bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Taxonomic and functional diversity of aquatic heterotrophs is sustained by dissolved organic matter chemodiversity

Sarahi L. Garcia [a,b#] (sarahi.garcia@su.se), Julia K. Nuy [a] (julenuy@gmail.com), Maliheh Mehrshad [c] (maliheh.mehrshad@slu.se), Justyna J. Hampel [a] (justyna.hampel@su.se), Vicente T. Sedano-Nuñez [d] (vicente.sedano.nunez@gmail.com), Moritz Buck [c] (moritz.buck@slu.se), Anna-Maria Divne [e] (anna-maria.divne@icm.uu.se), Eva S. Lindström [b] (eva.lindstrom@ebc.uu.se), Daniel Petras [f] (daniel.petras@uni-tuebingen.de), Jeffrey Hawkes [g] (jeffrey.hawkes@kemi.uu.se), Stefan Bertilsson [b,c] (stefan.bertilsson@slu.se)

[a] Department of Ecology, Environment, and Plant Sciences, Science for Life Laboratory, Stockholm University, 10691 Stockholm, Sweden

[b] Limnology, Department of Ecology and Genetics, Uppsala University, 75236, Uppsala, Sweden.

[c] Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, 75651 Uppsala, Sweden

[d] Molecular Evolution, Department of Cell and Molecular Biology, Uppsala University, 75237 Uppsala, Sweden

[e] Department of Cell and Molecular Biology, Uppsala University, 75237 Uppsala, Sweden

[f] CMFI Cluster of Excellence, Interfaculty Institute of Microbiology and Medicine, University of Tübingen

[g] Analytical Chemistry, Department of Chemistry, Uppsala University, 75237 Uppsala, Sweden

#Corresponding author

Article type: Original research article

The authors declare no competing interests

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Dissolved organic matter (DOM) is ubiquitous in aquatic ecosystems and fundamental for planetary processes and ecosystem functioning. While the link between microbial community composition and heterotrophic utilization of DOM has been recognized, the full diversity of organic compounds, their bioavailability, degradability and specific influences on the diversity and function of heterotrophs are still not clear. Here we experimentally investigate heterotrophic bacteria in thirty-three freshwater model communities. We identified 34 different heterotrophs growing in ambient lake DOM with taxonomic affiliations matching abundant freshwater bacterioplankton. We further describe 25 different heterotrophs growing in the phycosphere of *M. Aeruginosa* and 6 heterotrophs growing on the DOM produced by *M. Aeruginosa* with taxonomic affiliation in accord to phycosphere heterotrophs. In our experiment we observed that heterotrophs that live in the phycosphere remove more dissolved organic carbon than abundant freshwater heterotrophs. Moreover, phycosphere heterotrophs have bigger genomes than abundant lake bacteria. Altogether of the 4224 chemical features that were resolved by LC-MS, only 1229 were seen in all three treatments. None of the common/shared compounds were removed across all the model communities, suggesting contrasting niches of the studied taxa. Altogether our study highlights how each model community, with its unique taxonomic assemblages and organotroph functioning is upkept by the chemodiversity of DOM.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Dissolved organic matter (DOM) is ubiquitous in aquatic ecosystems and fundamental for planetary processes. DOM consists of a complex and variable mixture of compounds that serve as substrates for heterotrophic microorganisms (Azam 1998, Ferrer-Gonzalez et al 2021, Patriarca et al 2020b). As long-term carbon repository (~660 Gt C), DOM holds 200 times more carbon than aquatic living biomass and is equal to the carbon storage in the atmosphere (Bar-On et al 2018, Falkowski et al 2000, Hansell et al 2009, Koch et al 2008). The bulk of DOM in continental waters is typically allochthonous (derived from terrestrial sources), chemically complex, and biologically recalcitrant (Amon and Benner 1996, Koehler et al 2012) while autochthonous DOM is produced in-situ and typically more bioavailable and therefore rapidly consumed in a way that prevents build-up (Maki et al 2009, Patriarca et al 2020b). Aquatic microorganisms play central roles both as DOM producers and consumers and also connect these dissolved substrates to larger organisms in the food web via the microbial loop (Azam 1998, Fenchel 2008). Despite the significance of microorganisms as mediators of biogeochemical processes in aquatic environments, their ecological niches in the turnover of the DOM pool are still unresolved. This is at least in part because aquatic microbial diversity is high while many abundant microbes remain uncultivated (Rodríguez-Gijón et al 2022), and also because of the high complexity and dynamic nature of aquatic DOM (Patriarca et al 2020b).

Earlier studies have demonstrated that phylogenetically diverse marine phytoplankton species release distinct arrays of organic compounds into their phycosphere (Becker et al 2014). This attracts and selects for distinct groups of heterotrophic bacteria (Ferrer-Gonzalez et al 2021, Sarmento and Gasol 2012). Accordingly, the composition and diversity of freshly produced and released autochthonous DOM influences microbial community composition and diversity, and

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

62 those microorganisms that profit will in turn interactively drive DOM turnover in the oceans.
63 Neveretheless, the most abundant marine bacteria (Giovannoni et al 2005) are not particularly
64 abundant in the phycosphere (Buchan et al 2014, Seymour et al 2017) suggesting a more cryptic
65 and multifaceted role of autochthonous DOM as a factor controlling microbial community
66 structure. The same pattern is seen in freshwater lakes, where microbes that are numerically
67 abundant in planktonic communities (Newton et al 2011) are typically not prominent or abundant
68 in the phycosphere (Cai et al 2014, Eiler et al 2006) suggesting niche specialization with regards
69 to DOM use, possibly related to the much larger allochthonous component in the combined
70 DOM pool (Patriarca et al 2020b).

71 To experimentally uncover links between heterotrophs and their consumption of DOM, a
72 few pioneering studies have relied on mass spectrometry paired with simplified model systems
73 with a few identified keystone species of heterotrophs (Ferrer-Gonzalez et al 2021, Uchimiya et
74 al 2021). Furthermore, a study using fourteen radiolabeled substrates and microautoradiography
75 and fluorescence in situ hybridization (MAR-FISH) found niche partitioning with regards to
76 organic substrate use for different groups of abundant lake bacteria (Salcher et al 2013). While
77 bacterial community composition and substrate utilization patterns are tightly linked in aquatic
78 ecosystems (Logue et al 2016), the detailed relationship between the complexity and reactivity of
79 DOM and the functionally diverse heterotrophs is still not well understood.

80 With this study we advance the understanding of how specific heterotrophic bacteria act
81 as central agents in the microbial loop. As models we use model communities developed from
82 lake bacterioplankton under selection (i) in coculture with the cyanobacterium *Microcystis*
83 *aeruginosa* (ii) growing on DOM produced from an axenic culture of *M. aeruginosa*, and (iii)
84 growing in ambient lake DOM. After dilutions and serial transfer, cultivated model communities

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were established and characterized with a powerful combination of comparative metagenomics and high-resolution untargeted metabolite analysis. The model communities served as tractable model systems with reduced complexity (Garcia 2016) to assess the role of DOM quality in selecting for different microbial cohorts, the reactivity of different DOM compounds and how the heterotrophs differed with regards to organic compound utilization patterns.

Results and discussion

Microbial assemblages grew faster on freshly produced DOM than on lake DOM

In total, we inoculated 990 cultures, of which only 33 feature stable growth and could be upscaled (Table S1). Out of these 33 cultures, 5 originated from inoculation on spent media from axenic *M. aeruginosa* (M-DOM; initial DOC 6.2 mg/L), 21 cultures originated from filtered sterilized lakewater (L-DOM; initial DOC 10.7 mg/L) and 7 cultures were established as co-cultures with *M. aeruginosa* (M-co; DOC without heterotrophs 14.6 mg/L) (Figure 1). For growth characteristics, some general observations were made; (i) the M-DOM cultures reached the highest cell densities and also featured the highest growth rates (Figure 2E), (ii) M-co cultures reached similar cell densities as L-DOM cultures, (iii) among L-DOM and M-DOM communities, those that reached higher cell densities also grew faster than those that reached lower cell densities (Figure 2E). Finally, heterotrophs in M-co cultures used on average more DOC (dissolved organic carbon) per cell per day, while heterotrophs in M-DOM cultures used the least amount of DOC (Figure 2F). In our experiment we can observe how heterotrophs that live in the phycosphere and are not considered among the most abundant aquatic microorganisms, remove more dissolved organic carbon than abundant heterotrophs.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Treatments are imprinted in the heterotroph average genome size

We reconstructed 61 metagenome-assembled genomes (MAGs) from the community assemblies in L-DOM cultures (Figure 2 and Table S1). These 61 MAGs grouped into 34 metagenomic operational taxonomic units (mOTUs) calculated based on 95% average nucleotide identity (ANI). 95% ANI is a widely recognized and used threshold for operationally defining the microbial equivalent to species (Garcia et al 2018b, Jain et al 2018). We further reconstructed 9 MAGs from M-DOM cultures that grouped into 6 mOTUs and an additional 36 MAGs from M-co cultures that grouped into 25 mOTUs. For context, we included a synoptic genomic inventory of the microbial source community in Lake Erken which included data from public databases (Buck et al 2021a, Mondav et al 2020). All 14 Lake Erken samples compiled yielded 248 MAGs that clustered in 205 mOTUs (Figure 2). The 354 MAGs were of good quality at >50% completeness and <5% contamination (Figure 2D) (Parks et al 2015).

We compiled and compared the genomes of representative MAGs from each mOTU (Table S3) for the different experimentally derived cultures as well as for Lake Erken. We found that the average estimated genome size was similar between M-DOM and M-co cultures (Wilcoxon, p=0.71) and MAGs from these two treatments had a significantly larger estimated genome size as compared to genomes from L-DOM cultures (Figure 2A and Figure S1, Wilcoxon p=0.019 and p=0.0001). The estimated genome size in the Lake Erken genomic survey encompassed the entire genome size range observed in the cultures (Figure 2A) but in average was smaller than genomes in M-DOM (Wilcoxon, p=0.07) and significantly smaller than those in M-co-cultures (Figure S1, Wilcoxon p=0.001). In conclusion, this implies that large genome size is characteristic for heterotrophs in the phycosphere. It is known that larger genomes require

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

more maintenance (Giovannoni et al 2014), and the phycosphere provides an environment with labile and 'plentiful' resource.

Eight of the L-DOM treatment MAGs clustered in mOTUs with in-situ MAGs from Lake Erken (Figure 2B, MAGs marked in bold in Figure 3, Table S3 for details), while merely one of the MAGs from the M-DOM treatment clustered in an mOTU together with MAGs from Lake Erken. None of the MAGs from M-co cultures clustered together in mOTUs with in-situ MAGs from Lake Erken. Moreover, for most of the MAGs recovered from M-DOM and M-co culture treatment it was difficult to recruit reads from Lake Erken metagenomes that matched with more than 95% sequence identity (Figure 3). This suggests that microorganisms inhabiting the phycosphere are quantitatively scarce in community-level samples retrieved from the lake. The phycosphere is known as a hotspot for biogeochemical cycling (Seymour et al 2017) where a large proportion of the DOM released from phytoplankton cells are consumed (Smriga et al 2016).

Taxonomic composition of bacterial diversity is treatment-dependent

All three treatments fostered communities with a distinct taxonomic signature. In the L-DOM cultures we retrieved lineages that are typical for freshwater lakes (Newton et al 2011), such as *Polynucleobacter* (Hahn et al 2016), *Planktophila* (Mondav et al 2020), *Caulobacterales* (Hentchel et al 2019) and *Limnohabitans* (Hahn et al 2010). In the M-co cultures, we recovered diverse *Alphaproteobacteria* affiliated with the orders *Sphingomonadales* and *Rhizobiales* as well as *Bacteroida* within the *Cytophagales* and *Flavobacterials*, most of them uncharacterized at the species level. The diversity was much lower in the M-DOM cultures, where recovered taxa represented *Hydrogenophaga*, *Caulobacter*, *Brevundimonas* and *Flavobacterium omnivorum*. A

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

single mOTU included MAGs from both the M-DOM and M-co culture treatments (Figure2B, *Flavobacterium* DOM03 in bold in Figure 3). None of the mOTUs included MAGs from all three treatments. In our approach through dilution and serial transfers, we created reduced complexity sub-communities from the natural lake bacterioplankton assemblage. This approach enable diverse and abundant microorganisms to grow while maintaining interactions with partnering microorganisms that they also interact with in their natural environment (Garcia 2016, Pascual-Garcia et al 2020, Yu et al 2019). The observed taxonomic composition per treatment fit previous observations of abundant freshwater heterotrophs (Newton et al 2011, Salcher et al 2013) and the broader classes that are known to thrive on freshly produced phytoplankton DOM (Cai et al 2014, Eiler et al 2006, Ferrer-Gonzalez et al 2021).

Reconstructed good quality MAGs from cultures across all treatments grouped in 31 different genera. Amongst them, Flavobacterium with 22 representatives in 10 different mOTUs was the most prevalent. Flavobacterium representatives were recovered in 62% of the L-DOM cultures, 40% of M-DOM cultures, and 86% of the M-co cultures. The mOTUs of cultured Flavobacteria from L-DOM treatment were more abundant in the Lake Erken metagenomes collected in colder and deeper waters whereas the Flavobacterium from M-DOM and M-co cultures were hardly found in the lake (Figure 3). In contrast, only 2 additional Flavobacterium mOTUs were recovered from Lake Erken metagenomes and one of those was abundant in the warmer and upper water column (Figure S2). As the samples used to set up model communities of this study were collected in April 2019, their high prevalence in the emerging cultures is likely at least partly related to high abundances in the inoculum but may also be related to their previously recognized high capacity to exploit phytoplankton-derived organic matter (Williams

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

et al 2013, Zeder et al 2009) and this is also supported by their metabolic profile suggesting a heterotrophic lifestyle.

Polynucleobacter is another prevalent genus with 11 representatives among reconstructed MAGs. These MAGs clustered in five different mOTUs and grew in 12 cultures, all of which were from the L-DOM treatment. These MAGs had an estimated genome size in the range of 2 to 2.6 Mb and GC content in the range of 43 to 49%. Metabolic reconstruction of these MAGs suggests a photoheterotrophic lifestyle with the capacity for aerobic anoxygenic phototrophy via pufM/L genes. An overall view of the metabolic potential of MAGs belonging to this genus also suggests thiosulfate oxidation via SOX complex.

Genomes in different treatments vary in encoded functions

Comparative analysis of functional annotation of the MAGs recovered from each treatment show that more than 59% of all annotated functions (3516 out of 5918 unique KO identifiers) were shared among all three treatments (Venn diagram in Figure 4). These shared functions comprise 75.5, 89.2, and 66.3 percent, respectively, of the annotated functions in MAGs grown on L-DOM, Microcystis DOM, and M-co culture DOM treatments, and encode for core bacterial metabolic modules.

The 13.6% of annotated functions unique to MAGs reconstructed from M-co cultures encode for modules related to photosynthesis (found in the reconstructed Microcystis MAGs) and thiamine salvage pathway. Thiamin diphosphate that is the active form of thiamin, functions as the cofactor for transketolases, decarboxylases, and other enzymes that are involved in producing or breaking C-C bonds. At the genome level, this module (K00878 and K14153) is encoded in treatments M-co83 and M-co88 and in both cases in two representatives of the

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

198 *Polaromonas* genus that are clustering in separate mOTUs. In our dataset, no other MAG was
199 detected to encode for K00878 (hydroxyethylthiazole kinase), limiting this module to these two
200 MAGs. This would point to thiamine salvage pathway being a possible candidate for community
201 cooperation in the environment (Mondav et al 2020).

202 Functions shared between L-DOM and M-co cultures comprise 15% of total annotated
203 functions in M-co cultures and encode for genes involved in degradation of aromatic compounds
204 via a complete module of Catechol ortho-cleavage (K03381, K01856, K03464, K01055, and
205 K14727). Catechol is a central intermediate in the degradation of different aromatic compounds
206 and has previously been reported to also be produced in some algae (Singh et al 2017). At the
207 level of individual genomes, we detected the full module in bin-30 MAG from the M-co89
208 treatment affiliated to genus *Bosea*, as well as bin-08 and bin-0 from L-DOM treatments
209 taxonomically affiliated with Sphingomonadaceae and genus Ga0077559. Both of these MAGs
210 cluster in mOTU_023 with two other MAGs from the L-DOM treatments (ERK38_bin-14 and
211 ERK49_bin-02 with 67.6 and 97.3% completeness, respectively Table S2) where the latter two
212 are missing genes involved in the Catechol ortho-cleavage module. This hints at a role of ecotype
213 differentiation for DOM degradation potential (Mondav et al 2020). Interestingly members of
214 mOTU_023 seem to grow with different microbial partners in the 4 different L-DOM model
215 communities where they were present. Apart from the two Sphingomonadaceae MAGs,
216 individual genes involved in Cathecol-ortho cleavage were also detected in other MAGs but not
217 as a complete module.

218 The common annotated functions between M-co cultures and the M-DOM cultures (5%)
219 encode for genes such as beta lactamase class D encoded in MAGs CO86_bin.08, affiliated to
220 genus Hyphomonas and DOM03_bin-0 and DOM04_bin-10 clustering as mOTU_017 affiliated

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to genus Flavobacterium. Interestingly both of these Flavobacterium MAGs were reconstructed from M-DOM treatments that only contain one other MAG affiliated to Pseudorhodobacter_A (both of these also cluster in mOTU_150). The Pseudorhodobacter_A MAGs have photoheterotrophic metabolism with aerobic anoxygenic phototrophy via pufM/L genes. The Pseudorhodobacter_A MAGs also encode complete modules for biosynthesis of cobalamin and can synthetize cobyrinate a,c-diamide via the aerobic pathway as opposed to Microcystis that encode the anaerobic pathway for cobyrinate a,c-diamide production via sirohydrochlorin.

Among genes uniquely retrieved in the L-DOM treatments (9.5% of all annotated functions in these treatments) we observed several different Dioxygenases. One of the most important functions of dioxygenases is the cleavage of aromatic rings and they are widely found in nature (Harayama and Rekik 1989). L-DOM MAGs only share a few functions with the M-DOM MAGs, but this include some dioxygenases suggested to be involved in naphthalene degradation (Figure 4).

Genome-encoded functions unique to the M-DOM treatments (only 2% of the annotated functions in this treatments) encode for beta lactamases and biphenyl 2,3-dioxygenase ferredoxin component. The highest diversity of dioxygenases was detected in the L-DOM cultures. Apart from the fact that most of the reconstructed MAGs belong to the L-DOM cultures (n=61) compared to M-co cultures (n=36) and M-DOM cultures (n=9), this higher diversity could also be due to higher representation of aromatic compounds in the native DOM pool of the L-DOM treatment as compared to the cyanobacterial DOM treatments (Patriarca et al 2020b).

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chemodiversity of aquatic DOM sustain bacterial diversity

We discuss our data in terms of chemical features, which are MS/MS signals. One metabolite can lead to several features, so we refrain from naming them metabolites or compounds. Altogether 4224 LC-MS features were resolved and had sufficient peak intensity to trigger data-dependent MS/MS analysis in at least one sample in the dataset. Of these features, 3417 were present in at least one sample above 5x the intensity of the average of 4 blanks. Depending on the culture type (M-DOM, L-DOM or M-co cultures), a subset of these 3417 features were detected in the initial feed (see first row of Figure 5B). The M-DOM, L-DOM and M-co feed had 2422, 1872 and 2146 features respectively, with only a minor fraction unique to any individual feed (Figure 5C). 1229 features were found in all three feeds. Variable numbers of features were either removed or produced in the incubation replicates, leading to presence/absence and intensity differences in the sample extracts (Figure 5B). Pairwise sample dissimilarities were calculated with the Bray-Curtis metric, allowing samples to be compared and grouped in a principal coordinate diagram (PCoA) (Figure 5A). The L-DOM incubation samples were clearly separated from the M-DOM and M-co cultures along the first principal coordinate, while the latter two groups were separated along the second, indicating that the treatments varied more than the variance among biological replicates within each group. A PERMANOVA analysis of the Bray-Curtis distance matrix found an F-statistic of 22.1 for group membership (L-DOM, M-DOM or M-co), with a p value of 0.01 over 100 iterations of the PERMANOVA algorithm. As expected, the controls for M-DOM (medium incubated without heterotrophs) and the M-co cultures (*M. aeruginosa* cultivated without heterotrophs) were similar while they diverged along PCoA2 during the experimental incubations.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Of the 1229 features found in all three feeds, 372 remained in all incubations thoruhgout the experimental incubation. None of the common/shared feed features were removed across all of the cultures, indicating niche partitioning on substrate consumption by the taxa cultivated in the different treatments. However, for each feed type, there was a moderate number of features (between 51 and 131) that were removed in every culture with this experimental group (see third row of Figure 5B). We suggest that these features represent the most broadly bioavailable substrates in each treatment. Taken together, the presence/absence results indicate that variability between the different cultures and treatments was high and that each model community, with its unique taxa and functions is upkept by the chemodiversity of DOM.

We observed that heterotrophs in M-DOM cultures consumed more lipids, organic acids and organoheterocyclic compounds, whereas M-co cultures consumed more organic oxygen compounds (Table S4). Taxa in L-DOM cultures consumed less compounds that could be resolved and classified with the methods used and may rely more on the unresolved compounds typical for allochthonous DOM (Figure 2F and Table S4). We further calculated the nominal oxidation state of carbon (NOSC) for the features that were consumed and those that remained in all biological replicates for the respective treatment (Figure 5D-E). The mean and range of values was rather similar between feed types for consumed features but differed for the features that remained in all cultures between M-DOM/M-co cultures and the L-DOM cultures. The remaining features L-DOM cultures were substantially more oxidized than those remaining in the M-DOM or M-co cultures.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Relative oxidation state of labile and stable features

We expected most of the highly bio-labile features to be more reduced, and more stable features to be more oxidized, as more energy should be available from oxidation of reduced species (Chen et al 2021). The only treatment where this pattern was found was in L-DOM, where the features consumed in all replicates vs. stable in all replicates had a median NOSC of -0.96 and -0.4 respectively (Figure 5D-E). Generally speaking, and particularly in the treatments fed cyanobacterial DOM or co-cultures with Microcystis, the NOSC varied widely with many features having values -1.5 to 0, with higher variance in stable features compared with those consumed. This implies a diversity of metabolites with differing functionalities both in the bio-labile and stable groups.

Limitations and outlook

Here we report on heterotroph use of DOM in inoculated cultures amended either with cyanobacterial metabolites or the natural organic matter of lake water. The observed LCMS features derive from dissolved organic compounds in the samples that are either present in the initial feeds, produced by autotrophs or resulting from biomolecular transformation processes during the experimental incubations. Such DOM mixtures are extremely complex, and resolvable features such as those reported in the present study only represent a fraction of the total DOM. Additionally, the more labile low molecular metabolites autotrophically produced during the incubations may not be detected in chemical assays such as the LCMS analysis used here, simply because of their rapid assimilation and use while they can still be among the most important heterotrophic substrates.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Additionally, the bulk of DOM in freshwater lakes is unresolvable alicyclic carboxylic acids and lignin derived polymers that have limited value as heterotrophic substrates and originate from incomplete breakdown and molecular transformation of terrestrial organic matter in soils (Hertkorn et al 2006, Patriarca et al 2020a, Zherebker et al 2017). We found that many organic matter features in our dataset were resistant to removal across the relatively short timescale of our experiment, but this does not exclude the possibility that such recalcitrant DOM may degrades slowly over many months to years in natural inland waters (Koehler et al 2012, Mostovaya et al 2016). In conclusion, our approach is useful to compare metabolites in the experiment with DOM found in nature because the fragmentation pattern is like a signature that would otherwise be hidden in molecular formulas.

Cultivation experiments are known to bias against cultivation of many abundant microorganisms (Lewis et al 2020, Swan et al 2013). We used a dilution-to-extinction mixed cultivation approach (Garcia 2016) and obtained some abundant microorganisms in culture. However, the cultivation success was poorer than expected. One explanation could be that all our experimental incubations were done at 20°C and while this is optimal for *M. aeruginosa* growth, it is substantially higher that in situ conditions in the lake which was 4°C in February and 10°C in April. This explains why although our cultures hosted some of the most abundant freshwater bacterial groups, there was less success in establishing viable cultures as compared to similar experiments where incubation temperature was close to *in situ* conditions (Garcia et al 2018a). Earlier work has shown that freshwater bacteria feature microdiversity with variable temperature optima (Garcia-Garcia et al 2019). It is possible that at the time of sample collection cells that were more abundant in-situ were more viable at 4°C or 10°C.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Despite these limitations, our study was able to retrieve relevant microbial subsets and characterize these with metagenomics and high-resolution mass spectrometry to find the different microbes living and feeding of different types of DOM. We found that the phycosphere heterotrophs consume more dissolved organic carbon as well as had bigger genome sizes. In general, the bigger genome sizes in heterotrophs consuming autochthonous DOM allowed them to encode for cobalamin biosynthesis pathway and thiamine salvage pathway, which are a function rare in aquatic environments (Garcia et al 2015, Giovannoni 2012, Paerl et al 2018). This points to dependencies between heterotrophs of different kinds. Moreover, we different genes encoding for degradation of aromatic compounds, dioxygenases or beta lactamase in different heterotrophs feeding on lake DOM and primary produced DOM. And while the chemodiversity of DOM holds a niche for many different heterotrophs, we found in our study that these heterotrophs might not only depend on these chemodiversity, but also help upkeep it with their production. In our study, we confirm the high carbon consumption of heterotrophs in the phycosphere and speculate that their turnover in the microbial loop might also be high, since phycosphere microbes found to be abundant in aquatic ecosystems. Finally, in our study we cultivated thirty-three model communities, but we believe that high-throughput cultivation, sequencing and analytical chemical approaches will further reveal the roles of the diversity of microorganisms in aquatic environments.

Methods

Environmental sampling and DNA extraction

We collected environmental samples from the surface layer of Lake Erken on February 4 °C and April 2019 10 °C and DOC 10.7 mg/L. Lake water samples were collected both for

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

353 cultivation and to create a background genomic library (Supor-100 0.1 µm). Samples were
354 transported to the laboratory within one hour of collection and further processed by 0.1 µm
355 membrane filtration (Hollow fiber cartridge, GE healthcare) to produce media for the L-DOM
356 cultivations. Filters from lake water samples and harvested upscaled cultures were used to extract
357 DNA using the DNasy PowerWater kit (Qiagen) following the quick-start protocol.
358
359 *Cultivation of model communities*
360 All cells growing in model communities were collected from Lake Erken in February or
361 April 2019. To obtain simplified sub-communities we used dilution methods in 3 types of media:
362 0.1 µm filtered Lake Erken water, 0.1 µm filtrate of *Microcystis aeruginosa* that grew on $BG11_0$
363 media (Rippka et al 1979) or $BG11_0$ medium amended with actively growing *Microcyctis* cells
364 (Figure 1). The cyanobacterial cells were provided from the Pasteur Culture Collections of
365 Cyanobacteria (PCC, France). All cultures were grown with a photon flux density of 25 µmol
366 m−2 s−1 (Philips TL-D 18&30W 865 Super 80) under a 12-hour light/dark cycle at 20 °C. We
367 have two main types of cultivation setups (i) dilution of Lake Erken cells in order to cultivate
368 between 10 and 50 cells in 1 ml medium in a 96 well plate either in 0.1 µm filtrate from Lake
369 Erken (L-DOM cultures) or 0.22 µm filtrate from axenic Microcystis cultures (M-DOM cultures)
370 (Figure 1C-D), (ii) dilution of Lake Erken cells in order to grow between 10 and 10 000 cells
371 together with about 10^7 Microcystis cells for four 10% inoculum transfers in fresh $BG11_0$ media
372 (M-co cultures) (Figure 1 B). We set up dilution to extinction experiments in Feb 2019 but had
373 no growth in the wells. However, the pre-dilutions performed in February on 0.22 µm filtrate
374 from axenic Microcystis cultures were incubated at 20 °C and tested for cell count in April. One
375 of the pre-dilutions had 10^6 cells per mL and was considered pre-adapted inoculum one used to

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

set up model communities in April (Figure 1C). Another of the pre-dilutions showed 10^5 cells per mL and was considered a pre-adapted inoculum two.

The L-DOM and M-DOM cultures (Figure 1C-D) were screened for growth and selected for upscaling if cell densities were greater than 1000 cells per µl after 4 weeks incubation in 96-well plates. After selection of 50 such cultures for upscaling, these were grown in 100 ml of their respective media in Schott bottles under a 12-hour light/dark cycle at 20 °C. The upscaled cultures were monitored for growth by daily aseptic subsampling and subsequent flow cytometry cell counts (CytoFLEX, Beckman Coulter). Cultures were harvested when cell densities did not change during two consecutive days and was close to the cell density of the culture used to inoculate the upscaling. Cells were harvested by vacuum filtration onto a Supor-100 0.1 µm membrane filter (Pall, USA). M-co cultures were transferred three times about 12 days into cultivation under a 12-hour light/dark cycle at 20 °C in 10 ml volume flasks to ensure the selection of actively growing cells (Figure 1 B). M-co cultures were upscaled to 100 ml for the fourth transfer and harvested in 0.22 µm Sterivex filters (Millipore) on day 11 of the incubation.

Flow cytometry

All cell counts were performed in a CytoFLEX (Beckman Coulter). Acquisition settings, FSC 2500, SSC 2500, FITC 800, PC5.5 257. Primary threshold FITC Automatic. Flow rate 60 µl/min. Cells were stained using Syto13 (Invitrogen). Original Syto13 solution at 5 mM in DMSO. A Syto13 working solution was prepared from the original solution by diluting 1 to 200 in milliq water. Then 50 µl of sample were stained with 2.6 µl of working solution of Syto13.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Library preparations and sequencing

DNA samples were prepared in 1/10 of a volume of the standard library protocol using the Mosquito HV liquid handler from SPT Labtech (formerly TTP Labtech, Melbourn, UK). Libraries were prepared from extracted DNA of metagenomic samples in the range of 0.1 - 201 ng/µl using the Nextera Flex kit now known as Illumina DNA Prep kit (Illumina, San Diego, CA). Using the Mosquito HV (SPT Labtech, Melbourn, UK), reaction volumes were scaled down to a 1/10 of a volume compared to the standard Illumina protocol with the following exceptions: the post-tagmentation washing buffer (TWB) was added in a volume 8 ul (2 x 4 µl) and the 80% ethanol during library clean-up was added in a volume of 4,5 µl. Three µl of DNA diluted to an input of approximate 3-10 ng/µl was added to a premixed solution of bead linked transposons (BLT) and tagmentation buffer (TB1).

A 384-bottom plate magnet (SPT Labtech, Melbourn, UK) was used for all steps involving binding and washing of magnetic beads. Tagmented DNA samples with an input of 10-24 ng/µl were run using 8 library amplification cycles and those with 1- 9 ng/µl at 12 cycles using unique dual index combinations, IDT for Illumina DNA/RNA UD Indexes set A (Illumina, San Diego, CA) to mitigate index hopping. After double sided library clean-up using AMPure XP beads (Beckman Coulter, Brea, CA) libraries were resuspended in 10 µl, of which three µl was aspirated and diluted in nine µl, to account for quality control and quantification.
A subset of 20 samples were run on a bioanalyzer or fragment analyser, consistently showing an average total fragment length of 500-600 bp. Each library was quantified using the KAPA Library Quantification kit (ROCHE, Basel, CH) in three dilutions per sample for calculation of molarity and dilution before equimolar pooling of 1 nM each. Due to the low molarity in the total

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

421 volume of both pools, they were dried down prior to sequencing. The library pools were
422 sequenced on a NovaSeq 6000 SP flow cell, PE 2 x 150 bp with 10% PhiX.
423
424 *Metagenome assemblies and postprocessing*
425 The raw data was first trimmed using Trimmomatic (version 0.36; parameters:
426 ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15
427 MINLEN:36) (Bolger et al 2014). The trimmed data was assembled using Megahit (version
428 1.1.13) (Li et al 2015) with default settings. The relevant quality controlled reads were mapped
429 to all the assemblies using BBmap (Bushnell 2016) with default settings and the mapping results
430 were used to bin the contigs using Metabat (version 2.12.1, parameters --maxP 93 --minS 50 -m
431 1500 -s 10000) (Kang et al 2015). Genes of obtained MAGs were predicted using Prokka
432 (version 1.13.3, default settings) (Seemann 2014) and annotated using eggNOG-mapper (version
433 2.0.1, default settings) (Cantalapiedra et al 2021, Huerta-Cepas et al 2019). Prokaryotic
434 completeness and redundancy of all bins from Metabat and for all assembled single cells were
435 computed using CheckM (version 1.0.13) (Parks et al 2015). The MAGs were clustered into
436 metagenomic Operational Taxonomic Units (mOTUs) using mOTUlizer version (0.2.2) (Buck et
437 al 2021b) starting with 50% complete genomes with less than 5% contamination. Genome pairs
438 with ANI 95% or more were clustered into connected components. Additionally, less complete
439 genomes were recruited to the mOTU if its ANI similarity was above 95%. MAGs were
440 taxonomically annotated using GTDB-Tk (version 102 with database release 89) (Parks et al
441 2018). Mapping for the heatmap was done with Bowtie 2 for relative abundance estimates in the
442 samples ('--ignore-quals --mp 1,1 --np 1 --rdg 0,1 --rfg 0,1 --score-min L,0,-0.05') (version

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

443 2.3.5.1) (Langmead and Salzberg 2012). MAG coverage with breadth threshold of >= 50% was
444 used to confirm a true positive.
445
446 *Liquid Chromatography Tandem Mass Spectrometry analysis*
447 Metabolites were analyzed in the samples after solid phase extraction on a hydrophobic
448 sorbent at acidic pH. The extracted samples were separated and analyzed in duplicate using
449 ultrahigh performance liquid chromatography coupled to positive mode electrospray ionization
450 high resolution mass spectrometry (Orbitrap Q-Exactive). Analytes emerging from the column
451 were fragmented using a data dependent analysis routine, allowing for feature matching to
452 libraries of previously reported compounds and putative formula assignment, as detailed below.
453 Solid phase extraction was used to de-salt and concentrate analytes. Cartridges (Agilent PPL;
454 3ml, 100 mg) were rinsed with one column volume of methanol and 0.1% formic acid. Samples
455 (~5 ml), which were stored frozen in Falcon tubes, were thawed and acidified with 50µL 10%
456 formic acid containing 300ppb Capsaicin as an internal standard. The samples were loaded by
457 gravity onto the cartridges, which were then rinsed with a column volume of 0.1% formic acid
458 and dried by vacuum. The analytes were eluted with 1.3ml methanol by gravity, the methanol
459 was dried and the samples redissolved in 100 µL of 5:95 LCMS grade acetonitrile:water with
460 0.1% formic acid, containing 100 ppb Hippuric acid and Fusidic acid and 1 ppm Raffinose, as
461 internal standards.
462 Samples were injected at 10 µL into the UPLC system, and were separated on a
463 Phenomenex Kinetex C18 column (2x150 mm, 1.7 µm) at a flow rate of 0.4 ml/min. Mobile
464 phase A was 0.1% formic acid in LCMS grade water, B was 0.15% formic acid in LCMS grade
465 acetonitrile. A linear two step gradient started at 5 % B then started to increase at 30 sec from 5-

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

50% B at 7 min followed by an increase to 99% B at 10 min, a 3 min washout phase at 99% B, and a 4 min equilibration phase at 5% B (method length = 17 min).

MS1 resolution was set to 70000, MS2 resolution 17500. Data dependent analysis was set to be activated for the top 5 peaks, with a trigger range of 2-15 s and dynamic exclusion of 5 s, 1 Da isolation width and stepped collision energy at 20,30,40 eV. mzXML files were generated from Thermo .raw files with ReAdW.

Molecular formulas for the resulting MS2 features were generated using SIRIUS and ZODIAC software (Duhrkop et al 2019, Ludwig et al 2020). SIRIUS uses isotope pattern matching and MS/MS fragmentation trees (Bocker and Rasche 2008) to rank possible molecular formulas. ZODIAC uses molecular formulas generated via SIRIUS and re-ranks them based on the MS/MS network topology to each other. We used 10 ppm as default mass accuracy for both MS1 and MS/MS level and C, H, N, O, P, S as possible elements for de novo structure elucidation in addition to the molecular formulas in the biodatabases in SIRIUS.

The nominal oxidation state of carbon (NOSC) was calculated as in Equation 1:

$$NOSC = -((4*C + H - 3*N - 2*O + 5*P - 2*S)/(C)) + 4 \quad \text{Equation 1}$$

The minimum and maximum values for NOSC are -4 (e.g. CH_4) and +4 (e.g. CO_2).

Feature intensities were normalized per sample and used for sample dissimilarity calculation via the Bray-Curtis metric, which is common in ecology. The Bray Curtis dissimilarity matrix was subsequently used as the basis for a principal coordinate analysis via classical multidimensional scaling (cmdscale, MATLAB version 2017b).

Peak abundances are highly influenced by sample contents at the ESI spray, which can cause suppression or promotion, and may drift during a long analytical run. For this reason, four internal standards were added to the extracts (one before SPE (Capsaicin) and three after

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Hippuric acid, Fusidic acid and Raffinose)) in order to assess the suitability of the method for comparing feature abundances between samples. The standards were always detected, but the relative standard deviation of these four standards over 36 samples and 4 blanks (after duplicate averaging) was 40, 66, 24 and 26%, respectively. Due to this relatively large variability, which is likely due to variable suppression effects, we only consider features that were removed below or emerged above the detection limit when we mention metabolite consumption or production.

Data availability

All raw sequencing data is available under ERP124195, PRJEB37497 and PRJEB38681, and all accession numbers are found in Table S1. All raw mass spectrometry data is available through the MassIVE repository (massive.ucsd.edu) under the following accession number: MSV000086759.

References

Amon RMW, Benner R (1996). Bacterial utilization of different size classes of dissolved organic matter. *Limnology and Oceanography* **41:** 41-51.

Azam F (1998). OCEANOGRAPHY: Microbial Control of Oceanic Carbon Flux: The Plot Thickens. *Science* **280:** 694-696.

Bar-On YM, Phillips R, Milo R (2018). The biomass distribution on Earth. *PNAS* **115:** 6506-6511.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

512 Becker JW, Berube PM, Follett CL, Waterbury JB, Chisholm SW, Delong EF *et al* (2014).
513 Closely related phytoplankton species produce similar suites of dissolved organic matter.
514 *Frontiers in microbiology* **5:** 111.
515
516 Bocker S, Rasche F (2008). Towards de novo identification of metabolites by analyzing tandem
517 mass spectra. *Bioinformatics* **24:** i49-i55.
518
519 Bolger AM, Lohse M, Usadel B (2014). Trimmomatic: a flexible trimmer for Illumina sequence
520 data. *Bioinformatics* **30:** 2114-2120.
521
522 Buchan A, LeCleir GR, Gulvik CA, Gonzalez JM (2014). Master recyclers: features and
523 functions of bacteria associated with phytoplankton blooms. *Nature reviews Microbiology* **12:**
524 686-698.
525
526 Buck M, Garcia SL, Fernandez L, Martin G, Martinez-Rodriguez GA, Saarenheimo J *et al*
527 (2021a). Comprehensive dataset of shotgun metagenomes from oxygen stratified freshwater
528 lakes and ponds. *Sci Data* **8:** 131.
529
530 Buck M, Mehrshad M, Bertilsson S (2021b). mOTUpan: a robust Bayesian approach to leverage
531 metagenome assembled genomes for core-genome estimation. *bioRxiv*.
532
533 Bushnell B (2016). BBMap short read aligner: University of California, Berkeley.
534

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

535 Cai H, Jiang H, Krumholz LR, Yang Z (2014). Bacterial community composition of size-
536 fractioned aggregates within the phycosphere of cyanobacterial blooms in a eutrophic freshwater
537 lake. *PLoS One* **9:** e102879.
538
539 Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J (2021). eggNOG-
540 mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the
541 Metagenomic Scale. *Molecular biology and evolution* **38:** 5825-5829.
542
543 Chen Q, Chen F, Gonsior M, Li Y, Wang Y, He C *et al* (2021). Correspondence between DOM
544 molecules and microbial community in a subtropical coastal estuary on a spatiotemporal scale.
545 *Environ Int* **154:** 106558.
546
547 Duhrkop K, Fleischauer M, Ludwig M, Aksenov AA, Melnik AV, Meusel M *et al* (2019).
548 SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information.
549 *Nature methods* **16:** 299-302.
550
551 Eiler A, Olsson JA, Bertilsson S (2006). Diurnal variations in the auto- and heterotrophic activity
552 of cyanobacterial phycospheres (Gloeotrichia echinulata) and the identity of attached bacteria.
553 *Freshwater Biology* **51:** 298-311.
554
555 Falkowski P, Scholes RJ, Boyle E, Canadell J, Canfield D, Elser J *et al* (2000). The global
556 carbon cycle: a test of our knowledge of earth as a system. *Science* **290:** 291-296.
557

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

558 Fenchel T (2008). The microbial loop – 25 years later. *Journal of Experimental Marine Biology*
559 *and Ecology* **366:** 99-103.
560
561 Ferrer-Gonzalez FX, Widner B, Holderman NR, Glushka J, Edison AS, Kujawinski EB *et al*
562 (2021). Resource partitioning of phytoplankton metabolites that support bacterial heterotrophy.
563 *ISME J* **15:** 762-773.
564
565 Garcia SL, Buck M, McMahon KD, Grossart HP, Eiler A, Warnecke F (2015). Auxotrophy and
566 intrapopulation complementary in the "interactome' of a cultivated freshwater model community.
567 *Molecular ecology* **24:** 4449-4459.
568
569 Garcia SL (2016). Mixed cultures as model communities: hunting for ubiquitous
570 microorganisms, their partners, and interactions. *Aquatic Microbial Ecology* **77:** 79-85.
571
572 Garcia SL, Buck M, Hamilton JJ, Wurzbacher C, Grossart HP, McMahon KD *et al* (2018a).
573 Model Communities Hint at Promiscuous Metabolic Linkages between Ubiquitous Free-Living
574 Freshwater Bacteria. *mSphere* **3:** 103838.
575
576 Garcia SL, Stevens SLR, Crary B, Martinez-Garcia M, Stepanauskas R, Woyke T *et al* (2018b).
577 Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations.
578 *ISME J* **12:** 742-755.
579

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

580 Garcia-Garcia N, Tamames J, Linz AM, Pedros-Alio C, Puente-Sanchez F (2019).
581 Microdiversity ensures the maintenance of functional microbial communities under changing
582 environmental conditions. *ISME J* **13:** 2969-2983.
583
584 Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D *et al* (2005). Genome
585 Streamlining in a Cosmopolitan Oceanic Bacterium. *Science* **309:** 1242-1245.
586
587 Giovannoni SJ (2012). Vitamins in the sea. *Proceedings of the National Academy of Sciences of*
588 *the United States of America* **109:** 13888-13889.
589
590 Giovannoni SJ, Cameron Thrash J, Temperton B (2014). Implications of streamlining theory for
591 microbial ecology. *ISME J* **8:** 1553-1565.
592
593 Hahn MW, Kasalicky V, Jezbera J, Brandt U, Jezberova J, Simek K (2010). Limnohabitans
594 curvus gen. nov., sp nov., a planktonic bacterium isolated from a freshwater lake. *International*
595 *Journal of Systematic and Evolutionary Microbiology* **60:** 1358-1365.
596
597 Hahn MW, Jezberova J, Koll U, Saueressig-Beck T, Schmidt J (2016). Complete ecological
598 isolation and cryptic diversity in Polynucleobacter bacteria not resolved by 16S rRNA gene
599 sequences. *Isme Journal* **10:** 1642-1655.
600
601 Hansell D, Carlson C, Repeta D, Schlitzer R (2009). Dissolved Organic Matter in the Ocean: A
602 Controversy Stimulates New Insights. *Oceanography* **22:** 202-211.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Harayama S, Rekik M (1989). Bacterial aromatic ring-cleavage enzymes are classified into two different gene families. *The Journal of biological chemistry* **264:** 15328 - 15333.

Hentchel KL, Reyes Ruiz LM, Curtis PD, Fiebig A, Coleman ML, Crosson S (2019). Genome-scale fitness profile of Caulobacter crescentus grown in natural freshwater. *ISME J* **13:** 523-536.

Hertkorn N, Benner R, Frommberger M, Schmitt-Kopplin P, Witt M, Kaiser K *et al* (2006). Characterization of a major refractory component of marine dissolved organic matter. *Geochimica et Cosmochimica Acta* **70:** 2990-3010.

Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H *et al* (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* **47:** D309-D314.

Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* **9**.

Kang DD, Froula J, Egan R, Wang Z (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3:** e1165.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

624 Koch BP, Ludwichowski K-U, Kattner G, Dittmar T, Witt M (2008). Advanced characterization
625 of marine dissolved organic matter by combining reversed-phase liquid chromatography and FT-
626 ICR-MS. *Marine Chemistry* **111:** 233-241.
627
628 Koehler B, von Wachenfeldt E, Kothawala D, Tranvik LJ (2012). Reactivity continuum of
629 dissolved organic carbon decomposition in lake water. *Journal of Geophysical Research:*
630 *Biogeosciences* **117**.
631
632 Langmead B, Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods* **9:**
633 357-U354.
634
635 Lewis WH, Tahon G, Geesink P, Sousa DZ, Ettema TJG (2020). Innovations to culturing the
636 uncultured microbial majority. *Nature reviews Microbiology*.
637
638 Li D, Liu CM, Luo R, Sadakane K, Lam TW (2015). MEGAHIT: an ultra-fast single-node
639 solution for large and complex metagenomics assembly via succinct de Bruijn graph.
640 *Bioinformatics* **31:** 1674-1676.
641
642 Logue JB, Stedmon CA, Kellerman AM, Nielsen NJ, Andersson AF, Laudon H *et al* (2016).
643 Experimental insights into the importance of aquatic bacterial community composition to the
644 degradation of dissolved organic matter. *Isme Journal* **10:** 533-545.
645

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

646 Ludwig M, Nothias L-F, Dührkop K, Koester I, Fleischauer M, Hoffmann MA *et al* (2020).
647 Database-independent molecular formula annotation using Gibbs sampling through ZODIAC.
648 *Nature Machine Intelligence* **2:** 629-641.
649
650 Maki K, Kim C, Yoshimizu C, Tayasu I, Miyajima T, Nagata T (2009). Autochthonous origin of
651 semi-labile dissolved organic carbon in a large monomictic lake (Lake Biwa): carbon stable
652 isotopic evidence. *Limnology* **11:** 143-153.
653
654 Mondav R, Bertilsson S, Buck M, Langenheder S, Lindstrom ES, Garcia SL (2020). Streamlined
655 and Abundant Bacterioplankton Thrive in Functional Cohorts. *mSystems* **5**.
656
657 Mostovaya A, Koehler B, Guillemette F, Brunberg A-K, Tranvik LJ (2016). Effects of
658 compositional changes on reactivity continuum and decomposition kinetics of lake dissolved
659 organic matter. *Journal of Geophysical Research: Biogeosciences* **121:** 1733-1746.
660
661 Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S (2011). A Guide to the Natural
662 History of Freshwater Lake Bacteria. *Microbiol Mol Biol Rev* **75:** 14-49.
663
664 Paerl RW, Sundh J, Tan D, Svenningsen SL, Hylander S, Pinhassi J *et al* (2018). Prevalent
665 reliance of bacterioplankton on exogenous vitamin B1 and precursor availability. *Proc Natl Acad*
666 *Sci U S A* **115:** E10447-E10456.
667

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

668 Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015). CheckM: assessing the
669 quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome*
670 *Res*.
671
672 Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA *et al* (2018). A
673 standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of
674 life. *Nature Biotechnology* **36:** 996-1004.
675
676 Pascual-Garcia A, Bonhoeffer S, Bell T (2020). Metabolically cohesive microbial consortia and
677 ecosystem functioning. *Philosophical transactions of the Royal Society of London Series B,*
678 *Biological sciences* **375:** 20190245.
679
680 Patriarca C, Balderrama A, Moze M, Sjoberg PJR, Bergquist J, Tranvik LJ *et al* (2020a).
681 Investigating the Ionization of Dissolved Organic Matter by Electrospray. *Analytical chemistry*
682 **92:** 14210-14218.
683
684 Patriarca C, Sedano-Núñez VT, Garcia SL, Bergquist J, Bertilsson S, Sjöberg PJR *et al* (2020b).
685 Character and environmental lability of cyanobacteria-derived dissolved organic matter.
686 *Limnology and Oceanography* **66:** 496-509.
687
688 Rippka R, Stanier RY, Deruelles J, Herdman M, Waterbury JB (1979). Generic Assignments,
689 Strain Histories and Properties of Pure Cultures of Cyanobacteria. *Microbiology* **111:** 1-61.
690

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

691 Rodríguez-Gijón A, Nuy JK, Mehrshad M, Buck M, Schulz F, Woyke T *et al* (2022). A
692 Genomic Perspective Across Earth's Microbiomes Reveals That Genome Size in Archaea and
693 Bacteria Is Linked to Ecosystem Type and Trophic Strategy. *Frontiers in microbiology* **12**.
694
695 Salcher MM, Posch T, Pernthaler J (2013). In situ substrate preferences of abundant
696 bacterioplankton populations in a prealpine freshwater lake. *ISME J* **7:** 896-907.
697
698 Sarmento H, Gasol JM (2012). Use of phytoplankton-derived dissolved organic carbon by
699 different types of bacterioplankton. *Environ Microbiol* **14:** 2348-2360.
700
701 Seemann T (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30:** 2068-2069.
702
703 Seymour JR, Amin SA, Raina JB, Stocker R (2017). Zooming in on the phycosphere: the
704 ecological interface for phytoplankton-bacteria relationships. *Nat Microbiol* **2:** 17065.
705
706 Singh R, Parihar P, Singh M, Bajguz A, Kumar J, Singh S *et al* (2017). Uncovering Potential
707 Applications of Cyanobacteria and Algal Metabolites in Biology, Agriculture and Medicine:
708 Current Status and Future Prospects. *Frontiers in microbiology* **8:** 515.
709
710 Smriga S, Fernandez VI, Mitchell JG, Stocker R (2016). Chemotaxis toward phytoplankton
711 drives organic matter partitioning among marine bacteria. *Proc Natl Acad Sci U S A* **113:** 1576-
712 1581.
713

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

714 Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, Gonzalez JM *et al* (2013).
715 Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface
716 ocean. *Proc Natl Acad Sci U S A* **110:** 11463-11468.
717
718 Uchimiya M, Schroer W, Olofsson M, Edison AS, Moran MA (2021). Diel investments in
719 metabolite production and consumption in a model microbial system. *ISME J*.
720
721 Williams TJ, Wilkins D, Long E, Evans F, DeMaere MZ, Raftery MJ *et al* (2013). The role of
722 planktonic Flavobacteria in processing algal organic matter in coastal East Antarctica revealed
723 using metagenomics and metaproteomics. *Environ Microbiol* **15:** 1302-1317.
724
725 Yu X, Polz MF, Alm EJ (2019). Interactions in self-assembled microbial communities saturate
726 with diversity. *ISME J* **13:** 1602-1617.
727
728 Zeder M, Peter S, Shabarova T, Pernthaler J (2009). A small population of planktonic
729 Flavobacteria with disproportionally high growth during the spring phytoplankton bloom in a
730 prealpine lake. *Environmental Microbiology* **11:** 2676-2686.
731
732 Zherebker A, Kostyukevich Y, Kononikhin A, Kharybin O, Konstantinov AI, Zaitsev KV *et al*
733 (2017). Enumeration of carboxyl groups carried on individual components of humic systems
734 using deuteromethylation and Fourier transform mass spectrometry. *Analytical and bioanalytical*
735 *chemistry* **409:** 2477-2488.
736

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

We are grateful to John Paul Balmonte for helpful discussions. We thank Kamila Koprowska, application scientist at SPT Labtech for helpful technical assistance setting up the Mosquito HV protocol.

The work was primarily funded by Science for Life Laboratory, Knut and Alice Wallenberg Foundations (grant KAW 2013.0091), Kungl. Vetenskapsakademiens stiftelser (CR2019-0060) and the Swedish Research Council (grant 2017-04422). The authors would like to acknowledge support from the Genomics infrastructure services at Science for Life Laboratory (https://www.scilifelab.se) in Uppsala. The metagenomics libraries were prepared at the Microbial Single Cell facility and sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facilities are part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. The computations and data handling were enabled by resources in the project SNIC 2021/6-99 and SNIC 2021/5-133 provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX, partially funded by the Swedish Research Council through grant agreement no. 2018-05973. We thank the Deutsche Forschungsgemeinschaft for the support of D.P. through the CMFI Cluster of Excellence (EXC 2124).

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Author contribution

SLG, VTSN, MB and SB conceptualized the research idea. SLG and VTSN performed all cultivation experiments. SLG, JH, AMD, MB, JN, MM, and DP performed data analysis. SLG drafted the first manuscript. JH, MM and JJH wrote some sections of the manuscript. All authors contributed to writing and editing of the manuscript.

Competing interests

The authors declare no competing interests

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures and Supplementary Material

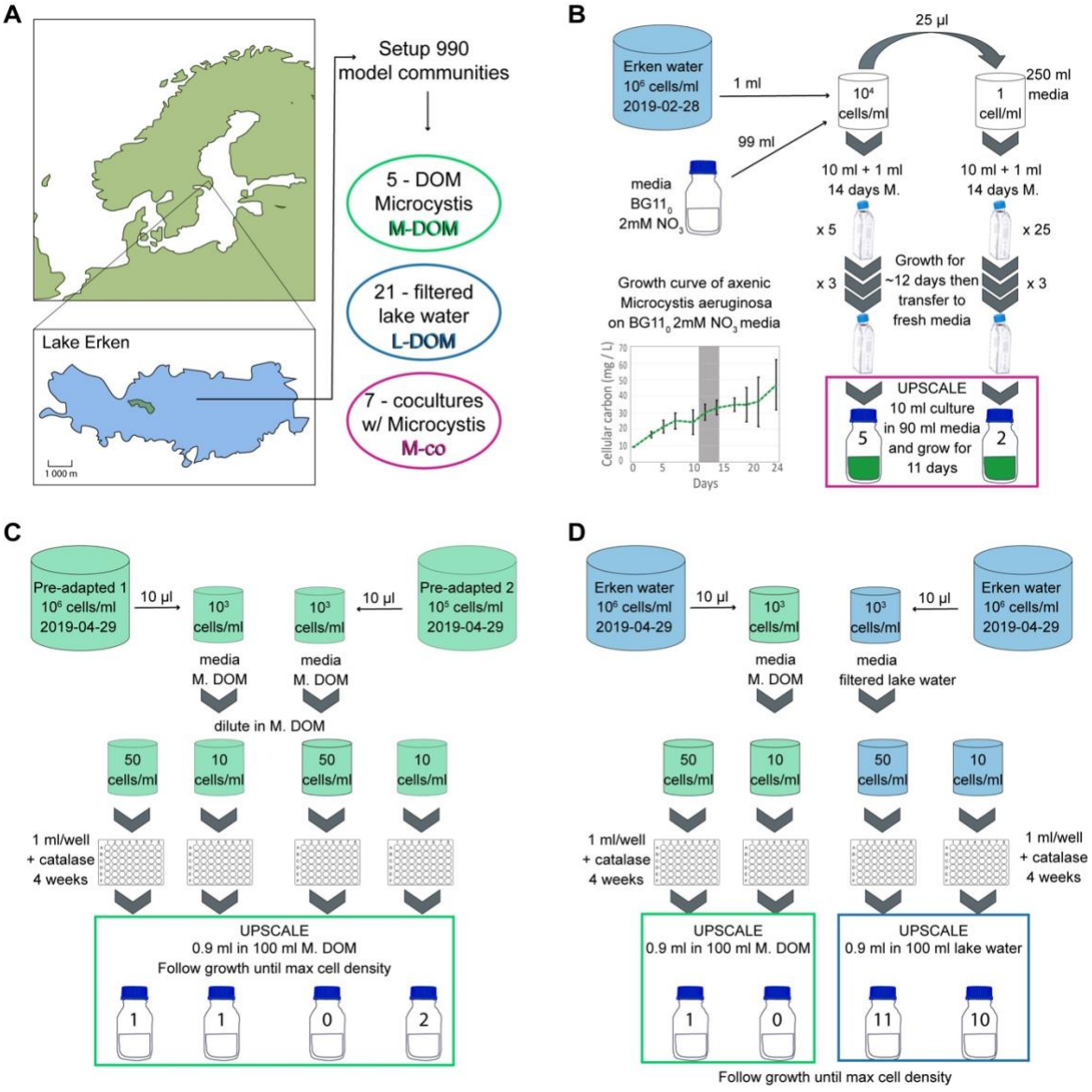


Figure 1. Overview of methods to establish model communities. Location of Lake Erken and overview of three treatments of model communities [A]. Seven model communities of Lake Erken cells grow together with Microcystis aeruginosa on $BG11_0$ 2 mM NO_3 media [B]. Pre-adaptation enrichment of cells from Lake Erken from February 2019 to M. aeruginosa DOM serve as inoculum for dilution cultivation of model communities in April 2019 yielding four model communities [C]. Cells from Lake Erken collected in April 2019 yield 21 model communities growing in Lake Erken DOM and 1 model community growing in DOM from M. aeruginosa [D].

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

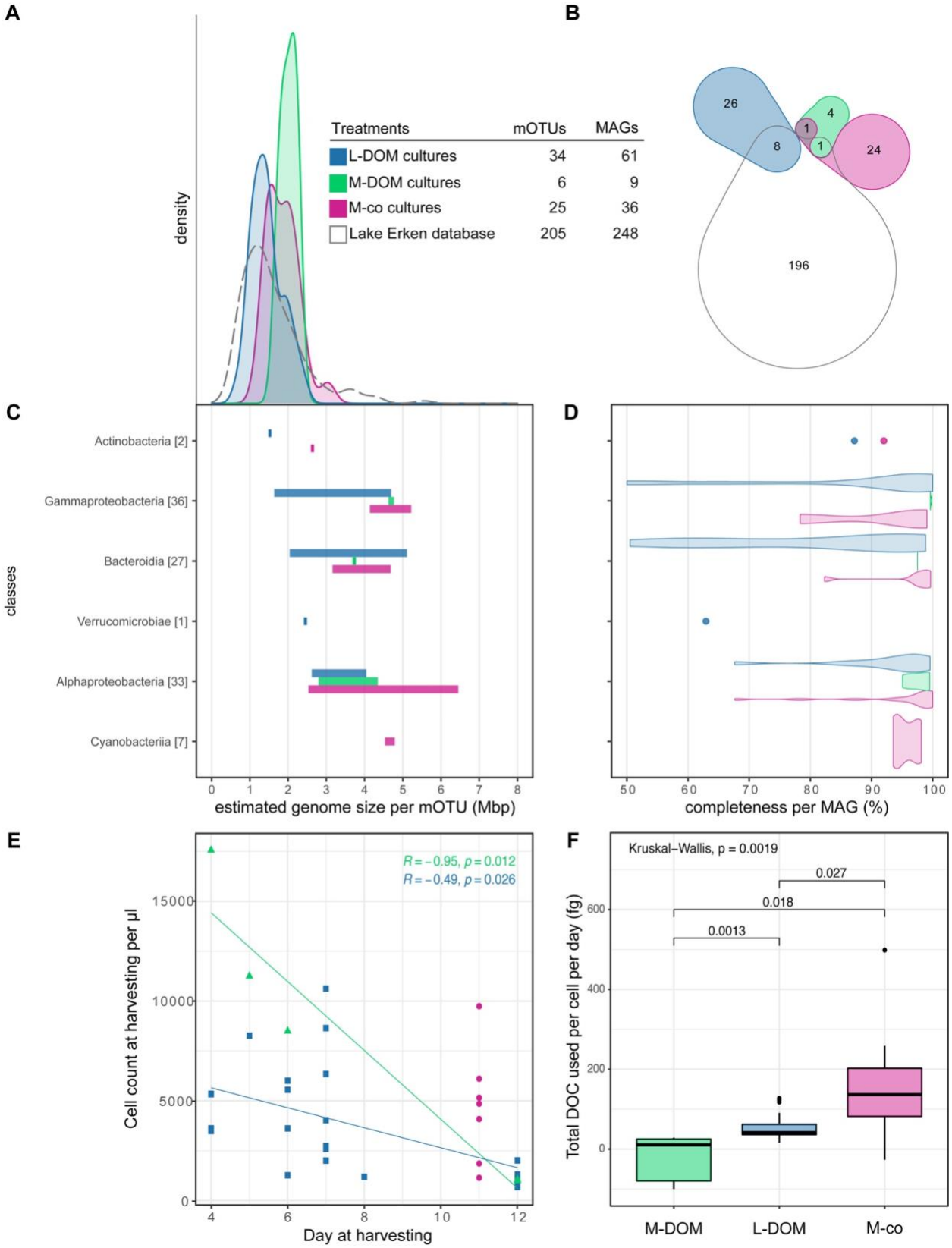


Figure 2. Overview of the genome size and growth across all model communities. Genome size distribution of bacteria in the 3 different treatments [A]. We use one representative metagenome

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

assembled genome (MAG) per metagenomic operational taxonomic unit (mOTU - defined by 95% ANI). To construct the figures, we plotted the estimated genome sizes which were calculated based on the genome assembly size and completeness estimation provided. Grey dashed line includes 6251 MAGs which are representative genomes in the biggest freshwater genome collection until date. Venn diagram of the intersection between the representative MAGs of the three treatments [B]. The intersection was calculated using FastANI (Jain et al 2018) and was determined with a threshold of 95%. Genome size distribution across different bacterial classes [C]. Completeness distribution of the different MAGs calculated using CheckM (Parks et al 2015) [D]. Final day of growth, followed by harvesting vs. number of cells at harvesting per µl of each model community [E]. Cell count in model communities in coculture with M. aeruginosa includes count of M. aeruginosa. Total dissolved organic carbon (DOC) used per cell per day in mg per µl [F].

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

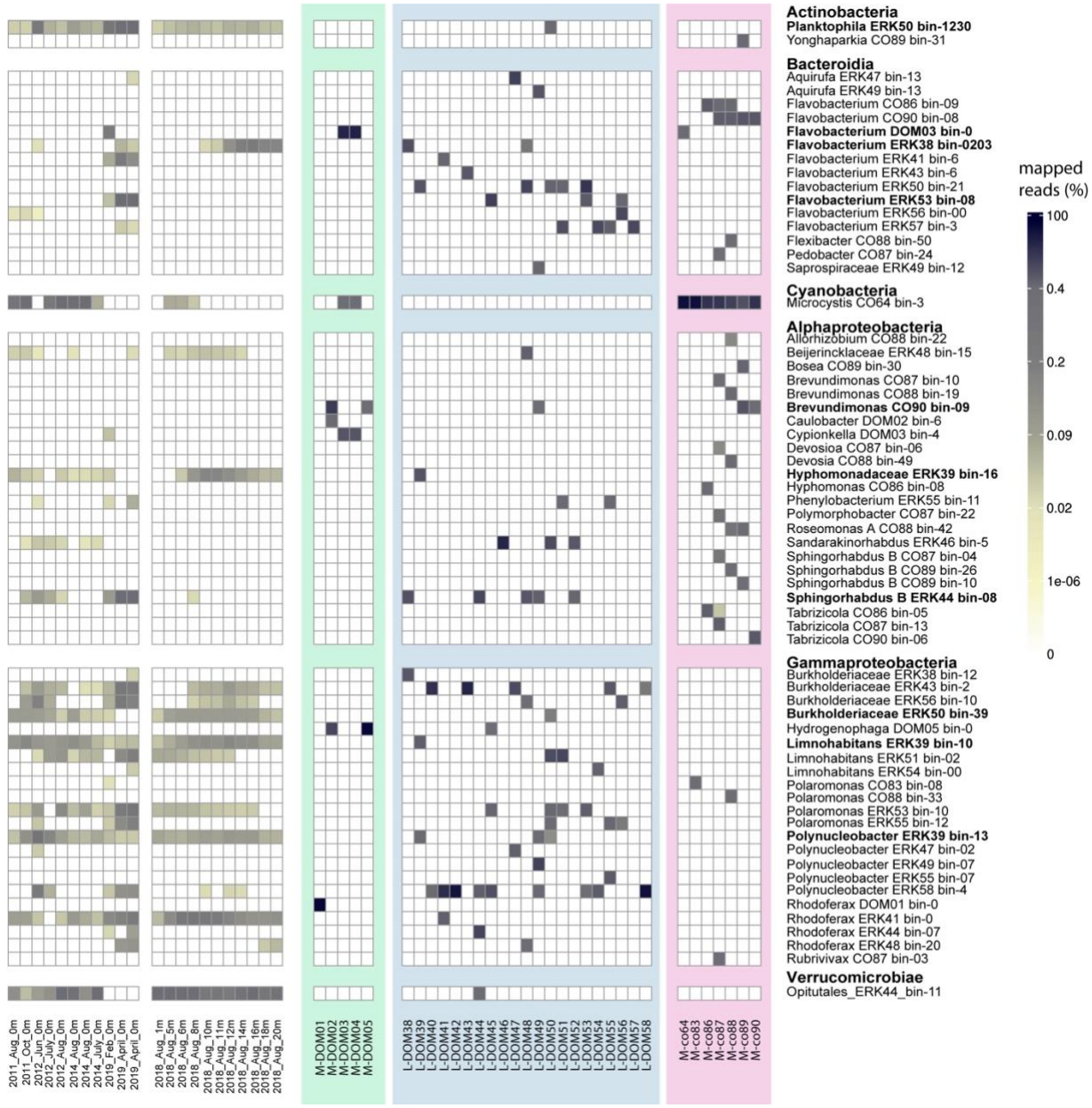


Figure 3. Relative read abundance for the 64 mOTU representatives of the 33 model communities and 22 Lake Erken metagenomes. The reads were normalized to the relative abundance of reads per metagenome. To show mapped reads, each representative MAG needed to have a breath coverage of 50%. Taxa are organized per Class and alphabetically. MAGs in bold are in the intersect between treatments or intersect with MAGs binned from the lake (Fig 2B).

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

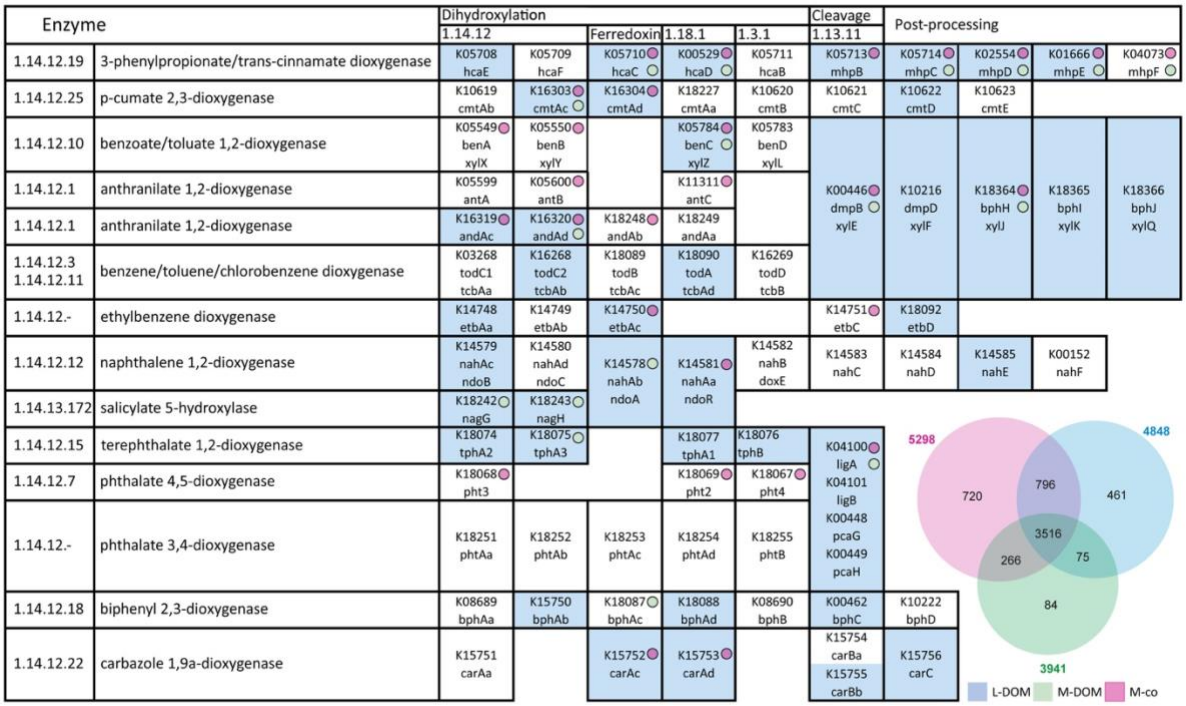
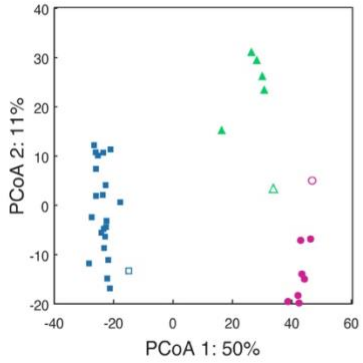


Figure 4. Distribution of dioxygenases involved in aromatic ring cleavage (KEGG brite br01602) in MAGs reconstructed from different treatments. Genes recovered from MAGs reconstructed form L-DOM cultures are highlighted in blue. Their presence in the other two treatments is also shown with circles. The Venn diagram shows the overlap of annotated functions in reconstructed MAGs from different treatments.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

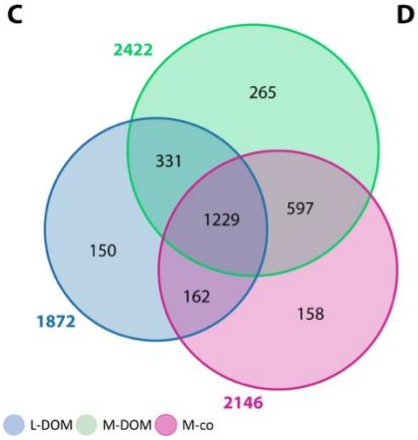
A



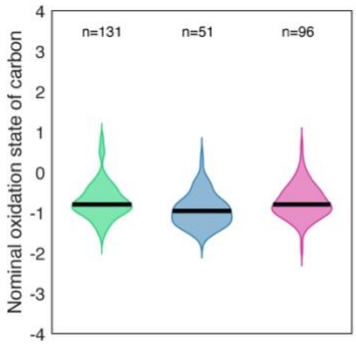
B

Number of chemical features	M-DOM cultures	L-DOM cultures	M-co cultures	Common to all three types
Initial mixture	2422	1872	2146	1229
Removed in some cultures	887	1083	676	145
Removed in all cultures	131	51	96	0
Remains in all cultures	1404	738	1374	372
Produced in some cultures	511	910	502	56
Produced in all cultures	106	3	41	0

C



D



E

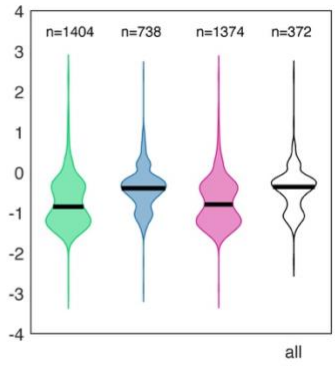
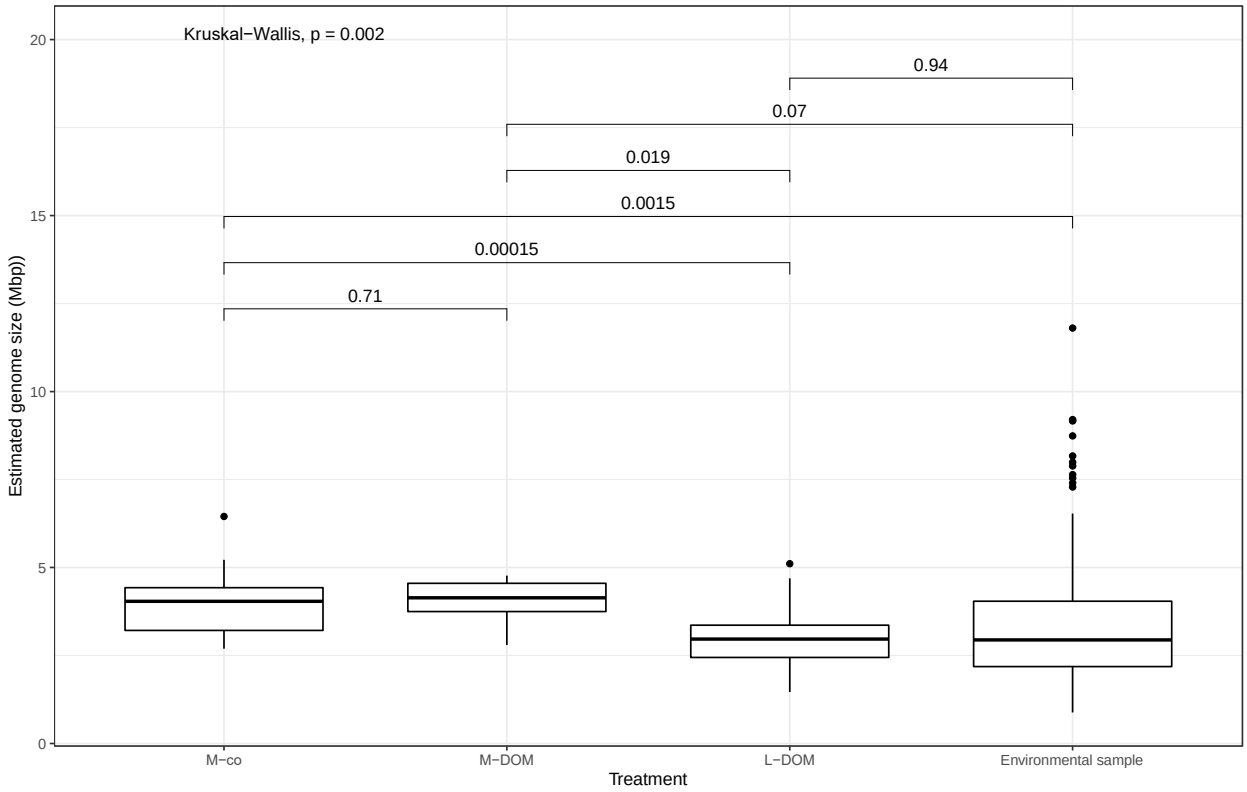


Figure 5. PCoA of the chemical composition of DOM in the model communities [A]. Filled symbols are each of the cultures, whereas empty symbols are the media control without heterotrophs. Table of number of metabolites in the initial mixture, removed in some or all cultures, remaining in all cultures or produced in some or all cultures [B]. Venn Diagram with the number of initial chemical features in the different treatments [C]. Violin plot of the oxidation state in the consumed metabolites [D]. Violin plot of the oxidation state of the metabolites that remain [E].

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

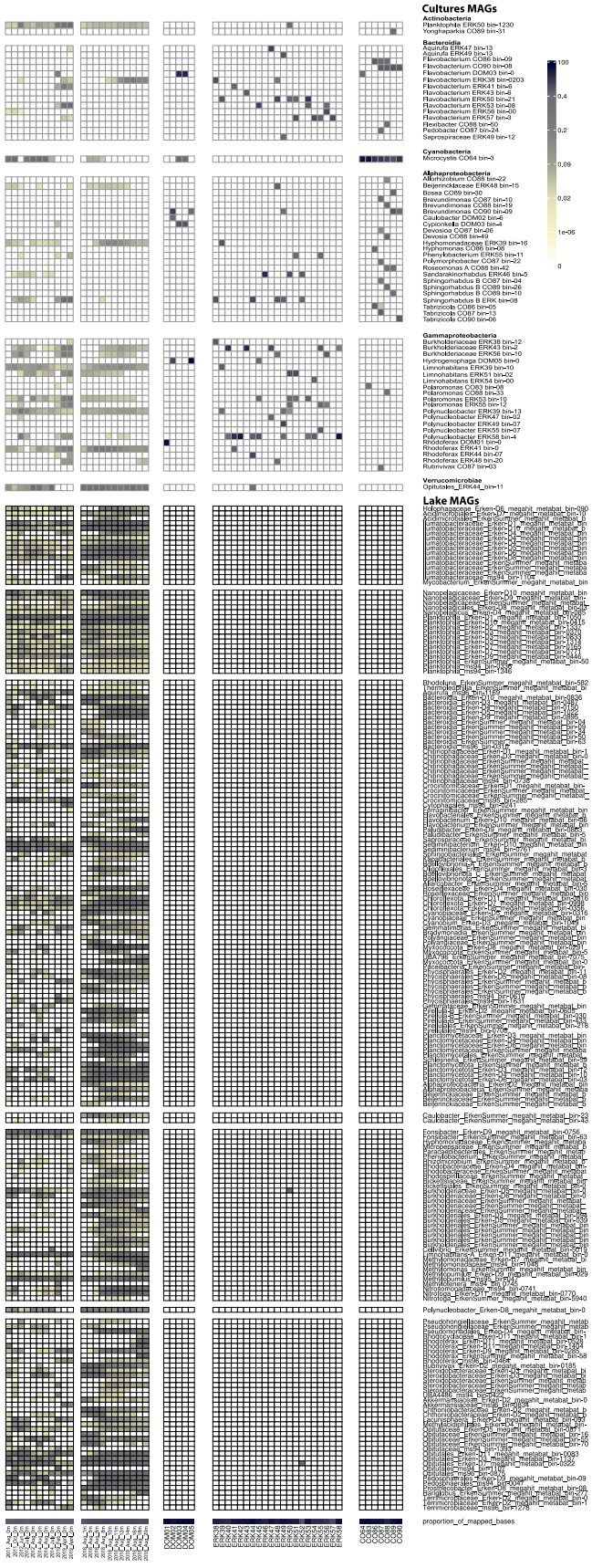
810 0.07



811

812 **Figure S1.** Average estimated genome size per treatment and Wilcoxon statistical comparison.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



813

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.21.485019; this version posted March 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S2. Relative read abundance for the 260 mOTU representatives of the 33 model communities and 22 Lake Erken metagenomes. The reads were normalized to the relative abundance of reads per metagenome. To show mapped reads, each representative MAG needed to have a breath coverage of 50%. Taxa are organized per Class and alphabetically.

Table S1. Accession numbers and metadata of 33 model community metagenomes and 3 Lake Erken metagenomes

Table S2. Information about all MAGs binned from the 33 model community metagenomes

Table S3. Information about all MAGs that were selected as representatives for the mOTUs in the study

Table S4. Compounds types removed in all cultures per treatment.