast origin) or archaeal were deliberately retained despite the

461 primers not specifically being selected to amplify such sequences. as they were considered
462 critical in the mixed-culture analyses from a community metabolism perspective. Samples
463 were normalized using Qiimes single rarefraction to even depth of 1500 reads per sample
464 for the time-series and 500 reads for the mixed-culture samples. These normalised OTU
465 tables were then used for all downstream analyses.
466
467

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

469 Statistical analyses were done using R v3.3.1 (54) in the RStudio IDE v0.99.903 (55) and
470 graphs processed for publication in Inkscape v0.91. Environmental parameters were
471 removed if more than 40% of the values were missing then remaining twenty-one tested
472 for co-linearity with the usdm v1.1.18 package with vifstep followed by vifcor functions
473 and removed if VIF >10 or correlations were over 0.7 sequentially starting with those with
474 highest number of missing values. The remaining 14 environmental predictors were tested
475 for statistical differences in connection to lake-cycle with Kruskal-Wallis (kruskal.test of
476 the base stats package) and graphed using beanplot v1.2 (56) and scales v0.4.1. Statistical
477 significance of differences between individual paired cycle points were determined post-
478 hoc with the Kruskal-Wallis multiple comparison (KWmc) using thekruskal.mc test of the
479 pgirmess package v1.6.4 (57). Environmental parameter correlates with selected taxa
480 relative abundances were screened using Pearsons correlation with an alpha level of 0.001
481 and p-values estimated and corrected for multiple comparisons with Benjamini, Hochberg,

482 and Yekutieli (BH) false discovery rate using the psyche package v1.6.12 corr.test function
483 and Pearsons moment r reported. Parameters with weak to moderate correlation (r >|0.3|)
484 were plotted in R to observe monotonicity.
485
486 **Network analyses**
487 The normalized time-series OTU table was further processed for network and correlation
488 analyses to remove rare OTUs that, while potentially important within the community, are
489 known to inhibit identification of ecological relationships between OTUs and also between
490 OTUs and their environment. Recommendations include OTU tables with 50% or less
491 sparcity and n_{eff} (inverse Simpsons) above an average of 10 and the use of multiple methods
492 (58). A range of filtering parameters based on frequency of OTU detection, from 2 samples
493 up to 46 (20% of samples) and total relative abundance ranging from 1 up to 150 (0.1%),
494 was used on the OTU table to remove type I errors in detection of co-occurrence networks
495 (58). The best filter combination was found with OTUs at a total relative abundance above
496 0.1% and occurring in at least 29 samples. All OTU tables were analysed in R (54) using
497 the vegan v2.4-1 package (59) diversity function to find the n_{eff} and the base stats v3.3.1
498 package to find min, max and mean of the n_{eff}. The reduced OTU table was analysed with
499 SparCC (60), Pearsons correlation coefficient, and SPIEC-EASI (61). SparCC analyses
500 were carried out in Rstudio (55) using both the rsparcc and spiec.easi packages with rsparcc
501 parameters: max.iter=100, th=0.1, exiter=10 and pseudo p-values generated with

502 spiec.easis sparccboot (100 bootstraps) and pval.sparccboot scripts (two-sided test).
503 Pearson correlations were performed using the corr.test of psyche package using the
504 optional p-values adjustment for multiple testing of Benjamini, Hochberg and Yekutieli
505 false discovery rate. SPIEC-EASI correlations were analysed using the spiec.easi package
506 with parameters: method="mb", sel.criterion="stars", lambda.min.ratio=1e-2,
507 nlambda=20, and 50 repetitions. Singletons were removed from the normalized mixed-
508 culture OTU table prior to network analysis. The reduced OTU table was analysed with
509 Pearsons correlation, SpiecEasi (61) and on the binary version of the table, Dice-Sørensen
510 index (62, 63).
511 Correlations between OTUs in the time-series and mixed cultures were retained if the
512 correlation co-efficient was greater than 0.3 absolute for Pearsons, SparCC, or Dice-
513 Sørensen, or if detected at all by SPIEC-EASI. Time-series and culture networks consisting
514 of OTU connections detected by at least two correlation methods were visualized in
515 Cytoscape. All further statistical analyses of the networks were performed in Cytoscape
516 except for assortativity analyses which were done in R using the igraph v1.1.2 package
517 nominal assortativity function. Sub-networks of the identified dominant genera ("*Ca.*
518 Planktophila", "*Ca*. Nanopelagicus", and "*Ca*. Fonsibacter") and their primary cohorts
519 were extracted from the two initial networks and visualized separately. Layout for the
520 networks were "edge-weighted spring-embedded" with edge weights calculated from the
521 average of the correlations. Correlations between *Nanopelagicales* OTU and "*Ca*.

522 Fonsibacter" OTU detection in the time-series versus the mixed cultures were investigated
523 by Pearson correlation of their relative abundance (log av.ra), prevalence (percent samples
524 or cultures detected in) and relative abundance versus prevalence to determine if detection
525 in culture was correlated to occurrence in the lake.
526
527 **Selection of samples for metagenomic sequencing**
528 Samples for metagenomic sequencing were selected by highest relative abundance of
529 clades of interest according to amplicon data, greatest available DNA mass, and greatest
530 variation in environmental parameters. This was done to examine samples with the greatest
531 potential divergence of environmental conditions under which microbes of interest were
532 most abundant. These included samples taken during lake circulation and summer
533 stratification which separately including hypo, meta and epi-limnionic samples. Two
534 samples per category were selected from different years. Lower biomass during ice cover
535 meant there was insufficient DNA for metagenomic sequencing.
536
537 **Metagenome bioinformatics**
538 HiSeq Illumina paired end reads were pre-processed first using SeqPrep v1.2 (64) to filter
539 out low quality reads using default settings resulting in removal of all reads less than 30b
540 in length, un-paired reads and initial clipping of adapter sequences. Next sequences were
541 cleaned using Trimmomatic v0.36 (65) with removal of any remaining Illumina adapters

542 allowing for 2 mismatches; followed by 4b sliding window trim with minimum average
543 quality of 15; then trimming of read ends with 'maxinfo' value less than 1. The eight
544 metagenomes were individually assembled with megahit (66) using iterative kmers from
545 21 to 121 at 10b intervals, minimum coverage of 2, and minimum length of 200.
546 Concatenated assemblies were cleaned of Illumina artifact low complexity (long
547 homopolymer) contigs with prinseq-lite v0.20.4 (67) removing contigs consisting of
548 greater than 80% one nucleotide.
549
550 **Investigation of katG gene and its product: catalase-peroxidase I**
551 Known impediments to axenic growth of freshwater isolates are multiple auxotrophies, and
552 slow growth and replication rates. The successful axenic cultivation of "*Ca*. Planktophila"
553 strains by addition of high functioning catalase to reduce oxidative stress, in addition to
554 catering for predicted auxotrophies, showed that there can be other considerations. To
555 investigate how both "*Ca*. Planktophila" and Nanopelagicus phylotypes grow in Lake
556 Erken and why Nanopelagicus grew more frequently in the minimalist mixed cultures, the
557 katG gene was investigated. First katG amino-acid sequences for species identified as
558 belonging to the Nanopelagicales and its primary cohort were retrieved from databases
559 along with those for whom the gene product has been assayed and/or the 3D crystal
560 structure is known. Next these sequences were used as BLAST (68) queries to retrieve
561 closely related (> 80% aa identity) sequences from Lake Erken metagenomes. The

562 collected amino-acid sequences were aligned using MAFFT e-insi and "leavegappy"
563 options (69). Sequences with greater than 99% identity were identified using CD-Hit and
564 removed (70) while sequences from uniRef useful for constructing a guide tree were then
565 also added (71, 72). The alignment was evaluated for best amino-acid substitution model
566 using IQ-tree v.1.6.12 modelfinder (73). This alignment was then analysed and used to
567 construct a tree in RAXML v.0.9.0 (74) with WAG+IG+R4 model and 1000 bootstraps and
568 catalase-peroxidase from *Oryza rufipogon* set as outgroup. To examine functionality
569 secondary structure was predicted and aligned to known models then tertiary structure
570 predicted in Phyre^2 (75). Further the location of katG genes in Nanopelagicales genomes
571 and MAGs was examined in IMG using the gene-neighbourhood function. Images were
572 downloaded with screen shot and processed in Inkscape.
573

574 **Data availability**

575 Time-series MiSeq sequences were deposited in ENA
576 (https://www.ebi.ac.uk/ena/browser/view/PRJEB20109) under accession numbers:
577 ERS1884053 to ERS1884254, sample details including MIDS are also in Table S6. Mixed-
578 culture sequences are available from ENA PRJEB35605
579 (https://www.ebi.ac.uk/ena/browser/view/PRJEB35605) with details also in Table S7. 454
580 data used in the time-series is available at SRA (https://trace.ncbi.nlm.nih.gov/Traces/sra/)
581 under accession SRR097418 (11). Assembled metagenomes are available at IMG

582 (https://img.jgi.doe.gov/cgi-bin/m/main.cgi) with TOIDs: 3300020141, 3300020151,
583 3300020159, 3300020160, 3300020161, 3300020162, 3300020172, & 3300020205.
584 Rscripts used for correlation and network analyses are available from
585 https://github.com/rmondav/publications.
586

587 **ACKNOWLEDGEMENTS**

588 RM, sequencing of the metagenomes, and publication supported by grants from the
589 Malméns Stiftelsen. SG was supported by VR (grant 2012-4592), Olsson-Borgh, Knut and
590 Alice Wallenberg Foundation (grant KAW 2013.0091) and SciLifeLab Fellowship. SB was
591 supported by grants from the Swedish Research Council (VR) and the Swedish Research
592 Council Formas. The sequencing of the Lake Erken 16S rRNA time-series dataset and
593 chemical data was supported by the Swedish Infrastructure for Ecosystem Science (SITES)
594 with technical assistance from Dr. Omneya Ahmed, Pilar López Hernández, Helena
595 Enderskog, Erika Bridell and Kristiina Mustonen. Sequencing of the mixed cultures was
596 supported by a collaborative (grant from the K&A2012-4592), Olsson-Borgh, Knut and
597 Alice Wallenberg foundation (PI. Lars Tranvik). Publicity was supported by Kungl.
598 Vetenskapsakademiens stiftelser (BS2017-0044). The funders had no role in study design,
599 data collection and interpretation, or the decision to submit the work for publication. We
600 gratefully acknowledge the computing resources provided by SNIC through Uppsala
601 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 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.

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

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

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

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

9. **Lymer D**, **Lindström ES**. 2010. Changing phosphorus concentration and subsequent prophage induction alter composition of a freshwater viral assemblage. Freshw. Biol. **55**:1984–1996.

10. **Burgmer T**, **Reiss J**, **Wickham S**, **Hillebrand H**. 2010. Effects of snail grazers and light on the benthic microbial food web in periphyton communities. Aquat. Microb. Ecol. **61**:163–178.

642 11. **Eiler A**, **Heinrich F**, **Bertilsson S**. 2012. Coherent dynamics and association
643 networks among lake bacterioplankton taxa. ISME J. **6**:330–42.
644 12. **Heinrich F**, **Eiler A**, **Bertilsson S**. 2013. Seasonality and environmental control of
645 freshwater SAR11 (LD12) in a temperate lake (Lake Erken, Sweden). Aquat.
646 Microb. Ecol. **70**:33–44.
647 13. **Eiler A**, **Bertilsson S**. 2004. Composition of freshwater bacterial communities
648 associated with cyanobacterial blooms in four Swedish lakes. Environ. Microbiol.
649 **6**:1228–1243.
650 14. **Henson MW**, **Lanclos VC**, **Faircloth BC**, **Thrash JC**. 2018. Cultivation and
651 genomics of the first freshwater SAR11 (LD12) isolate. ISME J. **12**:1846–1860.
652 15. **Newton RJ**, **Jones SE**, **Eiler A**, **McMahon KD**, **Bertilsson S**. 2011. A guide to the
653 natural history of freshwater lake bacteria.Microbiology and molecular biology
654 reviews : MMBR.
655 16. **Warnecke F**, **Amann R**, **Pernthaler J**. 2004. Actinobacterial 16S rRNA genes from
656 freshwater habitats cluster in four distinct lineages. Environ. Microbiol.
657 17. **Warnecke F**, **Sommaruga R**, **Sekar R**, **Hofer JS**, **Pernthaler J**. 2005.
658 Abundances, identity, and growth state of actinobacteria in mountain lakes of
659 different UV transparency. Appl. Environ. Microbiol.
660 18. **Glockner FO**, **Zaichikov E**, **Belkova N**, **Denissova L**, **Pernthaler J**, **Pernthaler**
661 **A**, **Amann R**. 2000. Comparative 16S rRNA analysis of lake bacterioplankton

662 reveals globally distributed phylogenetic clusters including an abundant group of
663 actinobacteria. Appl. Environ. Microbiol.
664 19. **Li J**, **Zhang J**, **Liu L**, **Fan Y**, **Li L**, **Yang Y**, **Lu Z**, **Zhang X**. 2015. Annual
665 periodicity in planktonic bacterial and archaeal community composition of eutrophic
666 Lake Taihu. Sci. Rep. **5**:1–14.
667 20. **Comte J**, **Lovejoy C**, **Crevecoeur S**, **Vincent WF**. 2016. Co-occurrence patterns in
668 aquatic bacterial communities across changing permafrost landscapes.
669 Biogeosciences **13**:175–190.
670 21. **Garcia SL**, **Buck M**, **Hamilton JJ**, **Wurzbacher C**, **Grossart H-P**, **McMahon**
671 **KD**, **Eiler A**. 2018. Model Communities Hint at Promiscuous Metabolic Linkages
672 between Ubiquitous Free-Living Freshwater Bacteria. mSphere **3**:1–8.
673 22. **Eckert EM**, **Baumgartner M**, **Huber IM**, **Pernthaler J**. 2013. Grazing resistant
674 freshwater bacteria profit from chitin and cell-wall-derived organic carbon. Environ.
675 Microbiol. **15**:2019–2030.
676 23. **Neuenschwander SM**, **Ghai R**, **Pernthaler J**, **Salcher MM**. 2017.
677 Microdiversification in genome-streamlined ubiquitous freshwater Actinobacteria.
678 ISME J. **12**:185–198.
679 24. **Eiler A**, **Mondav R**, **Sinclair L**, **Fernandez-Vidal L**, **Scofield DDG**, **Scwientek P**,
680 **Martinez-Garcia M**, **Torrents D**, **McMahon KDKD**, **Andersson SGESGE**,
681 **Stepanauskas R**, **Woyke T**, **Bertilsson S**, **Schwientek P**, **Martinez-Garcia M**,

682 **Torrents D**, **McMahon KDKD**, **Andersson SGESGE**, **Stepanauskas R**, **Woyke**
683 **T**, **Bertilsson S**. 2015. Tuning fresh: radiation through rewiring of central
684 metabolism in streamlined bacteria. ISME J. **10**:1–13.
685 25. **Kim S**, **Kang I**, **Seo J-HH**, **Cho J-CC**. 2019. Culturing the ubiquitous freshwater
686 actinobacterial acI lineage by supplying a biochemical 'helper' catalase. ISME J.
687 **112**:343640.
688 26. **Rodriguez-R L**, **Tsementzi D**, **Luo C**, **Konstantinidis K**. 2019. Iterative
689 Subtractive Binning of Freshwater Chronoseries Metagenomes Identifies of over
690 Four Hundred Novel Species and their Ecologic Preferences. bioRxiv 826941.
691 27. **Giovannoni SJ**, **Cameron Thrash J**, **Temperton B**. 2014. Implications of
692 streamlining theory for microbial ecology. ISME J. Nature Publishing Group.
693 28. **Tripp HJ**. 2013. The unique metabolism of SAR11 aquatic bacteria. J. Microbiol.
694 Seoul Korea **51**:147–53.
695 29. **Pernthaler J**, **Posch T**, **Šimek K**, **Vrba J**, **Pernthaler A**, **Glöckner FO**, **Nübel U**,
696 **Psenner R**, **Amann R**. 2001. Predator-Specific Enrichment of Actinobacteria from
697 a Cosmopolitan Freshwater Clade in Mixed Continuous Culture. Appl. Environ.
698 Microbiol.
699 30. **Salcher MM**. 2014. Same same but different: Ecological niche partitioning of
700 planktonic freshwater prokaryotes. J. Limnol. **73**:74–87.
701 31. **Dadon-Pilosof A**, **Conley KR**, **Jacobi Y**, **Haber M**, **Lombard F**, **Sutherland KR**,

Steindler L, **Tikochinski Y**, **Richter M**, **Oliver Glamp F**, **Suzuki MT**, **West NJ**, **Genin A**, **Yahel G**. 2017. Surface properties of SAR11 bacteria facilitate grazing avoidance. Nat. Microbiol. 1.

32. **Våge S**, **Storesund JE**, **Thingstad TF**. 2013. SAR11 viruses and defensive host strains. Nature **499**:E3-4.

33. **Morris JJ**, **Lenski RE**, **Zinser ER**. 2012. The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss. MBio.

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

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

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

37. **Brown JH**, **Whitham TG**, **Morgan Ernest SK**, **Gehring CA**. 2001. Complex

722 species interactions and the dynamics of ecological systems: Long-term
723 experiments. Science (80-.). American Association for the Advancement of
724 Science.
725 38. **Ducklow HW**, **Doney SC**, **Steinberg DK**. 2009. Contributions of Long-Term
726 Research and Time-Series Observations to Marine Ecology and Biogeochemistry.
727 Ann. Rev. Mar. Sci. **1**:279–302.
728 39. **Franklin JF**. 1989. Importance and Justification of Long-Term Studies in Ecology,
729 p. 3–19. *In* Long-Term Studies in Ecology. Springer New York.
730 40. **Philippot L**, **Andersson SGE**, **Battin TJ**, **Prosser JI**, **Schimel JP**, **Whitman WB**,
731 **Hallin S**. 2010. The ecological coherence of high bacterial taxonomic ranks. Nat.
732 Rev. Microbiol. **8**:523–9.
733 41. **Garcia SL**, **Salka I**, **Grossart HP**, **Warnecke F**. 2013. Depth-discrete profiles of
734 bacterial communities reveal pronounced spatio-temporal dynamics related to lake
735 stratification. Environ. Microbiol. Rep. **5**:549–555.
736 42. **Garcia SL**, **Stevens SLR**, **Crary B**, **Martinez-Garcia M**, **Stepanauskas R**,
737 **Woyke T**, **Tringe SG**, **Andersson SGE**, **Bertilsson S**, **Malmstrom RR**, **McMahon**
738 **KD**. 2017. Contrasting patterns of genome-level diversity across distinct co-
739 occurring bacterial populations. ISME J.
740 43. **D'Autréaux B**, **Toledano MB**. 2007. ROS as signalling molecules: Mechanisms
741 that generate specificity in ROS homeostasis. Nat. Rev. Mol. Cell Biol. Nature

742 Publishing Group.

743 44. **Koonin E V**, **Wolf YI**. 2008. Genomics of bacteria and archaea: the emerging
744 dynamic view of the prokaryotic world. Nucleic Acids Res.2008/10/25. **36**:6688–
745 6719.

746 45. **Bienert GP**, **Chaumont F**. 2014. Aquaporin-facilitated transmembrane diffusion of
747 hydrogen peroxide. Biochim. Biophys. Acta - Gen. Subj. **1840**:1596–1604.

748 46. **Zaremba-Niedzwiedzka K**, **Viklund J**, **Zhao W**, **Ast J**, **Sczyrba A**, **Woyke T**,
749 **McMahon K**, **Bertilsson S**, **Stepanauskas R**, **Andersson SGE**. 2013. Single-cell
750 genomics reveal low recombination frequencies in freshwater bacteria of the SAR11
751 clade. Genome Biol. **14**:R130.

752 47. **Tsementzi D**, **Rodriguez-R LM**, **Ruiz-Perez CA**, **Meziti A**, **Hatt JK**,
753 **Konstantinidis KT**. 2019. Ecogenomic characterization of widespread, closely-
754 related SAR11 clades of the freshwater genus "Candidatus Fonsibacter" and
755 proposal of Ca. Fonsibacter lacus sp. nov. Syst. Appl. Microbiol. **42**:495–505.

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

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

762 50. **Caporaso JG**, **Kuczynski J**, **Stombaugh J**, **Bittinger K**, **Bushman FD**, **Costello**
763 **EK**, **Fierer N**, **Peña AG**, **Goodrich JK**, **Gordon JI**, **Huttley GA**, **Kelley ST**,
764 **Knights D**, **Koenig JE**, **Ley RE**, **Lozupone CA**, **McDonald D**, **Muegge BD**,
765 **Pirrung M**, **Reeder J**, **Sevinsky JR**, **Turnbaugh PJ**, **Walters WA**, **Widmann J**,
766 **Yatsunenko T**, **Zaneveld J**, **Knight R**. 2010. QIIME allows analysis of high-
767 throughput community sequencing data. Nat. Methods **7**:335–336.
768 51. **Edgar RC**. 2010. Search and clustering orders of magnitude faster than BLAST.
769 Bioinformatics **26**:2460–2461.
770 52. **McDonald D**, **Price MN**, **Goodrich J**, **Nawrocki EP**, **DeSantis TZ**, **Probst A**,
771 **Andersen GL**, **Knight R**, **Hugenholtz P**. 2012. An improved Greengenes
772 taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria
773 and archaea. ISME J. **6**:610–8.
774 53. **Yilmaz P**, **Parfrey LW**, **Yarza P**, **Gerken J**, **Pruesse E**, **Quast C**, **Schweer T**,
775 **Peplies J**, **Ludwig W**, **Glöckner FO**. 2014. The SILVA and "all-species Living Tree
776 Project (LTP)" taxonomic frameworks. Nucleic Acids Res. **42**:643–648.
777 54. **R-Core-team**. 2011. R: A Language and Environment for Statistical Computing.
778 2.14.1. R Foundation for Statistical Computing, Vienna, Austria.
779 55. **RStudio**. 2012. RStudio: Integrated development environment for R. 0.97.248.
780 Boston, MA.
781 56. **Kampstra P**. 2008. Beanplot: A Boxplot Alternative for Visual Comparison of

782 Distributions. J. Stat. Softw. **28**:1–9.

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

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

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

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

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

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

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

801 64. 2014. SeqPrep. 1.2.

802 65. **Bolger AM**, **Lohse M**, **Usadel B**. 2014. Trimmomatic: A flexible trimmer for
803 Illumina sequence data. Bioinformatics **30**:2114–2120.
804 66. **Li D**, **Liu CM**, **Luo R**, **Sadakane K**, **Lam TW**. 2014. MEGAHIT: An ultra-fast
805 single-node solution for large and complex metagenomics assembly via succinct de
806 Bruijn graph. Bioinformatics **31**:1674–1676.
807 67. **Schmieder R**, **Edwards R**. 2011. Quality control and preprocessing of
808 metagenomic datasets. Bioinformatics **27**:863–864.
809 68. **Camacho C**, **Coulouris G**, **Avagyan V**, **Ma N**, **Papadopoulos J**, **Bealer K**,
810 **Madden TL**. 2009. BLAST+: architecture and applications. BMC Bioinformatics
811 **10**:421.
812 69. **Katoh K**, **Kuma KI**, **Toh H**, **Miyata T**. 2005. MAFFT version 5: Improvement in
813 accuracy of multiple sequence alignment. Nucleic Acids Res. **33**:511–518.
814 70. **Niu B**, **Fu L**, **Sun S**, **Li W**. 2010. Artificial and natural duplicates in pyrosequencing
815 reads of metagenomic data. BMC Bioinformatics **11**:187.
816 71. **Suzek BE**, **Huang H**, **McGarvey P**, **Mazumder R**, **Wu CH**. 2007. UniRef:
817 comprehensive and non-redundant UniProt reference clusters. Bioinformatics
818 **23**:1282–8.
819 72. **Magrane M**, **Consortium U**. 2011. UniProt Knowledgebase: a hub of integrated
820 protein data. Database (Oxford). **2011**:bar009.
821 73. **Kalyaanamoorthy S**, **Minh BQ**, **Wong TKF**, **Von Haeseler A**, **Jermiin LS**. 2017.

822 ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat.
823 Methods **14**:587–589.
824 74. **Kozlov AM**, **Darriba D**, **Flouri T**, **Morel B**, **Stamatakis A**. 2019. RAxML-NG: a
825 fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference.
826 Bioinformatics.
827 75. **Kelley LA**, **Mezulis S**, **Yates CM**, **Wass MN**, **Sternberg MJE**. 2015. The Phyre2
828 web portal for protein modeling, prediction and analysis. Nat. Protoc. **10**:845–858.
829
830
831
832 **FIGURE LEGENDS**
833 Fig. 1. Cooperation models shown in a lake setting. Ecotype Cooperation Model describes
834 anabolic variability as the main contributor to co-existence while allowing for
835 recombination and swapping of metabolic modules across strains. Community Cooperation
836 Model also encompasses anabolic complementarity but between rather than within species
837 and includes common goods ranging from expensive to costless. The Community Detrital
838 Model places complementarity both between and within species but describes catabolic
839 variation.
840

841 Fig. 2. OTU networks of the primary cohorts from amplicons of time-series (upper) and
842 mixed-cultures (lower). Nodes sized by average relative abundance in time-series and by
843 percentage of cultures detected in for mixed-cultures. Pink dotted lines indicate mutual
844 exclusion of OTUs while grey solid lines are co-occurrence.
845
846 Fig. 3. The relationship between OTU phylogenetic distance and correlation of the primary
847 cohort phylotypes to selected genera: *"Ca. Planktophila"* upper-left, "*Ca*. Nanopelagicus"
848 upper-right, "*Ca*. Fonsibacter" lower-left, and basal "*Ca*. Fonsibacter "lower-right.