Title: Model communities hint to promiscuous metabolic linkages between ubiquitous free-living freshwater bacteria

Sarahi L Garcia[a*], Moritz Buck[a], Joshua J Hamilton[b], Christian Wurzbacher[c], Magnus Alm Rosenblad[d], Katherine D McMahon[b,e], Hans-Peter Grossart[f,g], Falk Warnecke[h,+], Alexander Eiler[a]

a Department of Ecology and Genetics, Limnology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden
b Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA
c Department of Biological and Environmental Sciences, University of Gothenburg
d Deptartment of Marine Sciences, National Infrastructure of Bioinformatics, (NBIS), University of Gothenburg, Lundberg laboratory, Gothenburg, Sweden
e Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI, USA
f Leibniz-Institute for Freshwater Ecology and Inland Fisheries, Department Experimental Limnology, Stechlin, Germany
g Institute for Biochemistry and Biology, Potsdam University, Potsdam, Germany
h Jena School for Microbial Communication (JSMC) and Microbial Ecology Group at Friedrich Schiller University Jena, Jena, Germany
+ Deceased
* For correspondence and requests for materials E-mail S.L.G. (email: sarilog@gmail.com)

Key words: mixed-cultures / metagenomics / Actinobacteria / community / interactions

Abstract

Free-living microorganisms with streamlined genomes are very abundant in the environment. Genome streamlining results in losses in the cell's biosynthetic potential generating physiological dependencies between microorganisms. However, there exists no consensus on the specificity of these microbial associations. To verify specificity and extent of these associations, mixed cultures were established from three different freshwater environments. These cultures contained free-living streamlined organisms lacking multiple biosynthetic pathways. Among the co-occurring members of the mixed cultures, there was no clear recurring pattern of metabolic complementarity and dependencies. This, together with weak temporal co-occurrence patterns observed using time-series metagenomics, suggests that free-living freshwater bacteria form loose and unspecific cooperative loops. Comparative genomics suggests that the proportion of accessory genes in populations of streamlined bacteria allows for flexibility in interaction partners. Altogether this renders these free-living bacterial lineages functionally versatile despite their streamlining tendencies.

Introduction

Microorganisms can cooperate in many different ways and their relationships range from facultative to obligate dependencies [1]. On the far end of the dependency spectrum are endosymbionts (and endoparasites). These show obligate dependencies where genome reduction and associated loss of essential biosynthesis pathways are widespread. In most cases, a defined host - endosymbiont specificity is established. At the other end of the spectrum, free-living bacteria are considered to be widely autonomous. However, this image is now changing rapidly with frequent reports of reduced microbial genomes in the environment resulting in unique and singular auxotrophies [2].

In natural aquatic environments where nutrients generally occur in low concentrations, microbes produce many compounds that are costly, but promote survival and reproduction not only for themselves but also for neighboring cells [3]. In fact, auxotrophy for amino acids and vitamins has recently been shown for numerous free-living bacteria [4-6], pointing to critically important metabolic dependencies from other community members as exemplified in a simple model community [7]. Therefore, it would not be surprising that most natural aquatic

53 systems allow for frequent and complex metabolic interactions, for example, via continuous
54 mixing of the environment and molecular-scale diffusion facilitating distribution of public
55 goods. Hence, even when abundant environmental bacteria are seen as free-living, their
56 streamlined genomes render them tightly linked and dependent on other microorganisms [8] in
57 the community. The sheer number of bacteria and hence genetic variability in a population
58 could allow for high metabolic flexibility at the population level with an endless number of
59 metabolic interactions possible among free-living bacterial communities. However, the
60 character of these associations still remains largely unknown.
61 To investigate the specificity of the dependencies between free-living freshwater
62 microorganisms, we established dilution mixed cultures as model communities. These
63 cultures are a small subsample of the complex natural community [9] with about 12 initial cells.
64 About 600 of such mixed cultures were established and screened for a group of cosmopolitan
65 yet streamlined freshwater organisms from the Actinobacterial lineage acI. Here, we describe
66 six of these model communities and resolve their metabolic interaction potentials using
67 metagenomics. Finally, for a single mixed culture, we calculate co-occurrence patterns across
68 a metagenomics time-series in a lake to test if the observed dependencies are specific or if
69 the interaction-partners are flexible in more complex natural communities.
70

71 **Results and Discussion**

72 *Mixed cultures as model communities and their genomic reconstructions*
73 Ubiquitous freshwater Actinobacteria of the acI lineage were cultivated by diluting different
74 lake samples with triple-filtered sample water incubated in 96-well liquid culture plates. In
75 total, seven culture plates were set up and around forty-five wells contained acI. Of these,
76 only 6 cultures successfully propagated with densities comparable to those observed in the
77 environment (i.e. 10^6 cells ml^{-1}) for subsequent transfers and growth in larger volumes. This
78 resulted in 6 dilution-to-extinction mixed cultures: FNEF8, FNEB6, FNEB7, FNED7, FSWF8
79 and TBE6 (FNE refers to Northeast basin of Lake Grosse Fuchskuhle, FSW to Southeast
80 basin of Lake Grosse Fuchskuhle and TB to Trout Bog Lake). Characterizations of FNEF8
81 have been published, previously [7,10].

DNA was extracted from 4 L-cultures, and reads obtained by shotgun sequencing were assembled. Contigs > 1000 bp were considered for further analysis. In total, 77 metagenome assembled genomes (MAGs) were obtained from the assemblies (Supplemental). Of these, 31 MAGs each recruited more than 1% of the reads in the original metagenome. These MAGs have an average completeness of 83.5% as calculated by CheckM (Supplemental). It has been previously observed that because CheckM relies on lineage-specific marker genes, the completeness of genomes without lineage representation can often be underestimated [5,7]. However, since these MAGs have extremely high coverage and were assembled from model communities (low diversity of starter cells) we are confident of their high quality.

Some free-living microorganisms fulfill their metabolic needs from those they happen to encounter

Populations represented by the top 31 MAGs were assumed to be the dominant model community members and were analyzed further (mapping rates in Supplemental). All six cultures yielded at least one MAG of the cosmopolitan freshwater Actinobacteria acI (recruiting between 15% and 40% of reads), four cultures yielded a MAG of the freshwater Actinobacteria acIII (between 3% and 9% of the reads), two cultures yielded yeast MAGs (1 and 3% of the reads), two cultures featured MAGs affiliated with Bacteroidetes (2% and 6% of the reads), two cultures yielded a Polynucleobacter MAG (3% and 55% of the reads), two cultures yielded a Spirochaetes MAG (8% and 22% of the reads), one culture yielded an Alphaproteobacteria MAG (1% of the reads) and one an Acidimicrobiales MAG (48% of the reads). Knowing that acI Actinobacteria do not grow in pure culture [11], we now have strong support that the free-living acI is highly dependent on interactions with other microorganisms. Results from previous co-occurrence studies support the idea that free-living streamlined bacteria have very high connectivity in their environments and are critically dependent on metabolites that might be provided by the surrounding members of the planktonic communities [8,12]. Moreover, by creating an initial bottleneck [13] of approximately 12 cells , we obtained a first evidence that these plankton community partners do represent a diverse sets of community members. This hints towards a non-specific metabolic dependence of acI on other abundant freshwater bacteria, as no consistent interaction patterns were observed.

Since six mixed cultures might be not sufficient to generalize the character of the observed associations, we correlated the abundance of the mixed culture TBE6 MAGs with those obtained from a nine-year time-series shotgun metagenome from its source environment, Trout Bog Lake [14]. For this, we also used MAGs reassembled from this time-series data. The observed correlations indicate that acI interacts promiscuously and randomly (Figure 1, Figure S1 and S2). That is, no strong exclusive correlations can be discerned, either with the members of TBE6 or MAGs from the time-series, strengthening the hypothesis that acI is promiscuous in its interactions.

Tight metabolic dependencies between different free-living microorganisms

Canonical central metabolic pathways such as glycolysis, the pentose phosphate pathway, the citric acid cycle, ammonia assimilation and oxidative phosphorylation are observed in each of the abundant MAGs of the microbial model communities, suggesting a mainly aerobic heterotrophic lifestyle (Supplemental). When looking at all extant acI MAGs and single-cell genomes (SAGs) [15,16], auxotrophies for amino acids and vitamins were observed (Figure 2A). However, in the mixed cultures, at least one of the main members in each community was able to synthesize each of these metabolites, hinting to a high degree of metabolic dependency within the members of each model community of relatively low complexity. For example, in culture FNEB6, Bacteroidetes is the only producer of asparagine and so potentially supplies it to the other community members (see more examples in Figure 2B). Interactions at the level of amino acids and vitamins have been previously observed and described in several aquatic study systems [4-6]. Furthermore, one or two members of each model community had the ability to reduce sulfate, consistent with prior findings of the transfer of redox reaction products known as “metabolic handoffs” [17]. These kinds of observed dependencies merely represent a few examples of the many interactions that likely occur in the complex natural environment. In this respect, it was interesting that they also formed cross-domain linkages to two ubiquitous yeasts. We consider this as an indication that these interactions potentially can cross boundaries to eukaryotic plankton.

High degree of genomic diversity might support flexibility in interaction partners

Our findings from analysis of genome content combined with prior evidence of high genome-level diversity within acI [15] prompted us to examine the variability in gene content among acI genomes. Interestingly, even when our cultures harbor a reduced diversity, all of our cultures contained more than one genotype of acI. This likely reflects the large genomic plasticity of this common and dominant free-living freshwater bacterium. This genomic plasticity, however, cannot be resolved with the short 16S rRNA gene fragments obtained in normal amplicon sequencing (Supplemental). Using all 33 acI Actinobacteria MAGs from our 6 mixed cultures in addition to the 14 available acI single cell amplified genomes (SAGs) (Figure S3), ~8000 orthologous gene clusters (OGCs) were found. The distribution of gene clusters in the genomes revealed that ~800 OGCs form the core genome of acI (Figure S4 and S5). Assuming around 1600 genes per acI [15], this means that about half of the whole acI genome belongs to the flexible genome. This is around the same proportion as SAR11 [18]. Thus, both acI and SAR11 have quite larger flexible genomes than other ubiquitous free-living aquatic bacteria like Prochlorococcus [19]. The paradox of acI's streamlining tendencies is that their high number of auxiliary genes is likely to render populations of this free-living bacterial lineage functionally versatile.

In summary, as is the case for other streamlined free-living aquatic bacteria, acI Actinobacteria depend on numerous other abundant microorganisms for metabolic handouts. This kind of dependence seems to be non-taxa-specific or promiscuous since no consistent interaction patterns could be observed. These observations add to the knowledge on the structure of cooperation and dependencies between free-living aquatic microbes.

Methods

Mixed culture and DNA extraction

Mixed cultures that included the abundant and ubiquitous freshwater *Actinobacteria* were established in 2012 and 2013. Samples were collected from 5 different lakes: Lake Grosse Fuchskuhle (North East basin and South West basin), Lake Stechlin, Lake Dagow, Lake Mendota and from Trout Bog Lake (metadata reported in Supplementary Table). In total, seven plates were set up, two for Northeast basin (Lake Grosse Fuchskuhle) and one for

each of the other lakes (Southwest basin Lake Grosse, Lake Stechlin Lake, Lake Dagow, Lake Mendota and Trout Bog Lake. Cultivation method, media preparation, acI qPCR assays, general observations and DNA extractions were as previously described for mixed culture FNE-F8 [10]. Selected mixed cultures (based on cell density) were scaled up to a four-liter culture and DNA was extracted. DNA was sent to JGI and to Research and Testing Laboratory (RTL) (http://researchandtesting.com/) for shotgun metagenomic sequencing and 16S rRNA gene amplicon sequencing, respectively.

16S rRNA gene amplicon sequencing

RTL performed PCR amplification and MiSeq sequencing of the 16S rRNA gene amplicons. The bacterial primers (*E. coli* positions) were 357wF (5'-CCTACGGGNGGCWGCAG-3') and 926wR (5'-CCGTCAATTYMTTTRAGTTT-3'). All downstream analysis was done using the RTL standard pipeline from 2014. In brief all sequences were clustered at 4% divergence into OTUs using the UPARSE [20] algorithm. The centroid sequence from each cluster was then run against either the USEARCH global alignment algorithm or the RDP Classifier against a database of high quality sequences derived from the NCBI database. The output was then analyzed using an RTL internally developed python program that assigns taxonomic information to each sequence and then computes and writes the final analysis files.

Library preparation and sequencing

JGI performed the library preparation and the sequencing under Community Sequencing Project 1289. First, 100 ng of genomic DNA was sheared to 270 bp using a focused-ultrasonicator (Covaris). The sheared DNA fragments were size selected using SPRI beads (Beckman Coulter). The selected fragments were then end-repaired, A-tailed, and ligated of Illumina compatible adapters (IDT, Inc.) using KAPA-Illumina library creation kit (KAPA biosystems). The prepared sample library was quantified using KAPA Biosystem's next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified library was then prepared for sequencing on the Illumina HiSeq sequencing platform utilizing a TruSeq paired-end cluster kit, v3, and Illumina's cBot instrument to generate a clustered flowcell for sequencing. Sequencing of the flowcell was

202 performed on the Illumina HiSeq2000 sequencer using Illumina TruSeq SBS sequencing kits,
203 v3, following a 2x150 indexed high-output run recipe.
204
205 *Genome assembly, annotations and metabolic features analysis*
206 After reads had been filtered based on their quality scores using sickle (version 1.210) [21], the
207 reads were digitally normalized using khmer 1.4 [22] and assembled with megahit [23]. Coverage
208 was computed by mapping back the reads competitively to the assembly using bbmap 35.40
209 (sourceforge.net/projects/bbmap/) using default parameters. Mapped reads were indexed and
210 sorted using SAMtools 1.3 [24] while we removed duplicates using picard-tools (version 1.101).
211 Bedtools (version 2.18.2) [25] was used for computing coverage. For the binning metaBAT [26]
212 was used. Tribe/taxonomical affiliation were defined using whole genome information and the
213 public database using PhyloPhlAn [27] and using previously published genomes and SAGs [16]
214 as references. If the MAGs belong to one of the freshwater tribes [28] they have the name of the
215 tribe, otherwise the taxonomic name after NCBI is given.
216 For bacterial genomes, gene prediction analysis was performed within the Integrated
217 Microbial Genomes (IMG) [29,30] platform developed by the Joint Genome Institute, Walnut
218 Creek, CA, USA (http://img.jgi.doe.gov). Specific KEGG biosynthetic maps were inspected for
219 completeness (http://www.genome.jp/kegg/mapper.html) counting number of missing
220 enzymes from the most complete pathway. MAG completeness was calculated using CheckM
221 [31].
222 For fungal genomes, ab initio gene prediction was performed using augustus (v 2.5.5) with
223 Aspergillus nidulans (FSWF8-4) or Cryptococcus neoformans (FNED7-22) models. Gene
224 model ORFs were then annotated using matches to Pfam 30.0 (June 2016) by hmmscan
225 (HMMER 3.1b2). Functional annotation was done using predicted aminoacids in the KAAS
226 platform [32].
227
228 *Core genome computation*
229 Clusters of orthologous genes were identified using OrthoMCL [33] using default options
230 following an established protocol. Core genome of the acI was computed using all 33 acI
231 Actinobacteria MAGs from our 6 mixed cultures in addition to the 14 available acI single cell

amplified genomes (SAGs). Due to the incompleteness SAGs and MAGs, common methods, such as picking all the OGCs present in at least 90% of the genomes, are unsafe. To reduce the noise generated by incompleteness, a semi-supervised approach has been chosen. For this we compute a distance matrix based on the genomes that contain the different OGCs. Two OGCs will have a distance of 0.0 if they are found in the same genomes, and a distance of 1.0 if none of the genomes containing one also contains the other. The overlap between the genome sets of each cluster were computed and normalized to the smallest genome set. OGCs contained in less than 4 genomes were removed. The obtained distances are then used with a hierarchical clustering to compute a heatmap (Supplemental Figure acI core). This heatmap clearly shows a set of OGCs that overlaps significantly with all other OGCs, this cluster is selected as core. Core OGCs should have a strong overlap with all other OGCs irrelevant of their position in the heatmap.

Metagenome recruitment to MAGs

A total of 113 metagenome samples from Trout Bog lake [14] were used in this study. As above, after assembly and binning, bbmap was used to compute the coverage of all the MAGs (from the time-series as well as the cultures), for all the time-points. All coverages have been normalized to sequencing depth. Mean coverage vectors of all MAGs are then correlated using Spearman correlation. For the heatmaps, MAGs from Trout Bog lake were only kept if they correlate with at least 0.6 to at least one MAG from our mixed culture TBE6 (positive or negative correlation). MAGs assembled from Trout Bog lake have not been further analyzed as they might represent an average population genome.

Competing financial interests

The authors declare no conflict of interest

Acknowledgements

In memory of Falk Warnecke, who believed in the power of mixed cultures and envisioned this manuscript already a few years ago. This work was supported by: the Swedish research council VR (grant 2012-4592), the German Science Foundation (grants DFG GR1540/ 17-1

and 21-1), the JSMC, the United States National Science Foundation Microbial Observatories program (MCB-0702395), the Long Term Ecological Research program (NTL-LTER DEB-1440297), and INSPIRE award (DEB- 1344254). JJH was supported by a USDA NIFA AFRI Postdoctoral Fellowship, award number 2016-67012-24709. We are also grateful to the Joint Genome Institute for supporting this work through the Community Sequencing Program (CSP) and providing technical support. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Thanks to Alexandra Linz for her support with the TBE6 mixed culture. Thanks to staff at the Joint Genome Institute including Rex Malmstrom, Matthew Bendall, Tijana Glavina del Rio, Rob Eagan, and Susannah Tringe.

References

1 Schink, B. Synergistic interactions in the microbial world. *Anton Leeuw Int J G* **81**, 257-261, doi:Doi 10.1023/A:1020579004534 (2002).

2 Giovannoni, S. J., Cameron Thrash, J. & Temperton, B. Implications of streamlining theory for microbial ecology. *ISME J* **8**, 1553-1565, doi:10.1038/ismej.2014.60 (2014).

3 Oliveira, N. M., Niehus, R. & Foster, K. R. Evolutionary limits to cooperation in microbial communities. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 17941-17946, doi:10.1073/pnas.1412673111 (2014).

4 D'Souza, G. & Kost, C. Experimental Evolution of Metabolic Dependency in Bacteria. *PLOS Genetics* **12**, e1006364, doi:10.1371/journal.pgen.1006364 (2016).

5 Brewer, T. E., Handley, K. M., Carini, P., Gilbert, J. A. & Fierer, N. Genome reduction in an abundant and ubiquitous soil bacterium 'Candidatus Udaeobacter copiosus'. *Nature Microbiology* **2**, 16198, doi:10.1038/nmicrobiol.2016.198 (2016).

6 Giovannoni, S. J. Vitamins in the sea. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 13888-13889, doi:10.1073/pnas.1211722109 (2012).

7 Garcia, S. L. *et al.* Auxotrophy and intrapopulation complementary in the "interactome' of a cultivated freshwater model community. *Molecular ecology* **24**, 4449-4459, doi:10.1111/mec.13319 (2015).

8 Milici, M. *et al.* Co-occurrence Analysis of Microbial Taxa in the Atlantic Ocean Reveals High Connectivity in the Free-Living Bacterioplankton. *Frontiers in microbiology* **7**, doi:UNSP 649
10.3389/fmicb.2016.00649 (2016).

9 Garcia, S. L. Mixed cultures as model communities: hunting for ubiquitous microorganisms, their partners, and interactions. *Aquatic Microbial Ecology* **77**, doi:doi:103354/ame01796 (2016).

10 Garcia, S. L., McMahon, K. D., Grossart, H. P. & Warnecke, F. Successful enrichment of the ubiquitous freshwater acI Actinobacteria. *Environ Microbiol Rep* **6**, 21-27, doi:10.1111/1758-2229.12104 (2014).

11 Salcher, M. M. & Šimek, K. Isolation and cultivation of planktonic freshwater microbes is essential for a comprehensive understanding of their ecology. *Aquatic Microbial Ecology* **77**, 183-196, doi:10.3354/ame01796 (2016).

12 Zelezniak, A. *et al.* Metabolic dependencies drive species co-occurrence in diverse microbial communities (vol 112, pg 6449, 2015). *Proceedings of the National*

Academy of Sciences of the United States of America **112**, E7156-E7156, doi:10.1073/pnas.1522642113 (2015).

13 Chuang, J. S., Rivoire, O. & Leibler, S. Simpson's Paradox in a Synthetic Microbial System. *Science* **323**, 272-275, doi:10.1126/science.1166739 (2009).

14 Bendall, M. L. *et al.* Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *ISME J*, doi:10.1038/ismej.2015.241 (2016).

15 Ghylin, T. W. *et al.* Comparative single-cell genomics reveals potential ecological niches for the freshwater acI Actinobacteria lineage. *ISME J* **8**, 2503-2516, doi:10.1038/ismej.2014.135 (2014).

16 Garcia, S. L. *et al.* Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations. *bioRxiv*, doi:10.1101/080168 (2016).

17 Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* **7**, doi:ARTN 13219

10.1038/ncomms13219 (2016).

18 Grote, J. *et al.* Streamlining and Core Genome Conservation among Highly Divergent Members of the SAR11 Clade. *Mbio* **3**, doi:ARTN e00252-12

DOI 10.1128/mBio.00252-12 (2012).

19 Kashtan, N. *et al.* Single-Cell Genomics Reveals Hundreds of Coexisting Subpopulations in Wild Prochlorococcus. *Science* **344**, 416-420, doi:10.1126/science.1248575 (2014).

20 Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature methods* **10**, 996-+, doi:10.1038/Nmeth.2604 (2013).

21 Joshi, N. A. & Fass, J. N. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files [Software]. *Available at* https://github.com/najoshi/sickle.

22 Crusoe, M. R. *et al.* The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Res* **4**, 900, doi:10.12688/f1000research.6924.1 (2015).

23 Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674-1676, doi:10.1093/bioinformatics/btv033 (2015).

24 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

25 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:Doi 10.1093/Bioinformatics/Btq033 (2010).

26 Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165, doi:10.7717/peerj.1165 (2015).

27 Segata, N., Bornigen, D., Morgan, X. C. & Huttenhower, C. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat Commun* **4**, 2304, doi:10.1038/ncomms3304 (2013).

28 Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D. & Bertilsson, S. A Guide to the Natural History of Freshwater Lake Bacteria. *Microbiol. Mol. Biol. Rev.* **75**, 14-49, doi:10.1128/mmbr.00028-10 (2011).

29 Markowitz, V. M. *et al.* IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res* **42**, D560-567, doi:10.1093/nar/gkt963 (2014).

30 Markowitz, V. M. *et al.* IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res* **42**, D568-573, doi:10.1093/nar/gkt919 (2014).

31 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*, doi:10.1101/gr.186072.114 (2015).

32 Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* **35**, W182-W185, doi:10.1093/nar/gkm321 (2007).

33 Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research* **13**, 2178-2189, doi:10.1101/gr.1224503 (2003).

Data accessibility

The raw shotgun metagenome reads are publicly available in the JGI portal and the assembly is available in IMG/MER under the ER submission IDs 26656, 26658, 26650, 29729, 29808, 50227. The bacterial metagenome-assembled genomes (MAGs) are also available through IMG. The fungal MAGs have been deposited at DDBJ/ENA/GenBank. For taxon OID or accession number check supplementary material.

Author contribution

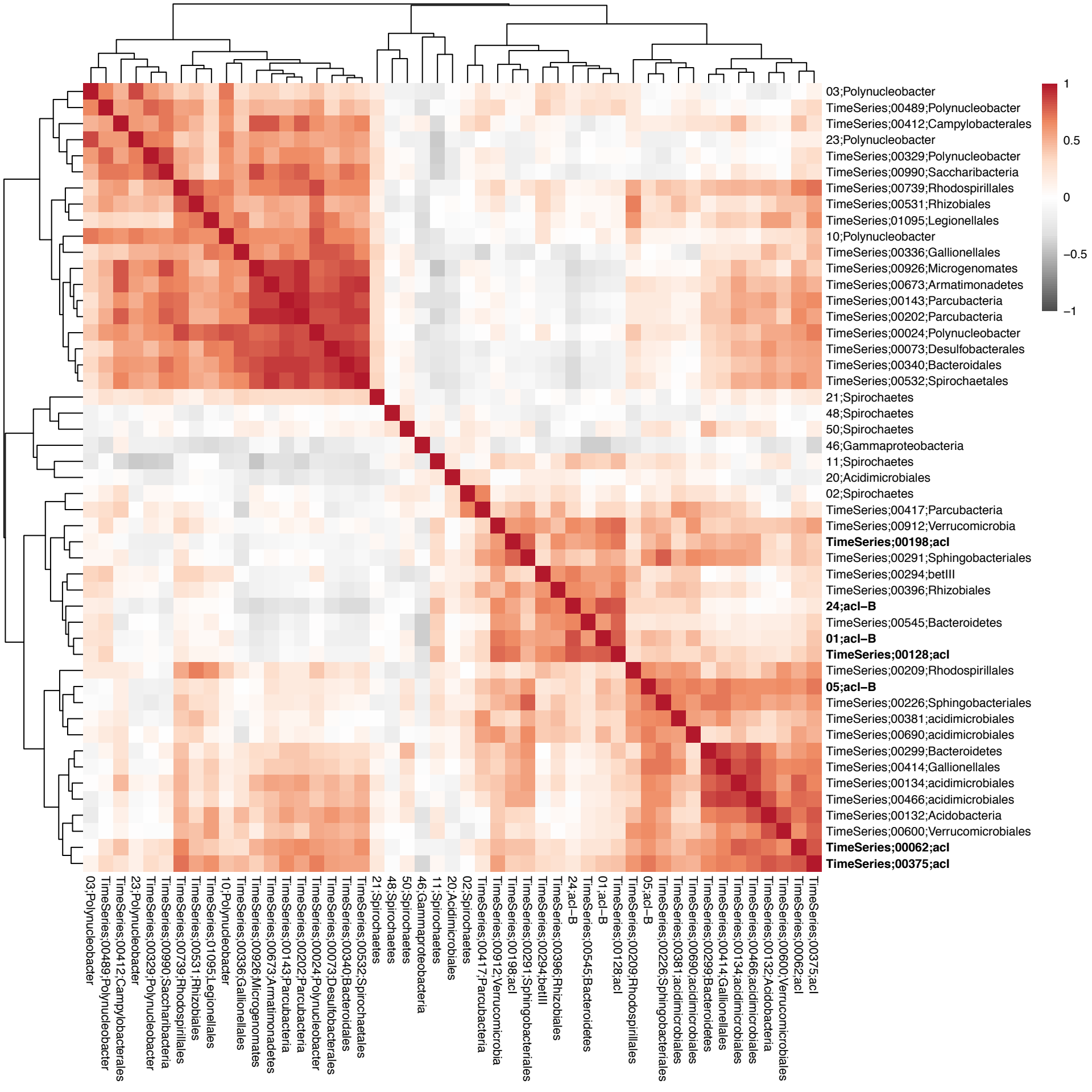
SLG and FW conceived the research. SLG and HPG collected and prepared the samples. SLG, MB, JJH, AE, CW and MAR analyzed the data. All authors wrote and/or revised the manuscript.

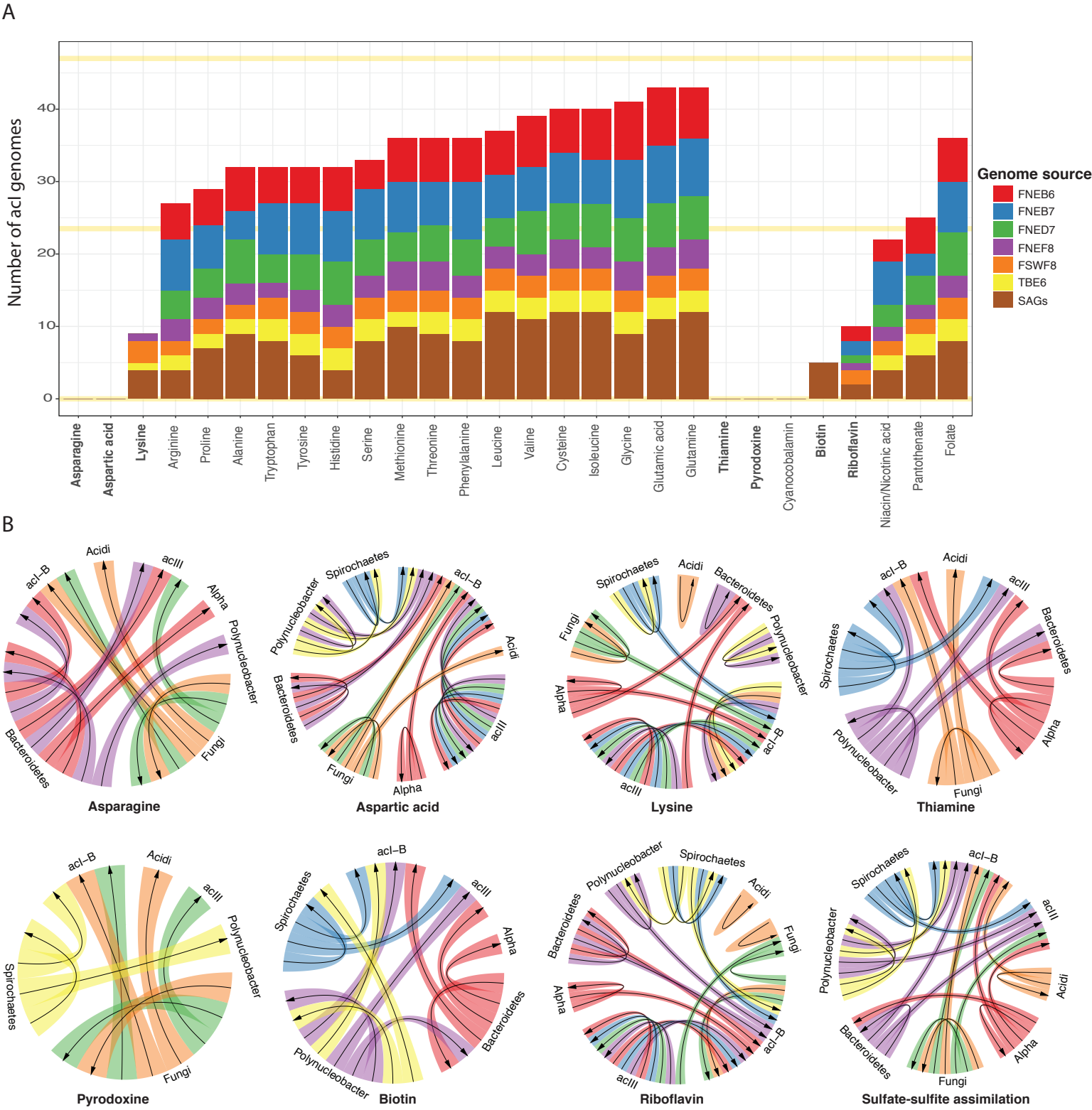
Conflict of interest Statement: The authors declare no conflict of interest.

Titles and legends to figures

Figure 1. Spearman correlation on normalized relative abundance between MAGs from TBE6 and taxa in the epilimnion of the environment of origin (Trout Bog Lake). Metagenome samples from time-points from all years were used (45 unique dates). Correlation is displayed via hierarchical clustering. MAG names that start with a number refers to a MAG from the mixed culture (for details on these MAGs look at supplemental tables). MAG names that start with the words "time series" were previously binned from the Trout Bog nine year metagenome series [14]. The acI MAGs are highlighted in bold. To look into the correlations of hypolimnion and both combined look into the supplemental.

Figure 2. A. Visualization of the amino acid and vitamin biosynthetic potential of all acI MAGs and SAGs analyzed. It includes acI information from all MAGs binned from each culture in this study plus the previously published SAGs. Yellow lines indicate 0%, 50% and 100% of genomes analyzed. Completeness of a metabolic pathway was considered if 80% of the pathway was present. **B.** Potential metabolic complementarity among major members of each of the cultures. Circle plots are displayed for metabolites highlighted in bold in panel A plus potential to assimilate sulfate and sulfite. Colors indicate each mixed culture using same color scheme as in panel A. Arrow shows the direction in which the metabolite would be "shared". Alpha – Alphaproteobacterium. Acidi – Acidimicrobium.


1
0.5
0
−0.5
−1
03;Polynucleobacter
TimeSeries;00489;Polynucleobacter
TimeSeries;00412;Campylobacterales
23;Polynucleobacter
TimeSeries;00329;Polynucleobacter
TimeSeries;00990;Saccharibacteria
TimeSeries;00739;Rhodospirillales
TimeSeries;00531;Rhizobiales
TimeSeries;01095;Legionellales
10;Polynucleobacter
TimeSeries;00336;Gallionellales
TimeSeries;00926;Microgenomates
TimeSeries;00673;Armatimonadetes
TimeSeries;00143;Parcubacteria
TimeSeries;00202;Parcubacteria
TimeSeries;00024;Polynucleobacter
TimeSeries;00073;Desulfobacterales
TimeSeries;00340;Bacteroidales
TimeSeries;00532;Spirochaetales
21;Spirochaetes
48;Spirochaetes
50;Spirochaetes
46;Gammaproteobacteria
11;Spirochaetes
20;Acidimicrobiales
02;Spirochaetes
TimeSeries;00417;Parcubacteria
TimeSeries;00912;Verrucomicrobia
TimeSeries;00198;acl
TimeSeries;00291;Sphingobacteriales
TimeSeries;00294;betIII
TimeSeries;00396;Rhizobiales
24;acI−B
TimeSeries;00545;Bacteroidetes
01;acI−B
TimeSeries;00128;acl
TimeSeries;00209;Rhodospirillales
05;acI−B
TimeSeries;00226;Sphingobacteriales
TimeSeries;00381;acidimicrobiales
TimeSeries;00690;acidimicrobiales
TimeSeries;00299;Bacteroidetes
TimeSeries;00414;Gallionellales
TimeSeries;00134;acidimicrobiales
TimeSeries;00466;acidimicrobiales
TimeSeries;00132;Acidobacteria
TimeSeries;00600;Verrucomicrobiales
TimeSeries;00062;acl
TimeSeries;00375;acl


A
Number of acl genomes
0
10
20
30
40
Genome source
FNEB6
FNEB7
FNED7
FNEF8
FSWF8
TBE6
SAGs
Asparagine
Aspartic acid
Lysine
Arginine
Proline
Alanine
Tryptophan
Tyrosine
Histidine
Serine
Methionine
Threonine
Phenylalanine
Leucine
Valine
Cysteine
Isoleucine
Glycine
Glutamic acid
Glutamine
Thiamine
Pyrodoxine
Cyanocobalamin
Biotin
Riboflavin
Niacin/Nicotinic acid
Pantothenate
Folate
B
Acidi
acI–B
acIII
Alpha
Polynucleobacter
Bacteroidetes
Fungi
Asparagine
Spirochaetes
Aspartic acid
Lysine
Thiamine
Pyrodoxine
Biotin
Riboflavin
Sulfate-sulfite assimilation