The role of land use types and water chemical properties in structuring the microbiome of a connected lakes system

Sophi Marmen[1], Lior Blank[2], Ashraf Al-Ashhab[1,3], Assaf Malik[4], Lars Ganzert[5], Maya Lalzar[4], Hans-Peter Grossart[5] and **Daniel Sher[1+]**

[1] Department of Marine Biology, Leon H. Charney School of Marine Sciences, University of Haifa, Israel; [2] Department of Plant Pathology and Weed Research, ARO, Volcani Center, Rishon Lezion, Israel; [3] Microbial Metagenomics Division, Dead Sea and Arava Science Center, Masada, Israel; [4] Bioinformatics Service Unit, University of Haifa, Israel; [5] Leibnitz Institute for Freshwater Ecology and Inland Fisheries, Department of Experimental Limnology, Stechlin, Germany

+ Corresponding author: Dr. Daniel Sher, Department of Marine Biology, Leon H. Charney School of Marine Sciences, University of Haifa, Haifa 31905, Israel.

email: dsher@univ.haifa.ac.il

Running title: Land use and lake microbiome structure

Summary

Aquatic microbial populations are intimately connected with the environmental conditions within their host water body, yet how they are affected by the surrounding land use (terrestrial-aquatic coupling) or by connectivity and transport between water bodies is less clear. We address these questions through an analysis of the land use surrounding 46 inter-connected lakes located within 7 drainage basins in northern Germany, and the microbiomes of these lakes during early summer. Land use within the drainage basin of each lake could predict, to some extent, NO_2+NO_3 and chlorophyll concentrations in the water, and these parameters (together with temperature and total phosphorus) affected water microbiome structure. In contrast, the specific drainage basin or the land use directly adjacent to the water body had little observable effect on water quality or the microbiome. Bacterial distribution was not limited by distance, suggesting high connectivity. Several microbial lineages, including cyanobacteria and verrucomicrobia, were differentially partitioned between lakes with different environmental conditions, attesting to the importance of local factors in determining the microbiome structure. As data on land use (e.g. from remote sensing) increases, identifying relationships between land use and aquatic microbial communities is a critical step in developing appropriate water management strategies.

Keywords: microbiome; lakes; water parameters; land use; distribution; Cyanobacteria

1. Introduction

The world's population growth is maintained by a suite of human enterprises such as agriculture, industry, fishing and international commerce (Vitousek*, et al.*, 1997). The use of land for such enterprises has greatly altered the way ecosystems interact with surrounding land, atmosphere and aquatic systems (Vitousek*, et al.*, 1997). Freshwater ecosystems, in particular, may be highly sensitive to human impact, mainly through eutrophication.

The microbial communities ("microbiomes") living in freshwater ecosystems are intimately connected with water quality. Microbes take up, utilize and recycle elements such as carbon, nitrogen, phosphorus and sulfur, and thus control and modulate local and global elemental cycles (Newton*, et al.*, 2011). Some aquatic microorganisms are potential pathogens of animals or humans, or can produce secondary metabolites or toxins (Abdelzaher*, et al.*, 2010, Cabral, 2010, Valério*, et al.*, 2010). Thus, significant effort has been invested in elucidating the environmental drivers that determine the composition and function of freshwater microbiomes (Newton*, et al.*, 2011). Environmental drivers of microbial community structure may act at the local scale, where the physical-chemical characteristics of the water lead to a "species sorting" process which enrich for bacteria adapted to occupy that specific niche (Jones and McMahon, 2009, Langenheder and Székely, 2011). Such physical - chemical properties can include, for example, nutrient concentrations, pH, temperature, salinity and light intensity (Crump*, et al.*, 2004, Horner-Devine*, et al.*, 2004, Yannarell and Triplett, 2005). Neutral assembly processes also play a role in determining community structure (Newton*, et al.*, 2011). Local processes can be affected by processes occurring at larger spatial and temporal scales. For example, physical connectivity between water bodies (e.g. within the same drainage basin) may lead to bacterial transfer between environments, a process termed "mass effect'', (Quinn*, et al.*, 1997, Lindström and Langenheder, 2012). In very strongly connected habitats, dispersal rates may be so high that

they lead to ecosystem homogenization (Leibold and Norberg, 2004). As the spatial scales increase, geographic distances and physical barriers may lead to limitations on bacterial dispersal, leading to “distance decay” patterns, whereby the difference between microbial communities increases with distance (Lear*, et al.*, 2013, Marcelino*, et al.*, 2017). Additionally, runoff from terrestrial ecosystems can affect water quality and the resulting microbiome (Witt*, et al.*, 2012). As the various processes affecting freshwater microbial ecosystems are identified and characterized, many questions remain: to what extent are local conditions within water bodies affected by those in the surrounding terrestrial system? What is the relative importance of local conditions in determining microbial community structure? What is the role of mass transport of microbes from upstream sites? Are all microbial taxa affected in the same manner by these processes? Answering some of these questions is critical as we aim to obtain predictive models of microbial ecosystem structure and function that can be used for science-based management of water resources.

In this study, we investigated the microbiomes of 46 lakes in Germany, located in a network of 7 drainage basins connected through streams. Our major goal was to understand what shapes microbial population in a system of connected natural water bodies, located within relatively short distances and surrounded by a variety of different land-use types. We analyzed the role of local water properties and land use at different geographic scales (basin-level and adjacent to the water body) to try and explain bacterial community structure, and investigated the association between land, specific water properties and specific bacterial lineages.

2. Results

89 2.1 General characterization of the sampled water bodies

90 During the summer of 2015, we sampled 46 water bodies (lakes and one fish pond) within a
91 relatively small area (~150 km by 100 km) of the lake-rich Brandenburg/Mecklenburg region in
92 northeastern Germany (Fig. 1, Fig. S1, Supplementary Excel table). The water bodies belonged
93 primarily to five major drainage basins, and varied widely in their sizes, nutrient concentrations
94 and trophic levels (Supplementary Excel table). While the microbial diversity, expressed as the
95 Shannon alpha-diversity index, was relatively similar between the basins (Fig. 2a), clear
96 differences were observed in community composition between the various sampling sites (Fig.
97 2b). Comparing the microbiome composition between the different sites (using the Bray- Curtis
98 dissimilarity index) revealed that sites can be broadly divided into two main groups: a cluster of
99 sites with a high relative abundance of Cyanobacteria, and a second cluster with a high relative
100 abundance of Actinobacteria, Proteobacteria and Bacteroidetes and low relative abundance of
101 Cyanobacteria (Fig. 2b). Interestingly, lakes from the same drainage basins did not cluster
102 together, despite relatively large differences between the drainage basins in terms of their size
103 and the land-use in their water shed (Fig. 2b, Fig. S2a, Supplementary Excel table). Neither was a
104 clear correlation observed between the size of the drainage basin of each water body and the
105 chemistry of the water (Fig. S3a-e) or the lake microbiome (Fig. S3f, g).

106 One major factor that might determine to what extent ecological communities differ, is the
107 geographical distance between them. This phenomenon, identified by a ''distance decay ''
108 relationship in community similarity (e.g. Nekola and White, 1999, Bell, 2010), has been
109 demonstrated for both macro- and micro-organisms at various scales of geographic distances
110 (Lear, *et al.,* 2013, Mercelino, et al.,2017). No distance decay relationships were observed in our
111 samples, neither within a single basin nor between samples belonging to different basins (Fig.
112 2c). Moreover, examination of the total species diversity in our landscape (i.e. γ-diversity)

113 showed that 27% of the OTUs are core OTUs that were present among all sites, and that
114 additional 22% OTUs were cosmopolitan, defined as being observed in >90% of the sites (Fig.
115 2d). Rare and site unique OTUs combined represented only 3.4% of all OTUs. These rare and
116 unique OTUs were also numerically scarce, representing at most 2.6% of the total sequence
117 reads of the total reads. The core microbiome was dominated by Actinobacteria and
118 Proteobacteria (43% and 39% of the OTUs respectively, Fig. S4a), and, surprisingly, contained
119 very few Cyanobacterial OTUs (1%). In contrast, Cyanobacteria was the dominant phylum in the
120 cosmopolitan OTUs (47%). Thus, it seems that the differences between the microbial
121 communities are primarily due to changes in the relative contribution of OTUs rather than to the
122 presence of unique or endemic OTUs in different regions.
123
124 2.2 Land use in the drainage basins affects water chemistry, and through it, the microbiome
125 We next asked whether environmental conditions within the water body (e.g. nutrients,
126 temperature etc', referred to here as “aquatic environmental conditions”), or the surrounding
127 land use, may affect water microbiomes. We hypothesize that the effect of land use could be
128 manifested at the level of the entire drainage basin (summing up the input of the entire region
129 draining into each water body), or, alternatively, at the level of local land use (i.e. the land
130 adjacent to the water body, defined here as a distance of 500 meters from lake shore). To test
131 these hypotheses, we first performed variation partition analysis to determine the explanatory
132 ability of land use types on the aquatic environmental conditions. As shown in Fig. 3a, the land
133 use of the drainage basins explained almost 30% of the variation in aquatic environmental
134 conditions, whereas the local land use had no discernable explanatory power of its own.
135 Nevertheless, combining local and basin land uses explained more of the variability than the
136 basin level land use alone. We next used GLMs to determine whether specific land use types

137 could be associated with specific aquatic environmental conditions. On the basin scale, six land
138 use types explained a high proportion of the variability of two specific variables: Chl *a* (RN^2=
139 0.98) and nitrite and nitrate concentrations (RN^2= 0.74, Fig. 4, Tables S1 and S2). For each of
140 these two models, one land use stood out in having a profoundly larger importance value: the
141 fraction of urban land use explained NO_3+NO_2 concentrations, whereas the fraction of forests
142 and natural areas cover explained Chl *a* (Fig. 4). The relationship between these two land use
143 types and the respective aquatic environmental parameter was positive, i.e. in basins with a
144 high proportion of forests and natural areas, Chl *a* concentration in the water bodies was higher,
145 while in more urban areas, nitrate and nitrite concentrations were higher (Table S1 and S2).
146 When quantifying the independent contribution of each different land use separately, the
147 variation in Chl *a* was significantly explained by arable land (25% independent contribution) and
148 forests and natural areas (29%), whereas urban land use was the only significant variable
149 explaining 68% of the variation in NO_3+ NO_2.

150 We next used RDA to explore to what extent basin-level and local land use types, as well as the
151 aquatic environmental conditions themselves, explain the composition of the aquatic
152 microbiome. As shown in Fig. 3b, aquatic environmental parameters were able to explain 19% of
153 the variation in microbiome compositions among all sites. However, land use types (at both
154 basin and local scale) had a relatively weak direct effect on the microbial community structure
155 (3.4%, Fig. 3b).

156

157

158 <u>2.3 Specific bacterial taxonomic groups are associated with sampling locations differing in their</u>
159 <u>aquatic environmental conditions</u>

160 If aquatic environmental conditions can explain a significant amount of variability in the
161 microbiome structure, which environmental conditions are most important for different
162 bacterial lineages? Canonical Correspondence Analysis (CCA) revealed that four aquatic
163 environmental parameters significantly affected the microbiome: i) concentration of $NO_3 + NO_2$
164 ($p < 0.01$), ii) total phosphorus ($P < 0.001$), iii) water temperature ($p < 0.0001$) and iv)
165 concentration of Chl *a* ($p < 0.0001$) (Fig. 5a-b). Notably, two of these parameters, ($NO_3 + NO_2$
166 and total Chl *a*), were identified by the GLMs as being significantly affected by land use in the
167 drainage basins.
168 We next analyzed OTU distribution patterns for different phyla by CCA ordination, in which the
169 location of each OTU on the plot corresponds to the "center of mass" of its distribution among
170 sites (Fig 5c-g). Thus, OTUs that are more abundant at specific sites will occur closer to these
171 sites in the ordination and OTUs that are found in all samples at similar relative abundance will
172 be located towards the center of the plot. Of the five most abundant phyla tested, three
173 (Actinobacteria, Proteobacteria and Cyanobacteria), were not randomly distributed in the CCA
174 plot, as determined by permutation analysis (Fig. 5 c-g, $p < 0.05$). Actinobacterial OTUs clustered
175 tightly and close to the intersection of the axes, slightly on the negative side of the CCA1 axis
176 (Fig. 5c). This suggests that, while the relative abundance of Actinobacteria was different
177 between different sites, within this phylum the Actinobacterial OTUs behaved similar, i.e. they
178 are found at similar ratios at each location (Fig. 2c). Proteobacteria showed wider, but not
179 random distribution (Fig. 5d). These results are in agreement with our observation that these
180 two phyla dominate the core microbiome of all sites (Fig. S4a), and that within each of these
181 phyla, a relatively large fraction of OTUs were core or cosmopolitan (Fig. S4b). In contrast,
182 Verrucomicrobia and Bacteroidetes were randomly distributed in the CCA plot and not
183 associated with any specific water parameter, i.e. the center of mass of the phylum was not in

184 the direction of any of the arrows representing the eigenvalues of the environmental conditions
185 (Fig. 5f, g). Unlike the above four phyla, which showed a generally symmetrical shape in the CCA
186 plots, Cyanobacteria were not symmetrically or evenly distributed (Fig. 5e). While Cyanobacteria
187 appeared in all sampling sites (Fig. 2b), only few Cyanobacteria OTUs were part of the core
188 microbiome (Fig. S4a), and this phylum had the highest fraction of OTUs in the ''common''
189 category among all analyzed phyla (Fig. S4b). This prompted us to explore whether patterns
190 could be observed at higher taxonomic resolution (class or order). Indeed, a clear partitioning
191 could be seen between Synechococcales, single-celled Cyanobacteria which appeared mostly
192 among sites with low phosphorus concentration (i.e. the negative side of CCA1), and Nostocales
193 as well as Pseudanabaenales (Fig. 6a). Nostocales were associated with high TP and Chl *a*
194 concentrations, whereas Pseudanabaenales were also associated with high nitrite and nitrate
195 levels. Differences in the ordination of groups at the class level were also observed within the
196 phylum Verrucomicrobia, with Methylacidiphilae clearly separated from most of the other
197 orders (Fig. 6b). No clear differences were observed at the class, order or family level in
198 Proteobacteria, Bacteroidetes or Actinobacteria (Fig. 6c-e)
199

200 **3. Discussion:**

201 In this study, we analyzed the microbiomes of a network of partly-interconnected natural lakes,
202 which belong to several different drainage basins. We aimed to identify to what extent the land-
203 use in the drainage basins of the lakes affects both the environmental conditions in the water
204 and the microbiome. Several previous studies have identified correlations between land-use and
205 various parameters of water chemistry, as well as to the abundance of specific members of the
206 microbial community (Savage*, et al.*, 2010, Katsiapi*, et al.*, 2012, Beaver*, et al.*, 2014, Davis*, et*
207 *al.*, 2015). However, few studies have attempted to determine whether land use controls the

208 microbiome structure as a whole (Glöckner*, et al.*, 2000, Van Rossum*, et al.*, 2015). Importantly,
209 in our analyses, we used tools that are relatively accessible to researchers and water quality
210 managers worldwide: land-use maps are available for many regions of the world, and are often
211 used for land management using GIS (Joerin*, et al.*, 2001, Zhang*, et al.*, 2015). The water quality
212 measurements were performed using hand-held instruments. Measurements of inorganic
213 nitrogen compounds and total phosphorus were performed using commonly-applicable and
214 standardized methods (Grasshoff and Kremling, 1983). We chose this experimental approach in
215 order to facilitate comparisons between this study and future analyses terrestrial-aquatic
216 coupling and connectivity in lake networks from other regions.

217 Our results suggest a conceptual model of the lakes we studied (schematically presented in Fig.
218 7) whereby two measures of water quality, namely NO_3+ NO_2 and total chl *a*, are related to
219 specific land use types within the drainage basin of each water body (urban areas and forests
220 and natural areas, respectively). These water quality parameters, in turn, together with total
221 phosphorus and temperature, are related to the relative abundance of specific microbial linages
222 at different phylogenetic resolutions. Below we first discuss specific links between land use
223 types, water quality, and microbiome composition, as well as the association between specific
224 lineages and water quality parameters. We then mention the caveats of our study and suggest
225 ways to improve future studies.

226

227

228

229 <u>3.1 What determines the microbiome - land use, mass transfer or biogeography?</u>

230 Our initial hypothesis was that two types of factors would structure the microbial communities:
231 i) environmental conditions within the water body, which, in turn, are determined (at least to
232 some extent) by land use within the drainage basin of each water body; ii) mass effects, due to
233 bacterial transport along the streams connecting the various sampled lakes. For example,
234 connected lakes in Alaska shared similar bacterioplankton communities that shifted gradually
235 down the catchment (Crump*, et al.*, 2007). Other studies suggest an influence of water retention
236 time on community composition among connected lakes (Mašín*, et al.*, 2003, Lindström and
237 Bergström, 2004). Our results show no separation of bacterial communities by basin. Nor did we
238 identify any distance-decay patterns, which have been observed in other ecosystems at
239 different scales (Fig. 2, (Lear*, et al.*, 2013, Marcelino*, et al.*, 2017)). Thus, in our study system,
240 mass transfer or limitations on microbial dispersal are not strong determinants of microbial
241 community composition, at least not at the time of sampling. Instead, environmental conditions
242 within each water body strongly affected the microbiome structure, and these were partly
243 determined by land-use in the drainage area of each water body (Fig. 3).
244 Several studies have investigated the association between land use types and water properties
245 (Quinn*, et al.*, 1997, Larned*, et al.*, 2004, Mehaffey*, et al.*, 2005, Schoonover and Lockaby, 2006,
246 Reimann*, et al.*, 2009). These studies generally suggest higher dissolved nitrogen and
247 phosphorus concentrations in waters surrounded by high coverage of urban areas and
248 agriculture lands (Wang and Yin, 1997, Ghai*, et al.*, 2014). In our study, urban land was an
249 important predictor (and positively correlated with) the concentration of NO_2+NO_3, whereas
250 forests and natural areas were positively correlated with Chl *a* concentrations (Fig. 4, Tables S1,
251 S2). The correlation between urban area (which includes also industry) and NO_2+NO_3
252 concentration is in agreement with the previous mentioned studies, and may be driven by the
253 discharge of waste and rain run-off from sealed surfaces, both containing high nitrogen levels. In

254 contrast, the positive association between forests and natural areas and Chl *a* concentration
255 (Fig. 4b) is unexpected. Generally, it has been suggested that phytoplankton blooms in lakes are
256 driven by excess nutrients such as nitrate and phosphate which are commonly used in
257 agricultural fertilizers (Carpenter*, et al.*, 1998, Xu*, et al.*, 2010). Thus, we would expect a positive
258 association between Chl *a* and agricultural or urban land, and perhaps a negative relationship
259 with forests and natural areas, presumed to be more pristine. Indeed, a significant negative
260 correlation was found between the percentage of forests and natural areas cover and both total
261 phytoplankton and Cyanobacterial biomass among water reservoirs at the Ile-de-France region
262 (Catherine*, et al.*, 2008). At this stage, we cannot provide an adequate explanation to our
263 observation, however, we note that the methodology used to infer the drainage basin of water
264 bodies is different between different studies (e.g. the study of the Ile-de-France region
265 (Catherine*, et al.*, 2008) used "hydrographical zones", which are much smaller due to their non-
266 cumulative nature than the drainage basins we have defined). Alternatively, the difference may
267 be due to differences in the regions in the type of soil, plants or agricultural practices.
268 While we were able to explain, to some extent, the water chemistry via land-use types, and the
269 microbiome composition via water chemistry, we were unable to identify direct links between
270 land use and the specific composition of the microbiome. This may be simply due to the limited
271 amount of variability in aquatic environmental parameters explained by the land use (~30-40%,
272 Fig. 3), and the limited explanatory power of the aquatic environmental conditions themselves
273 on the microbiome (~19%). An additional interpretation of these results is that while land use in
274 the drainage basin of each lake affects its microbiome indirectly, e.g. by impacting water
275 chemistry, there is a relatively low direct effect of land use on the microbiome, e.g. by being a
276 source for specific (presumably non-aquatic) microbes. In support of this interpretation,
277 relatively few OTUs belonged to Acidobacteria (Fig. S4a) regarded as a representative phylum of

278 soil bacteria (Janssen, 2006). Among these, only 0.04% of the total reads belonged to the
279 *Solibacteres*, which have been considered as marker soil bacteria (Battistuzzi and Hedges, 2009).
280 In the study of water bodies in Ile-de-France discussed above (Catherine*, et al.*, 2008), land use
281 was able to explain, to some extent, the Cyanobacterial biomass. Similarly, a land-use based
282 model was able to predict, to some extent, the presence of one of the genes involved in the
283 biosynthesis of the Cyanobacterial toxin microcystin (Marmen*, et al.*, 2016). In our dataset, even
284 though the total fraction of Cyanobacteria is correlated with Chl *a* (R^2 = 0.3 Fig. S5), and the
285 main groups of Cyanobacteria partition clearly into different niches based on the water
286 chemistry (Fig. S5, Fig. S6), we were unable to identify any model based on land-use that could
287 explain the variation in the relative abundance of these specific Cyanobacterial groups. Thus,
288 while the potential of land-use to predict the abundance of Cyanobacteria and their toxins
289 (parameters which are of importance to water managers) has been shown, further work is
290 needed to develop the appropriate methodologies and statistical tools that will identify linkages
291 between land use types and microbiome structure. Given the good accessibility of land-use data,
292 which for much of the globe (including third world countries) can be derived from remote
293 sensing, developing these methods should be of high priority for future research (Ndlela*, et al.*,
294 2016).

295 <u>3.2 Niche differentiation of freshwater bacterioplankton at the class and order levels</u>

296 An important question in microbial ecology is whether there is a phylogenetic resolution at
297 which environmental conditions can be best related to microbiome structure and function.
298 Many studies focus on the phylum level (e.g. Lindström*, et al.*, 2005, Briée*, et al.*, 2007, Wolz*, et*
299 *al.*, 2018), since, at a broad scale, bacterial phyla differ in some basic functional traits including
300 their ability to photosynthesize, or to utilize different carbon sources (Mitsui*, et al.*, 1986, Camp*,*
301 *et al.*, 2009, Newton*, et al.*, 2011)). Indeed, a clear differentiation could be observed, in our

302 dataset, between two groups of sites, those with a high relative abundance of Cyanobacteria,
303 and those with a high relative abundance of Actinobacteria and Proteobacteria (Fig. 2B).
304 However, it has been suggested that this resolution is far too coarse to reliably infer community
305 function, due to the high genomic and functional variation within phyla and even within
306 individual bacterial lineages (Newton*, et al.*, 2011, Singer*, et al.*, 2016). Our results suggest that
307 some bacterial lineages clearly partition between different environments already at low
308 phylogenetic resolution, namely that of bacterial classes and orders. Several of these patterns
309 were in agreement with our previous expectations. For example, in freshwater environments, it
310 is generally accepted that the relative contribution of pico-Cyanobacteria (e.g. Synechococcales)
311 to total phytoplankton decreases with higher trophic status. However, temporal patterns can be
312 also important (Fukushima*, et al.*, 2017). This may be attributed to the ability of pico-
313 Cyanobacteria to acquire major nutrients and trace metals at sub-micromolar concentrations, a
314 capability that enable them to occupy relatively oligotrophic niches (Callieri, 2017). Larger
315 phytoplankton (e.g. Nostocales) fills in niches with higher nutrient concentrations, and hence is
316 physiologically adapted to proliferate at rich nutrient waters like fishponds and hypertrophic
317 reservoirs (Cirés and Ballot, 2016). The presence and density of Nostocales are of great interest
318 from a water quality perspective due to their ability to produce a variety of secondary toxic
319 metabolites (Sukenik*, et al.*, 2012). In our samples, a clear separation could be observed in the
320 CCA analyses between Synechococcales, associated with conditions of low chl *a* and total P, and
321 Nostocales, associated with the opposite conditions (Fig. 6a). Pseudoanabaenales, which are
322 filamentous Cyanobacteria whose cells are smaller than those of the Nostocales*,* were
323 associated also with high NO_3+NO_3 concentrations. The physiology and ecology of
324 Pseudoanabaenalesm are relatively poorly known (Acinas*, et al.*, 2008), yet these organisms are

325 quite dominant in many of the lakes we sampled (Fig. S5), providing an incentive for further
326 study.
327 Potential niche differentiation was also observed within the phylum Verrocumicrobia, which is
328 ubiquitous but relatively less studied in freshwater environments (Fig. 6b, Newton*, et al.*, 2011).
329 Some members of this phylum were found to be associated with high nutrient concentrations or
330 algal blooms (Eiler and Bertilsson, 2004, Kolmonen*, et al.*, 2004, Haukka*, et al.*, 2006). In our
331 study, Methylacidiphilae (and possibly also Spartobacteria) clearly separated from the other
332 three classes (Pedosphaerae, Opitutae and Verrucomicrobias), and may be associated with high
333 NO_3 and NO_2 concentrations. At least some members of the Methylacidiphilae are
334 methanotrophs (Camp*, et al.*, 2009), and one bacterium from this class (*Methylacdiphilum*
335 *fumariolicum*) is a diazotroph, able actively fix atmospheric N_2 (Khadem*, et al.*, 2010). While
336 most Verrucomicrobial methanotrophs have been isolated in extreme hot and acidic
337 environments, the association of potential diazotrophs (Methylacidiphilae as well as
338 Pseudoanabaenales) with high NO_3+NO_2 in lakes is intriguing, and their potential contribution to
339 high inorganic N concentrations is worthy of additional studies.
340 The separation between different orders in the CCA plots suggests that, in freshwater bacteria,
341 niche differentiation may be discernable at intermediate phylogenetic resolution - higher than
342 that of phyla but lower than that of specific species or strains. We also searched for specific
343 OTUs that were associated with aquatic environmental conditions or land use categories, but
344 found very few results, and when correlations were observed, they were relatively weak
345 (maximum Pearson R^2 was 0.53, Table S3). Two OTUs, one belonging to Phycisphaerales (phylum
346 Planctomycetes) and one belonging to Saprospirales (phylum Bacteroidetes), were positively
347 associated with NO_3+NO_2 concentrations. Phycisphaerales have been found in annamox
348 processes (Hu*, et al.*, 2015), and Saprospirales are associated with specific soil parameters (King*,*

349 *et al.*, 2010, Hargreaves*, et al.*, 2015), yet the role of these taxa in freshwater is mostly unclear.
350 Thus, in our dataset, no specific OTU can serve as an indicator species for specific aquatic
351 environmental parameters or land uses.
352 As the amount of genetic data (e.g. 16S rDNA surveys and metagenomes) from freshwater
353 locations rapidly increases, it will be important to explore the best phylogenetic resolution at
354 which to identify correlations between water quality and microbiome structure. Such
355 microbiome data can then be linked to the relative abundance of specific microbial groups such
356 as those involved in nutrient cycling, toxicity or pathogenesis, which in turn are important traits
357 for water quality management.
358
359 <u>3.3 Some caveats, and what can be done better next time.</u>
360 While we could explain a significant amount of the variability in the microbiome (a maximum of
361 22.7%, Fig. 3b, similar to other studies, e.g. Lear*, et al.*, 2013), a large part of the variability
362 remains unexplained. This variability may be due to other aquatic environmental parameters we
363 did not measure, for example the quantity and composition of the dissolved organic carbon,
364 which feeds heterotrophic bacteria and clearly has an effect on aquatic microbiomes (Davis*, et*
365 *al.*, 2015, Logue*, et al.*, 2015, Wei*, et al.*, 2016). Nor did we measure hydrological properties of
366 the systems such as the lake depth, turbulence, stratification or retention times of the lakes,
367 which can also strongly affect microbial population structure (Kopylov*, et al.*, 2002, Nogueira*, et*
368 *al.*, 2002, Humayoun*, et al.*, 2003, Carini*, et al.*, 2005, Lindström*, et al.*, 2005, Percent*, et al.*,
369 2008). Groundwater discharge into the lakes may also be an important driver of both
370 biogeochemistry and population structure (Griebler and Lueders., 2009). While clearly of
371 importance, often many of these hydrological parameters are unknown, including for the
372 majority of the lakes we studied, even though several of them have been studied intensively for

373 over 50 years (e.g. Allgaier and Grossart, 2006, Casper, 2012). Finally, we measured each
374 location only once, relatively early during summer. Such a snap-shot does not represent the
375 dynamic nature of the aquatic microbiome. For example, in several lakes that we studied, large
376 Cyanobacterial blooms have been recorded, particularly during August (Mantzouki*, et al.*, 2018).
377 These blooms are often dominated by *Microcystis* species, yet these were relatively rare in our
378 samples. Clearly, a combination of extended time-series and comparative analysis such as the
379 one presented here, taking into account local conditions, terrestrial-aquatic coupling and lake
380 connectivity, are needed to gain a better understanding of the processes affecting the
381 microbiome.

382

383 **4. Experimental Procedures**

384 4.1 Study region and delineation of drainage basins

385 The sampling area lies within the region of Brandenburg, in northern Germany (Fig. 1, Fig. S1,
386 Supplementary text). The area is relatively flat (18-89 meters above sea level) and contains
387 hundreds of lakes, ponds and depressions that were formed by glacial action during the
388 Pleistocene. Three main basins drain this region, which are (from west to east) the Rhin, the
389 Havel and the Ucker. The two former flow south in the region of sampling, whereas the Ucker
390 flows north. These three main basins can be subdivided into smaller basins. The drainage basin
391 of each lake was defined as the line connecting the highest points from which the surface water
392 flow enters each of the lakes, including all of the upstream lakes. We note that groundwater
393 flow may be significant in many lakes, and often does not flow as surface water does in
394 accordance with topographic basins (Winter*, et al.*, 2003). We also assume that the water is fully
395 mixed horizontally within each separate water body, and that our sampling represents the

396 epilimnion if the lakes are stratified. A detailed description of the sampling region, methods of
397 delineating the basins, identification of the canals connecting the basins and GIS layer analysis,
398 can be found in the supplementary information.
399

400 4.2. Sampling and analysis of environmental data

401 Each location was sampled once from the edge of the water body. During sampling, dissolved
402 oxygen, temperature pH and turbidity were measured using a YSI multi probe (YSI 6920 V2-2; YSI
403 Inc). At all locations, 5 liters of surface water were collected using plastic jerry-cans that were
404 washed with double distilled water then with 70% ethanol and air dried prior to sampling.
405 Samples were brought to the lab and filtered within two hours. For chlorophyll *a* (Chl *a*)
406 measurement, water was filtrated on GF/C filters (Whatman glass fiber, 25 mm, nominal pore
407 size 1.2 µm) and placed into 1.5 mL sterile Eppendorf tubes. The water was filtered until the
408 filters were clogged, and the volume was recorded (50-800 mL). Chlorophyll *a* was extracted in
409 absolute methanol for 12 h at room temperature in the dark, and the concentration was
410 determined spectrophotometrically according to (Ritchie, 2008). For DNA extraction, water was
411 filtered using a vacuum pump on GF/F filters (Whatman glass fiber, 25 mm, nominal pore size
412 0.7 µm). The filters were placed into 1.5 mL sterile Eppendorf tubes with 1 mL of storage buffer
413 (40 mM EDTA, 50 mM Tris-HCl, 0.75 M Sucrose) and stored at -80°C until analyzed. The filtrate
414 from the GF/F filters was collected for measurement of dissolved nutrients, and unfiltered water
415 was used for total phosphorus measurements according to (Grasshoff and Kremling, 1983). DNA
416 extraction was performed using bead-beating followed by robotic extraction using a QiaCube
417 robot, the gene encoding the 16S subunit of the ribosomal RNA was amplified using primers
418 targeting the V3-V4 region, and the amplicons were sequenced using the 2×250 base pair format

419 on a MiSeq flow cell (V3 chemistry). More details on the DNA extraction and 16S amplification
420 procedures are found in the supplementary methods section.

421 4.3 16S Sequence analysis and statistical tests

422 Sequences received from the sequencing facility were quality filtered to remove PhiX DNA,
423 unmerged reads and low quality sequences, followed by classification into different Operational
424 Taxonomic Units (OTUs) based on 97% sequence similarity threshold using the
425 "pick_de_novo_otus.py" command in Qiime (Caporaso*, et al.*, 2010). OTUs with fewer than 5
426 reads across all samples were removed, resulting in 35,775 – 75,704 reads per sample. Alpha
427 diversity (Shannon index) was calculated using the ''diversity'' function, and Bray–Curtis
428 dissimilarity by ''vegdist'' function. Non-metric multidimensional scaling analysis (nMDS) was
429 carried out in R with ''metaMDS'' function and a Bray-Curtis matrix. In order to relate
430 environmental factors to specific OTUs we performed variation partitioning ("varpart" function
431 in R) followed by Canonical Correspondence Analaysis (CCA) using the "cca" and "envfit"
432 functions in the "vegan" R package. For additional information on these analyses, and on the
433 permutation analysis for the CCA, see the supplementary methods.

434 4.4 Generalized Linear Model Analysis (GLM)

435 We used logistic regression in the framework of GLMs to relate the six land use types to the
436 measured water chemical parameters. Multi-model inference based on the Akaike Information
437 Criterion (AIC) was used to rank the importance of variables (Burnham and Anderson, 2003,
438 Blank and Blaustein, 2014). The package ''glmulti'' was used to facilitate multi-model inference
439 based on every possible first-order combination of the predictors in each scale (Calcagno and de
440 Mazancourt, 2010). The coefficients associated with each variable and their relative importance
441 were assessed using a multi-model average. Parameters were estimated by averaging across

442 models, weighted by the probability that their originating models are the best in the set. Hence,
443 the relative importance of a parameter is not based on *P*-values, but rather on its occurrence in
444 the models that are best supported by the data. The importance value for a particular predictor
445 is equal to the sum of the weights for the models in which the variable appeared. Thus, a
446 variable that shows up in many of the models with large weights received a high importance
447 value.

448 **Acknowledgements**

449 We thank Elke Mach and Uta Mallok for chemical analyses, Liat Linde and Doron Fogel for the
450 DNA extraction and the MIBI group for valuable discussions

451 **Funding source**

452 This study was supported by a German-Israeli Cooperation in Water Technology Research Young
453 Scientists Exchange Program (to SM), by grant number 3-10342 from the Israeli Ministry of
454 Science and Technology (to DS) and by DFG project Microprime (GR1540/28-1, to HPG). Also, we
455 thank the Leibnitz Institute and the University of Haifa for additional funding

456 **Conflict of Interest Statement**

457 The authors declare no conflict of interest

458 **Author contributions:**

459 SM, HPG and DS designed study. SM, LG, HPG and DS collected samples, LB, LG, HPG and DS
460 analyzed the geographic data, SM, LB, AA, AM, ML and DS analyzed the genetic data, SM and DS
461 wrote the manuscript with contributions from all co-authors.

462

463 ## References

464 Abdelzaher, A.M., Wright, M.E., Ortega, C.*, et al.* (2010) Presence of Pathogens and Indicator
465 Microbes at a Non-Point Source Subtropical Recreational Marine Beach. *Applied and*
466 *Environmental Microbiology* **76**: 724-732.

467 Acinas, S.G., Haverkamp, T.H.A., Huisman, J., and Stal, L,J. (2008) Phenotypic and genetic
468 diversification of Pseudanabaena spp. (Cyanobacteria). *The Isme Journal* **3**: 31.

469 Allgaier, M., and Grossart, H.P. (2006) Diversity and seasonal dynamics of Actinobacteria
470 populations in four lakes in northeastern Germany. *Applied and Environmental Microbiology* **72**:
471 3489-3497.

472 Battistuzzi, F.U., and Hedges, S.B. (2009) A Major Clade of Prokaryotes with Ancient Adaptations
473 to Life on Land. *Molecular Biology and Evolution* **26**: 335-343.

474 Beaver, J.R., Manis, E.E., Loftin, K.A., Graham, J.L., Pollard, A.I., and Mitchell, R.M. (2014) Land
475 use patterns, ecoregion, and microcystin relationships in U.S. lakes and reservoirs: A preliminary
476 evaluation. *Harmful Algae* **36**: 57-62.

477 Bell, T. (2010) Experimental tests of the bacterial distance–decay relationship. *The Isme Journal*
478 **4**: 1357.

479 Blank, L., and Blaustein, L. (2014) A multi-scale analysis of breeding site characteristics of the
480 endangered fire salamander (Salamandra infraimmaculata) at its extreme southern range limit.
481 *Hydrobiologia* **726**: 229-244.

482 Briée, C., Moreira, D., and López-García, P. (2007) Archaeal and bacterial community
483 composition of sediment and plankton from a suboxic freshwater pond. *Research in*
484 *Microbiology* **158**: 213-227.

485 Burnham, K.P., and Anderson, D.R. (2003) *Model selection and multimodel inference: a practical*
486 *information-theoretic approach*. Springer Science and Business Media.

487 Cabral, J.P.S. (2010) Water Microbiology. Bacterial Pathogens and Water. *International Journal*
488 *of Environmental Research and Public Health* **7**: 3657.
489 Calcagno, V., and de Mazancourt, C., (2010) glmulti: an R package for easy automated model
490 selection with (generalized) linear models. *Journal of statistical software* **34**: 1-29.
491 Callieri C (2017) Synechococcus plasticity under environmental changes. *FEMS microbiology*
492 *letters* **364**: fnx229.
493 Camp, H.J.M.O.d., Islam, T., Stott, M.B.*, et al.* (2009) Environmental, genomic and taxonomic
494 perspectives on methanotrophic Verrucomicrobia. *Environmental Microbiology Reports* **1**: 293-
495 306.
496 Caporaso, J.G., Kuczynski, J., Stombaugh, J.*, et al.* (2010) QIIME allows analysis of high-
497 throughput community sequencing data. *Nature Methods* **7**: 335.
498 Carini, S., Bano, N., LeCleir, G., and Joye, S.B. (2005) Aerobic methane oxidation and
499 methanotroph community composition during seasonal stratification in Mono Lake, California
500 (USA). *Environmental Microbiology* **7**: 1127-1138.
501 Carpenter, S.R., Caraco, N.F., Correll, D.L., Howarth, R.W., Sharpley, A.N., and Smith, V.H. (1998)
502 Mompoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications* **8**:
503 559-568.
504 Casper, S.J. (2012) *Lake Stechlin: A temperate oligotrophic lake*. Springer Science and Business
505 Media.
506 Catherine, A., Troussellier, M. and Bernard, C. (2008) Design and application of a stratified
507 sampling strategy to study the regional distribution of Cyanobacteria (Ile-de-France, France).
508 *Water Research* **42**: 4989-5001.

509 Cirés, S., and Ballot, A. (2016) A review of the phylogeny, ecology and toxin production of
510 bloom-forming Aphanizomenon spp. and related species within the Nostocales (Cyanobacteria).
511 *Harmful Algae* **54**: 21-43.
512 Crump, B.C., Hopkinson, C.S., Sogin, M.L., and Hobbie, J.E. (2004) Microbial Biogeography along
513 an Estuarine Salinity Gradient: Combined Influences of Bacterial Growth and Residence Time.
514 *Applied and Environmental Microbiology* **70**: 1494-1505.
515 Crump, B.C., Adams, H.E., Hobbie, J.E., and Kling, G.W. (2007) Biogeography of bacterioplankton
516 in lakes and streams of an Arctic Tundra chtchment. *Ecology* **88**: 1365-1378.
517 Davis, T.W., Bullerjahn, G.S., Tuttle, T., McKay, R.M. and Watson, S.B. (2015) Effects of Increasing
518 Nitrogen and Phosphorus Concentrations on Phytoplankton Community Growth and Toxicity
519 During Planktothrix Blooms in Sandusky Bay, Lake Erie. *Environmental Science and Technology*
520 **49**: 7197-7207.
521 Eiler, A., and Bertilsson, S. (2004) Composition of freshwater bacterial communities associated
522 with Cyanobacterial blooms in four Swedish lakes. *Environmental Microbiology* **6**: 1228-1243.
523 Fukushima, M., Tomioka, N., Jutagate, T.*, et al.* (2017) The dynamics of pico-sized and bloom-
524 forming Cyanobacteria in large water bodies in the Mekong River Basin. *PLOS ONE* **12**:
525 e0189609.
526 Ghai, R., Mizuno, C.M., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2014) Key roles for
527 freshwater Actinobacteria revealed by deep metagenomic sequencing. *Molecular Ecology* **23**:
528 6073-6090.
529 Glöckner, F.O., Zaichikov, E., Belkova, N., Denissova, L., Pernthaler, J., Pernthaler, A., and Amann,
530 R. (2000) Comparative 16S rRNA Analysis of Lake Bacterioplankton Reveals Globally Distributed
531 Phylogenetic Clusters Including an Abundant Group of Actinobacteria. *Applied and*
532 *Environmental Microbiology* **66**: 5053-5065.

533 Grasshoff, K.E., and Kremling, M. (1983) Methods of seawater analysis. Verlag Chemie.
534 *Weinheim, Germany* **4**: 9.
535 Griebler, C., and Lueders, T. (2009) Microbial biodiversity in groundwater ecosystems.
536 *Freshwater Biology* **54**: 649 – 677.
537 Hargreaves, S.K., Williams, R.J., and Hofmockel, K.S. (2015) Environmental Filtering of Microbial
538 Communities in Agricultural Soil Shifts with Crop Growth. *PLOS ONE* **10**: e0134345.
539 Haukka, K., Kolmonen, E., Hyder, R.*, et al.* (2006) Effect of Nutrient Loading on Bacterioplankton
540 Community Composition in Lake Mesocosms. *Microbial Ecology* **51**: 137-146.
541 Horner-Devine, M.C., Lage, M., Hughes, J.B., and Bohannan, B.J.M. (2004) A taxa–area
542 relationship for bacteria. *Nature* **432**: 750.
543 Hu, S., Zeng, R.J., Haroon, M.F., Keller, J., Lant, P.A., Tyson, G.W. and Yuan, Z. (2015) A laboratory
544 investigation of interactions between denitrifying anaerobic methane oxidation (DAMO) and
545 anammox processes in anoxic environments. *Scientific Reports* **5**: 8706.
546 Humayoun, S.B., Bano, N., and Hollibaugh, J.T. (2003) Depth Distribution of Microbial Diversity in
547 Mono Lake, a Meromictic Soda Lake in California. *Applied and Environmental Microbiology* **69**:
548 1030-1042.
549 Janssen, P.H. (2006) Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S rRNA and
550 16S rRNA Genes. *Applied and Environmental Microbiology* **72**: 1719-1728.
551 Joerin, F., Thériault, M., and Musy, A. (2001) Using GIS and outranking multicriteria analysis for
552 land-use suitability assessment. *International Journal of Geographical Information Science* **15**:
553 153-174.
554 Jones, S.E., and McMahon, K.D. (2009) Species-sorting may explain an apparent minimal effect
555 of immigration on freshwater bacterial community dynamics. *Environmental Microbiology* **11**:
556 905-913.

557 Katsiapi, M., Mazaris, A.D., Charalampous, E., and Moustaka-Gouni, M. (2012) Watershed land
558 use types as drivers of freshwater phytoplankton structure. *Hydrobiologia* **698**: 121-131.
559 Khadem AF, Pol A, Jetten MSM and Op den Camp HJM (2010) Nitrogen fixation by the
560 Verrucomicrobial methanotroph 'Methylacidiphilum fumariolicum' SolV. *Microbiology* **156**:
561 1052-1059.
562 King, A.J., Freeman, K.R., McCormick, K.F., Lynch, R.C., Lozupone, C., Knight, R., and Schmidt, S.K.
563 (2010) Biogeography and habitat modelling of high-alpine bacteria. *Nature Communications* **1**:
564 53.
565 Kolmonen, E., Sivonen, K., Rapala, J., and Haukka, K. (2004) Diversity of Cyanobacteria and
566 heterotrophic bacteria in Cyanobacterial blooms in Lake Joutikas, Finland. *Aquatic Microbial*
567 *Ecology* **36**: 201-211.
568 Kopylov, A.I., Kosolapov, D.B., Romanenko, A.V., and Degermendzhy, A.G. (2002) Structure of
569 planktonic microbial food web in a brackish stratified Siberian lake. *Aquatic Ecology* **36**: 179-204.
570 Langenheder, S., and Székely, A.J. (2011) Species sorting and neutral processes are both
571 important during the initial assembly of bacterial communities. *The Isme Journal* **5**: 1086-1094.
572 Larned, S.T., Scarsbrook, M.R., Snelder, T.H., Norton, N.J., and Biggs, B.J.F. (2004) Water quality
573 in low-elevation streams and rivers of New Zealand: Recent state and trends in contrasting land-
574 cover classes. *New Zealand Journal of Marine and Freshwater Research* **38**: 347-366.
575 Lear, G., Washington, V., Neale, M., Case, B., Buckley, H., and Lewis, G. (2013) The biogeography
576 of stream bacteria. *Global Ecology and Biogeography* **22**: 544-554.
577 Leibold, M.A., and Norberg, J. (2004) Biodiversity in metacommunities: Plankton as complex
578 adaptive systems? *Limnology and Oceanography* **49**: 1278-1289.

579 Lindström, E.S., and Bergström, A.K. (2004) Influence of inlet bacteria on bacterioplankton
580 assemblage composition in lakes of different hydraulic retention time. *Limnology and*
581 *Oceanography* **49**: 125-136.
582 Lindström, E.S., and Langenheder, S. (2012) Local and regional factors influencing bacterial
583 community assembly. *Environmental Microbiology Reports* **4**: 1-9.
584 Lindström, E.S., Kamst-Van Agterveld, M.P., and Zwart, G. (2005) Distribution of Typical
585 Freshwater Bacterial Groups Is Associated with pH, Temperature, and Lake Water Retention
586 Time. *Applied and Environmental Microbiology* **71**: 8201-8206.
587 Logue, J.B., Stedmon, C.A., Kellerman, A.M*., et al.* (2015) Experimental insights into the
588 importance of aquatic bacterial community composition to the degradation of dissolved organic
589 matter. *The Isme Journal* **10**: 533.
590 Mantzouki, E., Lürling, M., Fastner, J*,. et al.* (2018) Temperature Effects Explain Continental Scale
591 Distribution of Cyanobacterial Toxins. *Toxins* **10**: 156.
592 Marcelino, V.R., van Oppen, M.J.H,. and Verbruggen, H. (2017) Highly structured prokaryote
593 communities exist within the skeleton of coral colonies. *The Isme Journal* **12**: 300.
594 Marmen, S., Aharonovich, D., Grossowicz, M., Blank, L., Yacobi, Y.Z., and Sher, D.J. (2016)
595 Distribution and Habitat Specificity of Potentially-Toxic Microcystis across Climate, Land, and
596 Water Use Gradients. *Frontiers in Microbiology* **7**.
597 Mašín, M., Jezbera, J., Nedoma, J., Straškrabová, V., Hejzlar, J., and Šimek, K. (2003) Changes in
598 bacterial community composition and microbial activities along the longitudinal axis of two
599 canyon-shaped reservoirs with different inflow loading. *Hydrobiologia* **504**: 99-113.
600 Mehaffey, M.H., Nash, M.S., Wade, T.G., Ebert, D.W., Jones, K.B., and Rager, A. (2005) Linking
601 Land Cover and Water Quality in New York City'S Water Supply Watersheds. *Environmental*
602 *Monitoring and Assessment* **107**: 29-44.

Mitsui, A., Kumazawa, S., Takahashi, A., Ikemoto, H., Cao, S., and Arai, T. (1986) Strategy by which nitrogen-fixing unicellular Cyanobacteria grow photoautotrophically. *Nature* **323**: 720.

Ndlela, L.L., Oberholster, P.J., Van Wyk, J.H., and Cheng, P.H. (2016) An overview of Cyanobacterial bloom occurrences and research in Africa over the last decade. *Harmful Algae* **60**: 11-26.

Nekola, J.C., and White, P.S. (1999) The distance decay of similarity in biogeography and ecology. *Journal of Biogeography* **26**: 867-878.

Newton, R.J., Jones, S.E., Eiler, A., McMahon, K.D., and Bertilsson, S. (2011) A Guide to the Natural History of Freshwater Lake Bacteria. *Microbiology and Molecular Biology Reviews* **75**: 14-49.

Nogueira, R., Melo, Ls.F., Purkhold, U., Wuertz, S., and Wagner, M. (2002) Nitrifying and heterotrophic population dynamics in biofilm reactors: effects of hydraulic retention time and the presence of organic carbon. *Water Research* **36**: 469-481.

Percent, S.F., Frischer, M.E., Vescio, P.A.*, et al.* (2008) Bacterial Community Structure of Acid-Impacted Lakes: What Controls Diversity? *Applied and Environmental Microbiology* **74**: 1856-1868.

Quinn, J.M., Cooper, A.B., Davies-Colley, R.J., Rutherford, J.C. and Williamson, R.B. (1997) Land use effects on habitat, water quality, periphyton, and benthic invertebrates in Waikato, New Zealand, hill-country streams. *New Zealand Journal of Marine and Freshwater Research* **31**: 579-597.

Reimann, C., Finne, T.E., Nordgulen, Ø., Sæther, O.M., Arnoldussen, A., and Bank,s D. (2009) The influence of geology and land-use on inorganic stream water quality in the Oslo region, Norway. *Applied Geochemistry* **24**: 1862-1874.

626 Ritchie, R.J. (2008) Universal chlorophyll equations for estimating chlorophylls a, b, c, and d and
627 total chlorophylls in natural assemblages of photosynthetic organisms using acetone, methanol,
628 or ethanol solvents. *Photosynthetica* **46**: 115-126.
629 Savage, C., Leavitt, P.R., and Elmgren, R. (2010) Effects of land use, urbanization, and climate
630 variability on coastal eutrophication in the Baltic Sea. *Limnology and Oceanography* **55**: 1033-
631 1046.
632 Schoonover, J.E., and Lockaby, B.G. (2006) Land cover impacts on stream nutrients and fecal
633 coliform in the lower Piedmont of West Georgia. *Journal of Hydrology* **331**: 371-382.
634 Singer E, Bushnell B, Coleman-Derr D*, et al.* (2016) High-resolution phylogenetic microbial
635 community profiling. *The Isme Journal* **10**: 2020.
636 Sukenik, A., Hadas, O., Kaplan, A., and Quesada, A. (2012) Invasion of Nostocales
637 (Cyanobacteria) to Subtropical and Temperate Freshwater Lakes – Physiological, Regional, and
638 Global Driving Forces. *Frontiers in Microbiology* **3**.
639 Valério, E., Chaves, S., and Tenreiro, R. (2010) Diversity and Impact of Prokaryotic Toxins on
640 Aquatic Environments: A Review. *Toxins* **2**: 2359.
641 Van Rossum, T., Peabody, M.A., Uyaguari-Diaz, M.I*., et al.* (2015) Year-Long Metagenomic Study
642 of River Microbiomes Across Land Use and Water Quality. *Frontiers in Microbiology* **6**.
643 Vitousek, P.M., Mooney, H.A., Lubchenco, J., and Melillo, J.M. (1997) Human Domination of
644 Earth's Ecosystems. *Science* **277**: 494-499.
645 Wang, X., and Yin, Z.Y. (1997) Using GIS to assess the relationship between land use and water
646 quality at a watershed level. *Environment International* **23**: 103-114.
647 Wei, Y., Wei, Z., Cao, Z*., et al.* (2016) A regulating method for the distribution of phosphorus
648 fractions based on environmental parameters related to the key phosphate-solubilizing bacteria
649 during composting. *Bioresource Technology* **211**: 610-617.

650 Winter, T.C., Rosenberry, D.O., and LaBaugh, J.W. (2003) Where Does the Ground Water in Small
651 Watersheds Come From? *Groundwater* **41**: 989-1000.
652 Witt, V., Wild, C., and Uthicke, S. (2012) Terrestrial Runoff Controls the Bacterial Community
653 Composition of Biofilms along a Water Quality Gradient in the Great Barrier Reef. *Applied and*
654 *Environmental Microbiology* **78**: 7786-7791.
655 Wolz, C.R.M., Yarwood, S.A., Grant, E.H.C., Fleischer, R.C., Lips, K.R., and Hoye, B. (2018) Effects
656 of host species and environment on the skin microbiome of Plethodontid salamanders. *Journal*
657 *of Animal Ecology* **87**: 341-353.
658 Xu, H., Paerl, H.W., Qin, B., Zhu, G., and Gaoa, G. (2010) Nitrogen and phosphorus inputs control
659 phytoplankton growth in eutrophic Lake Taihu, China. *Limnology and Oceanography* **55**: 420-
660 432.
661 Yannarell, A.C., and Triplett, E.W. (2005) Geographic and Environmental Sources of Variation in
662 Lake Bacterial Community Composition. *Applied and Environmental Microbiology* **71**: 227-239.
663 Zhang, J., Su, Y., Wu, J., and Liang, H. (2015) GIS based land suitability assessment for tobacco
664 production using AHP and fuzzy set in Shandong province of China. *Computers and Electronics in*
665 *Agriculture* **114**: 202-211.

666

667

668

669

670

671

Table and Figure legends:

Figure 1. Land use and drainage basins in the sampling area. A total of 46 samples were collected from lakes and a fish-pond in the Brandenburg region of northern Germany. Five types of land use (urban areas, agriculture, forests, wetlands and water bodies), mapped using GIS, are shown, with black lines superimposed indicating drainage basins. The arrows indicate the dominant flow direction within each basin. The color used for the sampling locations from each basin is used in all other figures. A map with the identities of each sampling location is presented in Fig. S1, and the detailed information on each sampling location can be found in Supplementary Excel table.

Figure 2. An overview of the microbial populations in the sampled lakes. The columns are organized by the dendogram in panel 2B and are colored (bars in panel A) by drainage basin, as shown in Figure 1. **A) α-diversity of the microbial populations in the different lakes, measured by the Shannon index.** No clear patterns are observed. **B) Community composition and community clustering based on the Bray-Curtis similarity index**; Colors represent different phyla, lines within the same color represent OTUs at the 97% identity cutoff. The sampling sites can be divided into two groups - one dominated by Cyanobacteria, and the other by Actinobacteria, Proteobacteria and Bacteroidetes. **C) No distance decay patterns are observed**. Each point represents the Bray-Curtis dissimilarity between locations within the same basin (blue) or locations found in two separate basins (red). **D) Most of the OTUs are found in more than 50% of the sampling sites**. The diversity (γ diversity) of all sampling locations combined is presented, divided into five categories representing core OTUs (found in all sites), cosmopolitan OTUs (> 90% of the sites), common OTUs (> 50% of the sites), rare OTUs (< 10% of the sites), and unique OTUs (present at one site only).

Figure 3. Land use affects water chemistry, which, in turn, influences the microbiome composition. Variation Partition Analysis (VPA) was performed with 2 data sets (basin and local

696 land use) to explain water chemistry, and with 3 data sets (basin and local land use as well as
697 water chemistry) to explain the variation in the microbiome composition**. A) The combination of**
698 **local and basin scale land use were able to explain almost 40% of the variability in water**
699 **chemical properties**, with basin land use as the main explanatory variable. **B) Water chemical**
700 **properties represent the main explanatory variable for the microbiome composition (ca. 19%**
701 **among all sites).** The combination of basin scale land use and water quality explain 22.7% of
702 microbiome variability.

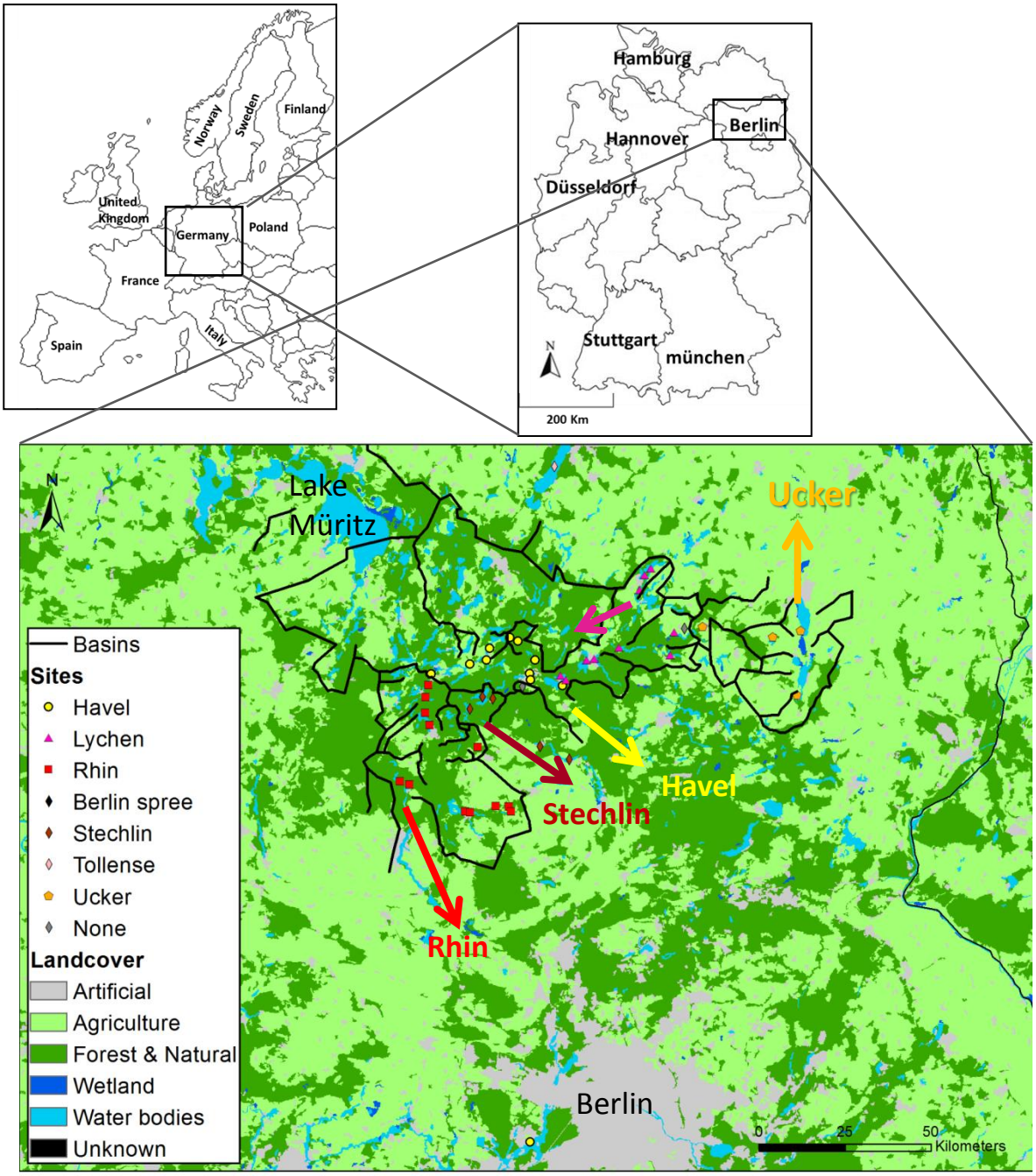
703 **Figure 4. The importance of different land use types for determining water chemical**
704 **parameters.** The results of General Linear Models (GLMs) analysis for the land use for each of
705 the water chemical properties are shown: **A) Urban areas are positively related to NO3+NO2**
706 **concentrations.** A strong correlation (R^2= 0.74) is observed between land uses and nitrite and
707 nitrate concentrations, with the fraction of "urban areas" being the most important variable
708 contributing to the correlation; **B) Forests and natural areas are positively related with**
709 **chlorophyll-a concentrations.** A very strong correlation between land use and *chl a*
710 concentration (R^2= 0.98), with "forests and natural areas" being the most important variable
711 contributing to the correlation.

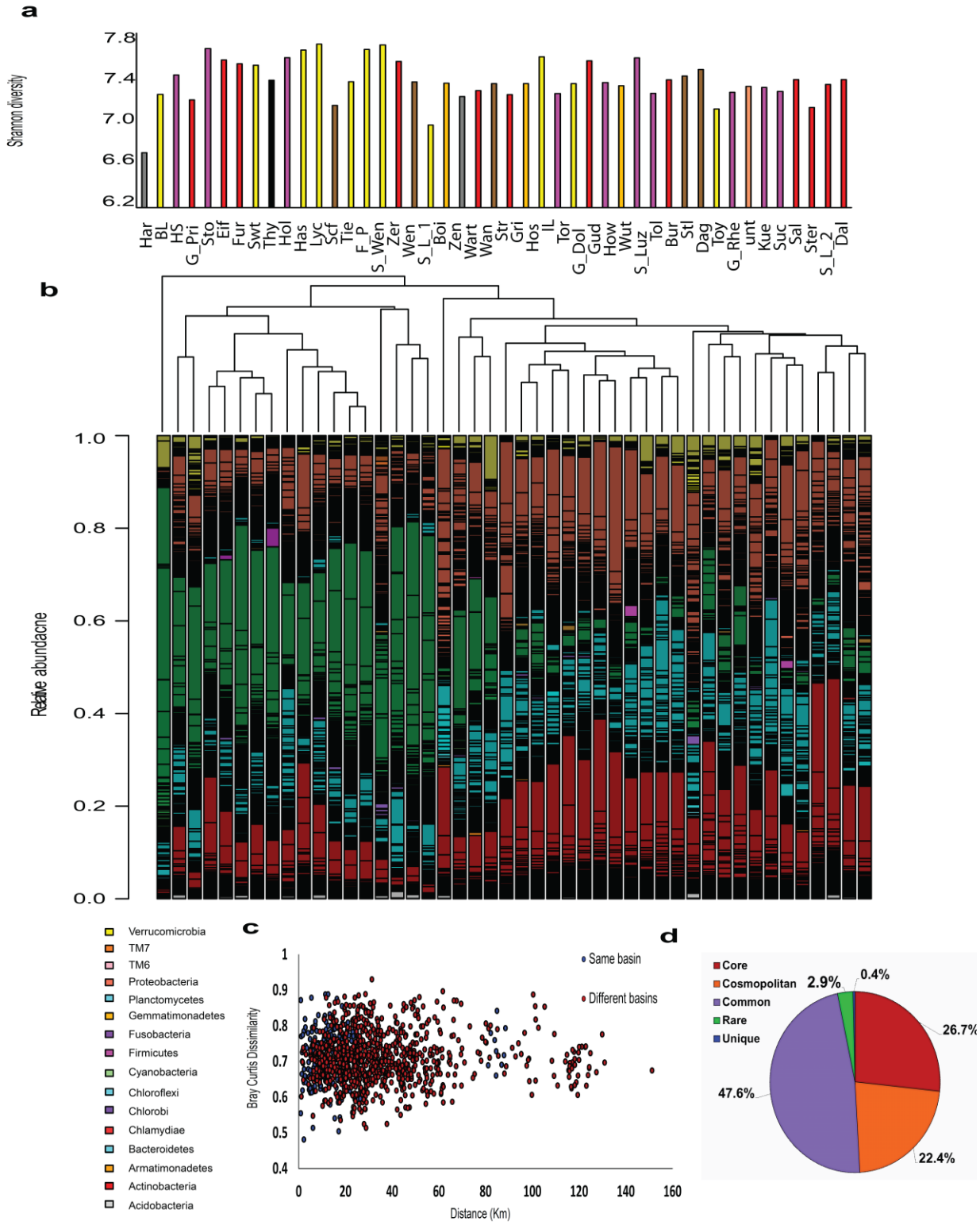
712 **Figure 5. Differences between bacterial phyla in their association with different aquatic**
713 **environmental conditions**. Canonical Correspondence Analysis (CCA) plots are shown, with
714 sampling locations and OTUs ordinated based on the four aquatic environmental parameters
715 shown to significantly affect the microbiome composition. **A) An overview of the ordination,**
716 **showing the 2845 most abundant OTUs. B) The same ordination as in A, with the OTUs**
717 **removed** to show the relative location of the eigenvectors for the aquatic environmental
718 parameters and the ordination of the sampling sites. **C-G) The same plot as in A, with specific**

719 **microbial phyla highlighted.** Numbers represent the results of permutation tests to determine
720 whether the distribution of each phyla is significantly un-random.

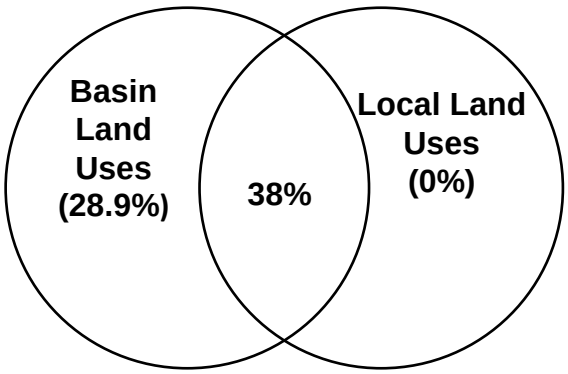
721 **Figure 6. Separation in the CCA ordination between bacterial lineages at the taxonomic levels**
722 **of order, class or family. a) Cyanobacterial classes.** Note the clear separation between the
723 Synechococcoales and the Pseudoanabaenales and Nostocales along the first CCA axis (r^2 = 0.2,
724 Adj p val = 0.03, Adonis test), with some differentiation between the latter two orders primarily
725 along the second CCA axis. **b) Verrucomicrobial orders.** Note the Methylacidifilae, which are
726 limited to the upper-right hand quadrant, possibly in association with high NO3+NO2 levels, and
727 clearly separated from the other orders, except of the Spartobacteria (r^2 = 0.2, Adj p val = 0.01 ,
728 Adonis test). **c) Proteobacterial orders. d) Bacteroidetes order. e) Actinobacteria family.** No
729 clear differentiation is seen between the orders or families.

730 **Figure 7. Conceptual model summarizing the observed effects of land use on water chemistry,**
731 **and of water chemistry on the microbiome.** Solid arrows represent important correlations
732 between land use and aquatic environmental parameters, based on the GLMs. Dashed arrows
733 represent positive relationships between aquatic environmental parameters and specific
734 microbial lineages inferred from the CCA plots, and blunt dashed lines represent negative
735 relationships inferred in the same manner. Forest and urban land uses correlate with high *Chl a*
736 and NO_3+NO_2 concentrations, respectively. These parameters in turn, are part of the aquatic
737 environmental conditions which determine, to some extent, the relative distribution of specific
738 Cyanobacterial classes and Verrucomicrobial orders.

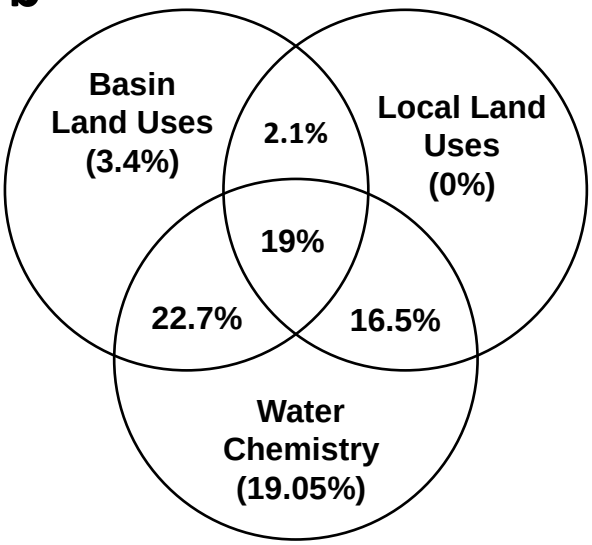

Norway
Sweden
Finland
United Kingdom
Germany
Poland
France
Spain
Italy
Hamburg
Berlin
Hannover
Düsseldorf
Stuttgart
münchen
N
200 Km
N
Lake Müritz
Ucker
Havel
Stechlin
Rhin
Berlin
0
25
50
Kilometers
Basins
Sites
Havel
Lychen
Rhin
Berlin spree
Stechlin
Tollense
Ucker
None
Landcover
Artificial
Agriculture
Forest & Natural
Wetland
Water bodies
Unknown

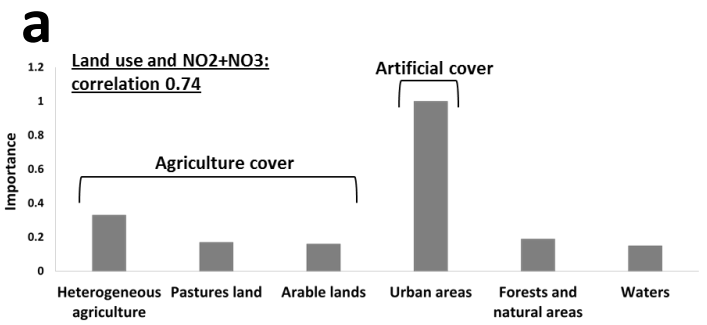

a
Shannon diversity
7.8
7.4
7.0
6.6
6.2
Har BL HS G_Pri Sto Eif Fur Swt Thy Hol Has Lyc Scf Tie F_P S_Wen Zer Wen S_L_1 Boi Zen Wart Wan Str Gri Hos IL Tor G_Dol Gud How Wut S_Luz Tol Bur Stl Dag Toy G_Rhe unt Kue Suc Sal Ster S_L_2 Dal
b
Relative abundance
1.0
0.8
0.6
0.4
0.2
0.0
Verrucomicrobia
TM7
TM6
Proteobacteria
Planctomycetes
Gemmatimonadetes
Fusobacteria
Firmicutes
Cyanobacteria
Chloroflexi
Chlorobi
Chlamydiae
Bacteroidetes
Armatimonadetes
Actinobacteria
Acidobacteria
c
Bray Curtis Dissimilarity
1
0.9
0.8
0.7
0.6
0.5
0.4
0
20
40
60
80
100
120
140
160
Distance (Km)
Same basin
Different basins
d
Core
Cosmopolitan
Common
Rare
Unique
2.9%
0.4%
26.7%
47.6%
22.4%

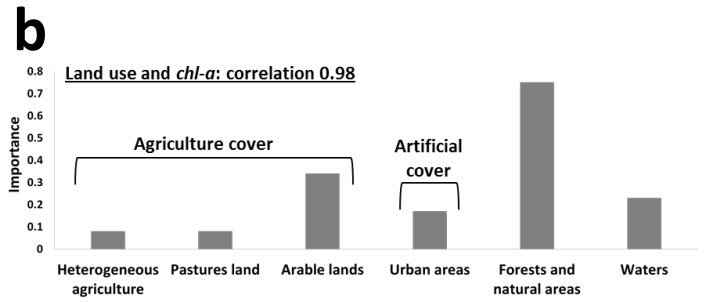
a

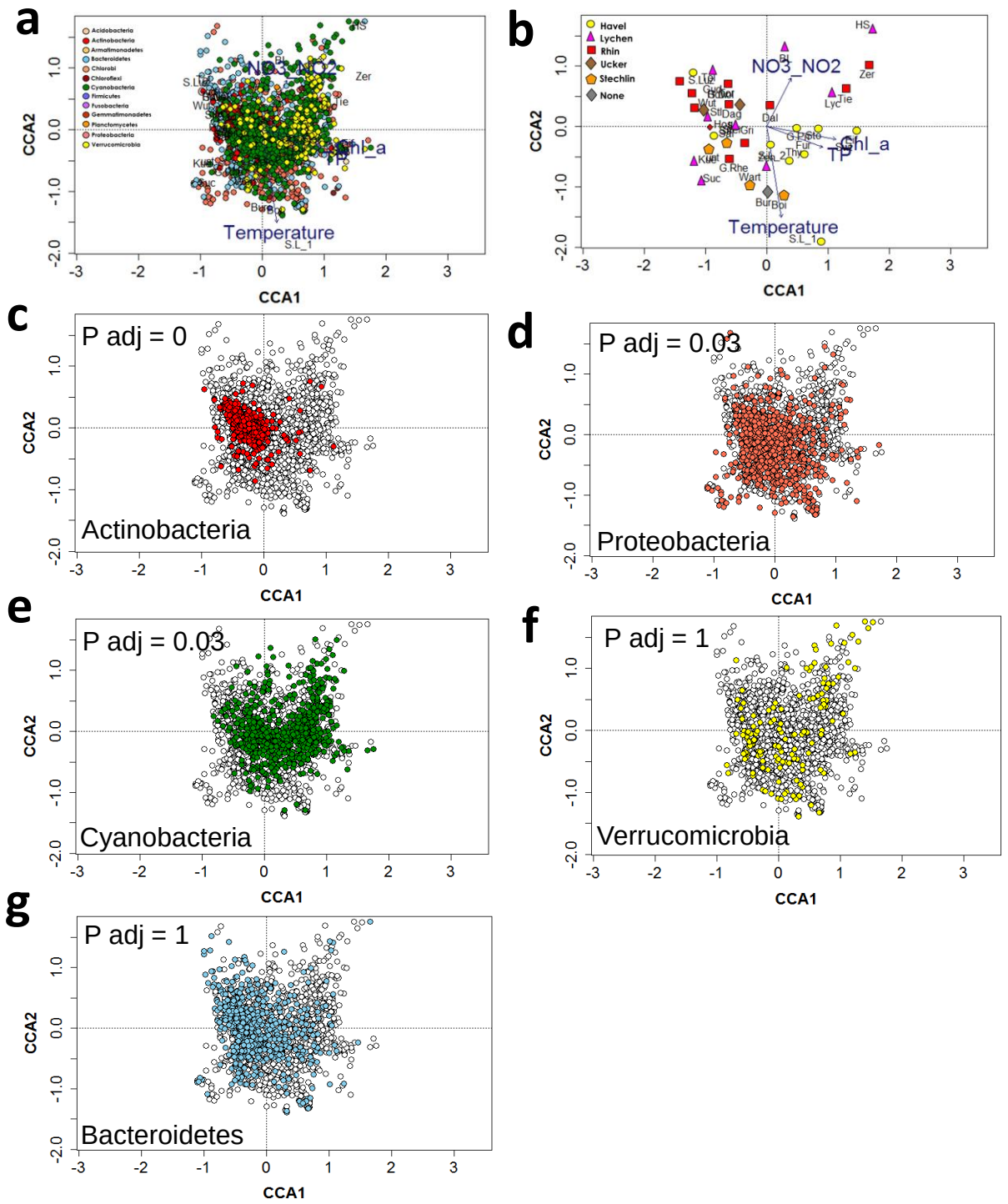

Basin
Land
Uses
(28.9%)
38%
Local Land
Uses
(0%)

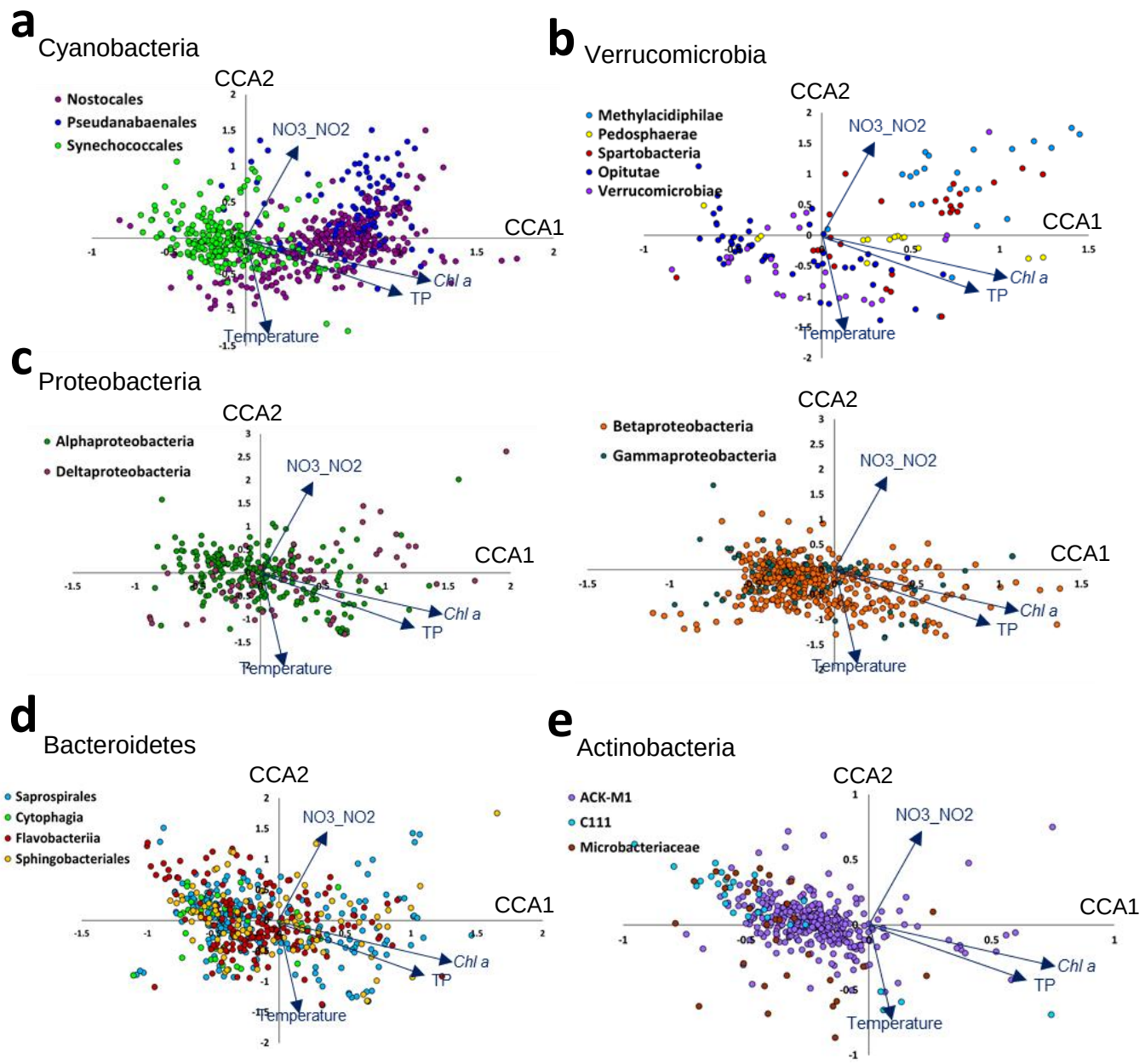
b

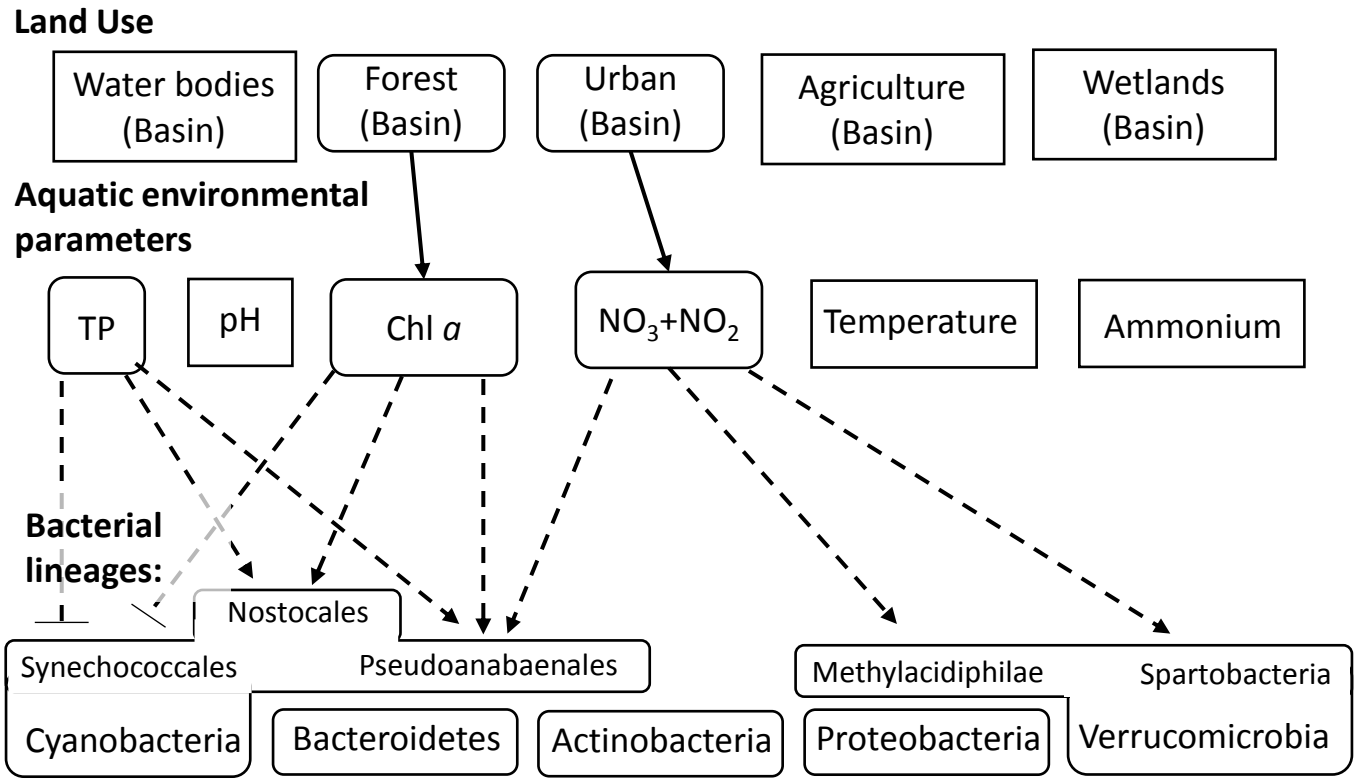

Basin
Land Uses
(3.4%)
2.1%
Local Land
Uses
(0%)
19%
22.7%
16.5%
Water
Chemistry
(19.05%)


a
Land use and NO2+NO3: correlation 0.74
Artificial cover
Agriculture cover
Importance
1.2
1
0.8
0.6
0.4
0.2
0
Heterogeneous agriculture
Pastures land
Arable lands
Urban areas
Forests and natural areas
Waters


b
Land use and *chl-a*: correlation 0.98
Agriculture cover
Artificial cover
Importance
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Heterogeneous agriculture
Pastures land
Arable lands
Urban areas
Forests and natural areas
Waters


a
Acidobacteria
Actinobacteria
Armatimonadetes
Bacteroidetes
Chlorobi
Chloroflexi
Cyanobacteria
Firmicutes
Fusobacteria
Gemmatimonadetes
Planctomycetes
Proteobacteria
Verrucomicrobia
NO3_NO2
Chl_a
TP
Temperature
CCA1
CCA2
b
Havel
Lychen
Rhin
Ücker
Stechlin
None
NO3_NO2
Chl_a
TP
Temperature
CCA1
CCA2
c
P adj = 0
Actinobacteria
d
P adj = 0.03
Proteobacteria
e
P adj = 0.03
Cyanobacteria
f
P adj = 1
Verrucomicrobia
g
P adj = 1
Bacteroidetes


a
Cyanobacteria
CCA2
Nostocales
Pseudanabaenales
Synechococcales
NO3_NO2
CCA1
Chl a
TP
Temperature
b
Verrucomicrobia
CCA2
Methylacidiphilae
Pedosphaerae
Spartobacteria
Opitutae
Verrucomicrobiae
NO3_NO2
CCA1
Chl a
TP
Temperature
c
Proteobacteria
CCA2
Alphaproteobacteria
Deltaproteobacteria
NO3_NO2
CCA1
Chl a
TP
Temperature
CCA2
Betaproteobacteria
Gammaproteobacteria
NO3_NO2
CCA1
Chl a
TP
Temperature
d
Bacteroidetes
CCA2
Saprospirales
Cytophagia
Flavobacteriia
Sphingobacteriales
NO3_NO2
CCA1
Chl a
TP
Temperature
e
Actinobacteria
CCA2
ACK-M1
C111
Microbacteriaceae
NO3_NO2
CCA1
Chl a
TP
Temperature


Land Use
Water bodies (Basin)
Forest (Basin)
Urban (Basin)
Agriculture (Basin)
Wetlands (Basin)
Aquatic environmental parameters
TP
pH
Chl *a*
NO_3+NO_2
Temperature
Ammonium
Bacterial lineages:
Nostocales
Synechococcales
Pseudoanabaenales
Methylacidiphilae
Spartobacteria
Cyanobacteria
Bacteroidetes
Actinobacteria
Proteobacteria
Verrucomicrobia