environmental microbiology

Environmental Microbiology (2017) **00**(00), 00–00 doi:10.1111/1462-2920.13749

Contribution of different bacterial dispersal sources to lakes: Population and community effects in different seasons

Jérôme Comte,[1†] Mercè Berga,[1,2] Ina Severin,[1‡] Jürg Brendan Logue [1§] and Eva S. Lindström [1*]

[1]*Department of Ecology and Genetics/Limnology, Uppsala University, Norbyvägen 18D, Uppsala, 75236, Sweden.*

[2]*Biological Oceanography, Leibniz-Institute for Baltic Sea Research, Warnemünde (IOW), Seestrasse 15, Rostock, Germany.*

Summary

The diversity and composition of lake bacterial communities are driven by the interplay between local contemporary environmental conditions and dispersal of cells from the surroundings, i.e. the metacommunity. Still, a conceptual understanding of the relative importance of the two types of factors is lacking. For instance, it is unknown which sources of dispersal are most important and under which circumstances. Here, we investigated the seasonal variation in the importance of dispersal from different sources (mixing, precipitation, surface runoff and sediment resuspension) for lake bacterioplankton community and population dynamics. For that purpose, two small forest lakes and their dispersal sources were sampled over a period of 10 months. The influence of dispersal on communities and populations was determined by 454 sequencing of the 16S rRNA gene and SourceTracker analysis. On the community level direct effects of dispersal were questionable from all sources. Instead we found that the community of the preceding sampling occasion, representing growth of resident bacteria, was of great importance. On the population level, however, dispersal of individual taxa from the inlet could be occasionally important even under low water flow. The effect of sediment resuspension and precipitation appeared small.

Received 20 September, 2016; revised 28 March, 2017; accepted 2 April, 2017. *For correspondence. E-mail Eva.Lindstrom@ebc.uu.se; Tel. +46184716497. Present addresses: [†]Environment and Climate Change Canada, Canada Centre for Inland Waters, Science and Technology - Watershed Hydrology and Ecology Research Division, Burlington, ON L7S 1A1, Canada; [‡]Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, Berlin, 12587, Germany; [§]Marine Microbiology, Centre for Ecology and Evolution in Microbial Model Systems (EEMiS), Linnaeus University, 39231 Kalmar, Sweden. The authors declare no conflicts of interest.

Introduction

Within the study of microbial biogeography it is of central interest to understand to what extent communities are shaped by local contemporary conditions (species sorting) and/or dispersal processes (e.g. Martiny *et al.*, 2006; Hanson *et al.*, 2012; Lindström and Langenheder, 2012). While a variety of environmental factors have been identified as being important for the structuring of microbial communities, knowledge of dispersal, such as sources and routes, and its importance are generally lacking. With regard to lake communities, micro-organisms may disperse from outside to the lake, for example via precipitation (Jones *et al.*, 2008; Jones and McMahon, 2009) and inlet water flow (Lindström *et al.*, 2006). On the other hand, dispersal may also occur from another habitat within the lake, such as the sediment via resuspension (Lear *et al.*, 2014) or another stratum through water column mixing (Shade *et al.*, 2007). Moreover, the relative contribution of these different sources may not be constant over time and among lakes. For instance, water retention time may influence the role of dispersal via surface and ground water flow (Lindström *et al.*, 2006; Logue and Lindström, 2010; Read *et al.*, 2015; Ruiz-González *et al.*, 2015; Niño-García *et al.*, 2016), whereas the lake's morphometry (e.g. depth or area) could influence the importance of dispersal from sediment resuspension and precipitation. Thus, sediment resuspension should be greater in shallow lakes, while lakes with a larger surface area should allow for more dispersal via precipitation. Then again, seasonal events, such as spring floods (e.g. Adams *et al.*, 2014), rain storms and water column mixing, could also affect rates of dispersal. Hence, the relative importance of different dispersal sources is expected to vary seasonally.

Whether the importance of dispersal differs among bacterial taxa is generally little studied, but has received

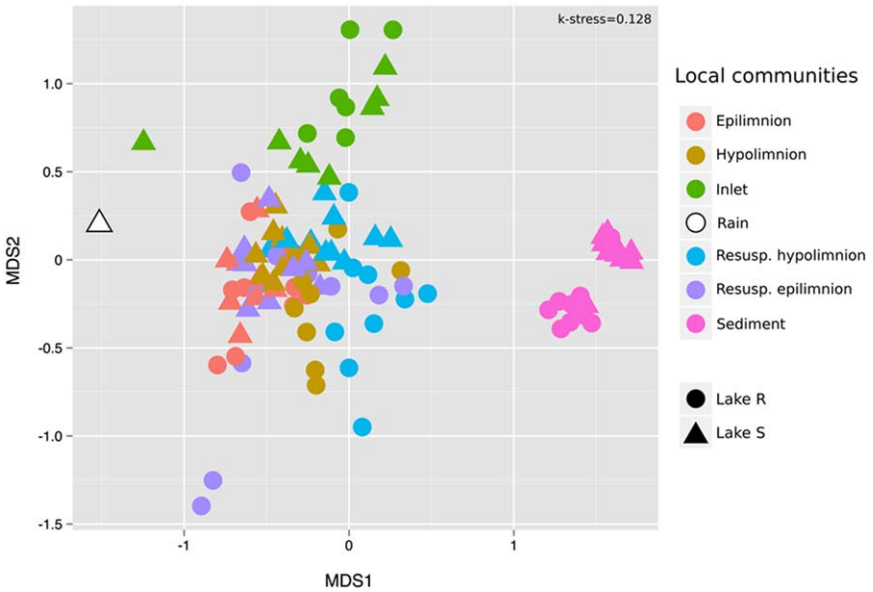

Fig. 1. Non-metric multidemsional scaling (NMDS) ordination visualization of bacterial community compositions (Morisita-Horn distances) sampled between February and November 2011 in Lakes Ramsjön (Lake R) and Siggeforasjön (Lake S). Resusp. refers to communities collected from sediment resuspension traps placed 1 meter below the surface (Resusp. epilimnion) or 1 meter above the sediment floor (Resusp. hypolimnion) in each lake.

some attention (Barberán and Casamayor, 2010; Székely and Langenheder, 2013; Ruiz-González *et al*., 2015). Different taxa may vary in their dispersal efficiency due to, for example, the ability to survive during dispersal (Bissett *et al*., 2010; Fahlgren *et al*., 2010; DeLeon-Rodriguez *et al*., 2013; Fahlgren *et al*., 2015). Furthermore, dispersed taxa appear to have different abilities to colonize a new environment. As an example, in the transition from soils to lake water some taxa may compete less well for new carbon sources in the lake than others (Crump *et al*., 2012; Shabarova *et al*., 2013; Adams *et al*., 2014). Therefore, although dispersal effects are not revealed on a community level using standard statistics, population specific effects are possible. However, this aspect of dispersal is little explored in natural bacterial communities, but should be facilitated by the combined use of high-throughput sequencing tools for characterization of bacterial communities and new computational methods, such as SourceTracker analysis (Knights *et al*., 2011). SourceTracker allows identifying the source of communities and populations (e.g. Newton *et al*., 2013), but is so far little used in community assembly studies (but see for example de Oliveira and Margis, 2015; Staley *et al*., 2015; Peter and Sommaruga, 2016).

The aim of this study is to evaluate the role of dispersal from external and internal sources for community composition as well as population dynamics of different taxa of bacteria in lake water (termed 'sink' hereafter). For this purpose, we investigated the importance of dispersal via precipitation, inlet inflow, sediment resuspension and mixing for the community composition of the epilimnion and hypolimnion of two forest lakes over a period of ten months. We hypothesized that (1) mass effect (i.e. immediate changes in community composition due to massive dispersal) would lead to changes in community composition especially at high dispersal events (e.g. during spring flood from inlets, heavy rains or mixing), (2) the relative contributions of the different sources differ between lakes depending on lake morphometry (i.e. depth and surface area) and (3) since colonization abilities differ among taxa, the effect of dispersal events also differs among taxa. These hypotheses were tested by analysis of the composition of bacterial communities in sources and sinks, using 454 sequencing of the 16S rRNA gene. We applied SourceTracker to evaluate the role of mass effects from different dispersal sources (precipitation, inlet inflow, sediment resuspension and mixing). We also included the resident 'past' bacterial community as a source in the SourceTracker analysis and propose this to serve as a measure of the role of growth of autochthonous bacteria.

Results

Bacterial community composition and diversity: general patterns

The bacterial community composition was very different in sediments as compared with samples obtained from lake and inlet water (Fig. 1). This was due to a greater dominance of Proteobacteria and Actinobacteria in the water samples compared with in the sediments (Supporting Information Fig. S4). The precipitation sample also deviated from all other samples (Fig. 1). The bacterial richness was generally higher in inlets and in sediments as opposed to epi- and hypolimnion samples (Supporting Information Fig. S5).

Bacterial community composition in epilimnion and hypolimnion

Bacterial community composition was very similar within strata of both lakes and there were also great similarities between epilimnion and hypolimnion of the respective lakes (Fig. 1), the minimum Morisita-Horn (MH) distance between strata being only 0.01. We used SourceTracker analysis to evaluate the proportion of different dispersal sources for the community composition of the epi- and hypolimnia (i.e. the sinks). Our SourceTracker models always included inlet and sediment communities as well as the other stratum of the lakes as sources. In the first model (Model 1), we also included the community of the respective stratum at the preceding sampling occasion as a source ('past'), while Model 2 was run without 'past'. The results of Model 1 showed that in all four strata a very important source of bacteria was 'past' (Fig. 2A). The other stratum of the lake, i.e. dispersal from the epilimnion to the hypolimnion and vice versa, also appeared important, except during the period of summer stratification in both lakes (Fig. 2A).

Removing 'past' from the SourceTracker Model (i.e. Model 2) led to an increase in the importance of the fraction 'unknown' (Fig. 2B). Thus, there were OTUs in the lakes, whose presence could not be explained by dispersal from the other sources but should mainly be sustained in the lake by internal growth, i.e. represented by 'past' in Model 1. Furthermore, removing 'past' also increased the importance of the other stratum of the lakes. Thus, there are a number of OTUs that have several possible sources of dispersal, and statistical analysis alone cannot reveal which one is the predominant. However, the general pattern of a decreased importance of the other stratum during summer stratification remained in Model 2. The SourceTracker analysis, thus, indicated that mass effect between epi- and hypolimnia could have occurred during periods of mixing in the lakes.

Since a lake-mixing event is not only dispersing cells but also changes the environmental conditions drastically, the SourceTracker analysis by itself is no proof of mass effect. Therefore, we evaluated the relative importance of dispersing cells and environmental conditions by distance-based redundancy analysis (db-RDA). In this model, we included as explanatory variables physico-chemical conditions (see Experimental procedures) and in addition a dummy variable representing mixing or non-mixing events. The rationale behind this procedure is that any variation explained by this dummy variable 'mixing' independent of physico-chemical conditions would represent other effects of mixing events including dispersal of cells. Separate db-RDA models were run for the two lakes (Fig. 3). Adjusted r^2 of the models were 0.26 and 0.41 for Lakes S and R, respectively ($p < 0.001$, $F = 4.46$ and 5.28, respectively). In both cases there was only a marginally significant effect of 'mixing' ($p = 0.09$ and 0.07 for Lakes S and R, respectively). Thus, in summary, although mixing events had clear effects on community composition it can be questioned if it was due to dispersal of cells between the two strata or rather due to the change in environmental conditions due to mixing followed by species sorting.

Community level results: inlets and precipitation

Similarities in BCC between inlets and lakes varied over time (Fig. 1). Dispersal from the inlet was identified by both SourceTracker models as a source to all strata (Fig. 2A and B). The influence of dispersal from the inlet appeared to be strongest for the epilimnia, especially in spring when the estimated water flow, and, thus, dispersal rates were highest (Supporting Information Fig. S1). At the highest inflow rate in April, dispersal rates to the epilimnion were estimated to be 3% of standing stock cell numbers over 48 h in Lake S and 11% in Lake R (Supporting Information Table S2). There was, however, no significant correlation between MH-distances between the communities in inlets and epilimnia and the dispersal rates from the inlet to the epilimnion in any of the lakes (Spearman rank correlation $p > 0.05$), suggesting an overall small importance of mass effect via the inlet at the community level. Further, there was no difference between the two lakes in the degree of importance of the inlet as a source of bacteria (two-tailed Mann-Whitney U-test of MH-distances between inlet and epilimnion, $p > 0.05$). There was a slightly greater importance of the inlets in SourceTracker Model 2 compared with Model 1, suggesting that for some of the OTUs both the 'past' and the inlet are possible sources.

Rates of dispersal via precipitation were negligible in both lakes (<0.01%; Supporting Information Table S2). We only had one BCC sample for precipitation samples, which was very different from the respective epilimnion community (Fig. 1). It was, however, not included in the SourceTracker models.

Community level results: sediments

As mentioned above, sediment communities and water column communities were always very different from each other (Fig. 1). However, differences between water and sediments were significantly smaller for Lake R than for Lake S (one-tailed Mann-Whitney *U*-test, $p = 0.001$ and 0.004 for epilimnion and hypolimnion, respectively). SourceTracker analysis (both models) consequently indicated a greater influence of the sediment community on the hypolimnion community in Lake R compared with lake S (Fig. 2A and B). Higher estimated dispersal rates in Lake R (up to 18% of standing stock cell numbers over 48 h) compared with Lake S (up to 5%) (Supporting Information

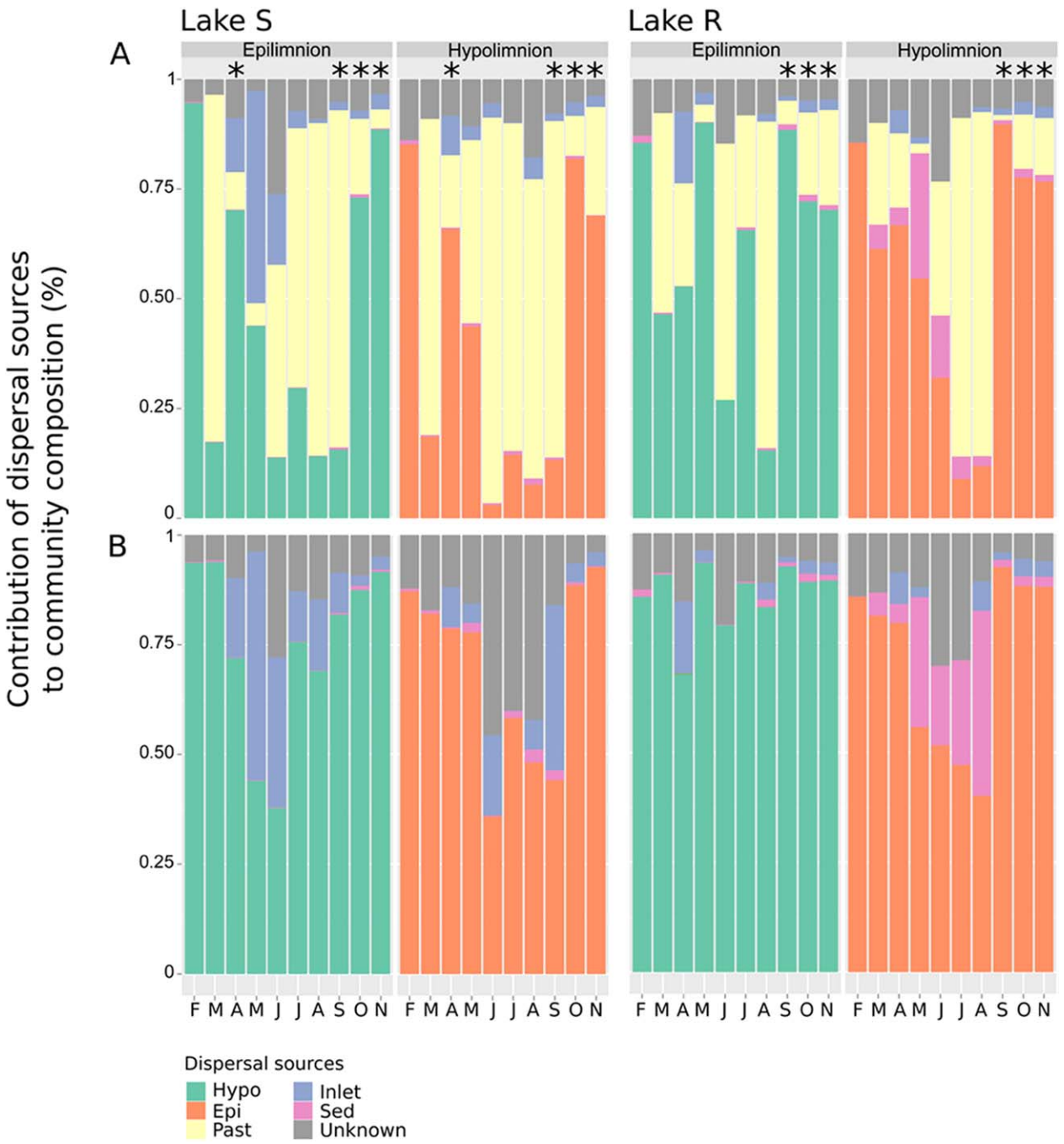

Fig. 2. Results from SourceTracker analysis showing the contribution of the different source communities to the epilimnion and the hypolimnion of Lakes Ramsjön (Lake R) and Siggeforasjön (Lake S). A: Model 1 including inlet, sediment, the other strata of the lake, and the previous sampling occasion ('past') as sources. B: Model 2, the same as A but without 'past'. Letters along the x-axes denotes the month of the year 2011. Asterisks indicate times of lake internal circulation.

Table S2) could be seen as support of mass effects due to sediment resuspension in Lake R. However, the estimated rates of dispersal due to sediment resuspension are probably over-estimating the true dispersal rates and should therefore be handled with caution (see Supporting Information Text S1).

Further, water chemistry analysis (results not shown) indicated that the May water sample from Lake R had been contaminated by the sediment upon sampling and, thus, this result is not reliable. We also discovered a shortcoming of the SourceTracker analysis in that sources and sinks are *a priori* defined and in some cases, it appears as

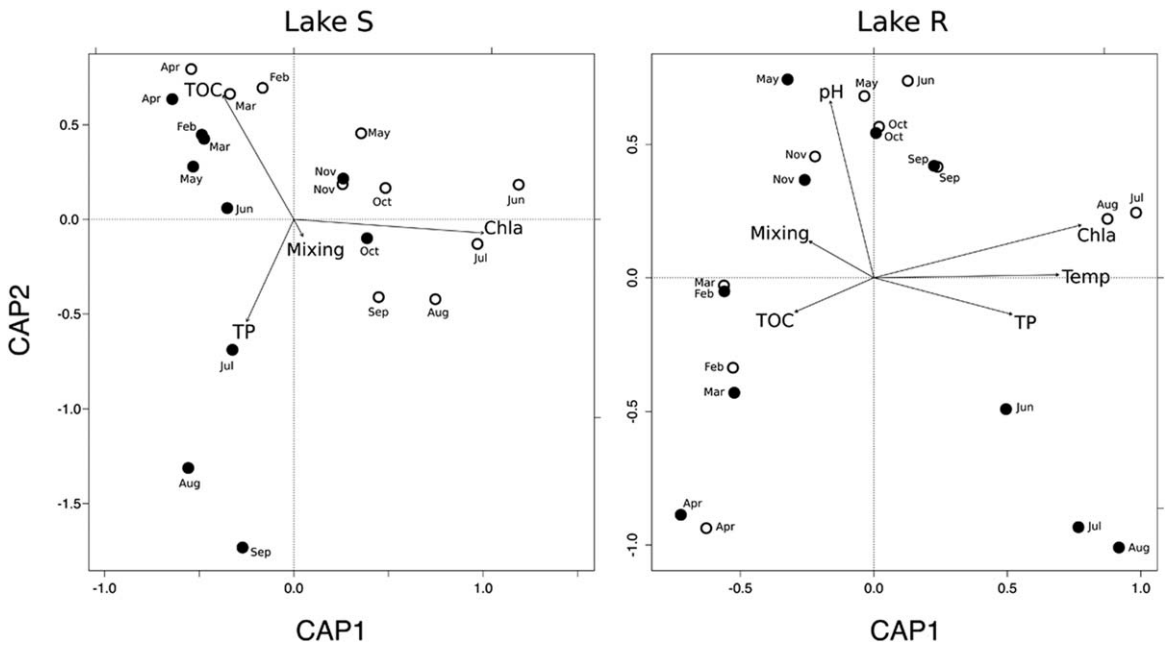

Fig. 3. Distance-based redundancy analysis of the bacterial community composition of epilimnion and hypolimnion of the two sampled lakes over the sampling season. Each circle represents individual observations of Lake R and Lake S. Open and black circles refer to epilimion and hypolimnion, respectively. The arrows represent the significant environmental vectors resulting from the forward selection analysis, and the vectors correspond to total phosphorus (TP), total organic carbon (TOC), chlorophyll-a (Chla), temperature (Temp) and water conductivity (Cond), pH and mixing events.

what is recorded by SourceTracker as a dispersal from the sediment to the lake could rather be because of dispersal in the other direction, thus, overestimating the role of sediment resuspension (Supporting Information Text S2). In summary, there was little evidence of an importance of mass effects due to sediment resuspension for the communities in epi- and hypolimnion.

Population level results: inlets and the other stratum of the lake

Concluding that mass effects from the different sources had little effect on community composition we continued investigating the potential effects of dispersal on the population level, i.e. if different OTUs could be affected differently over the season. We chose here to focus on the importance of dispersal from hypolimnia and inlets for bacteria in the epilimnia, since sediment and precipitation samples were so different from the lake samples (Fig. 1) and, thus, should have a lower number of relevant populations for this analysis. We here used SourceTracker model 2, which puts a greater focus on 'external' dispersal sources rather than internal growth.

For 57 selected OTUs (see selection criteria in Experimental Procedures below) the probabilities of dispersal into the epilimnion from the hypolimnion were in general greater than the probabilities of dispersal from the inlet into both lakes (Fig. 4, Supporting Information Table S4). Still, several OTUs showed a high probability of having the inlet as a source to Lake S, for example several Actinobacteria and several Betaproteobacteria (mostly belonging to *Limnohabitans;* betI and *Polynucleobacter*, betII sensu Newton *et al.* 2011) (Fig. 4, Supporting Information Fig. S3).

Further, for Lake S, the results showed that probabilities of dispersal from the inlet were generally higher in spring. Yet, several OTUs also showed high probabilities of dispersal from the inlet almost during the entire season, i.e. also when the inlet water flow was low and, thus, total dispersal rates were low. Among these were especially OTU25 (Actinobacterium, Luna1;Luna1-A;Luna1-A2), OTU8585 (Actinobacterium, acI;acI-A;acI-A6) and OTU7 (Betaproteobacterium, betI;betI-A;Lhab-A3) (Fig. 4 and Supporting Information Fig. S3). Some of these OTUs showed extremely high abundances in the inlets (Supporting Information Table S3).

For Lake S, the relative number of reads of these three OTUs in the epilimnion were analysed in relation to calculated dispersal rates of the respective OTU from the inlet as well as lake water chemistry and mixing regime by Partial least square (PLS) analysis (Fig. 5). PLS regression models identify latent factors that allow explaining a given fraction of the variance in Y (here the relative abundance of the three OTUs) and X variables (estimated flow rate of these OTUs from the inlet (see Experimental procedures for how this was calculated), water chemistry and mixing events). Here, the first two latent factors of the PLS models collectively

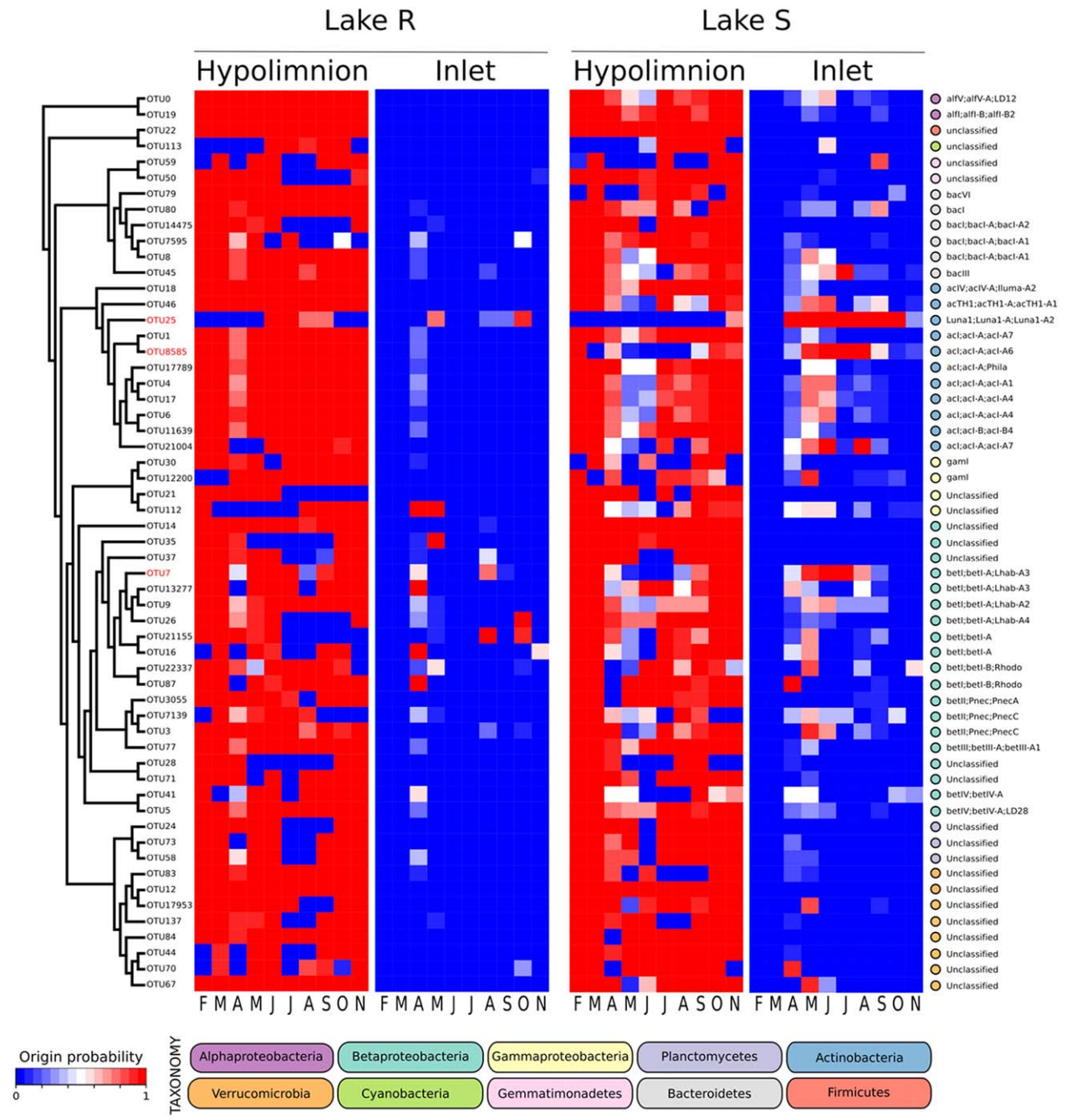

Fig. 4. Heatmaps showing the probability (according to SourceTracker analysis) of the inlet and the hypolimnion being sources of dispersal to the epilimnia for the 57 selected OTUs of Lakes Ramsjön (Lake R) and Siggeforasjön (Lake S). The OTUs in the plot are ordered according to their phylogenetic affiliation (EPA placement of OTUs short reads onto a phylogenetic reference tree as presented in Supporting Information Figure S3). In both lakes the probability of dispersal from the inlet was by definition zero in February and March since the inlets were frozen. The same applies for Lake R in June and July since its inlet had dried up.

explained 57.3, 92.1 and 92.6% of the variance in relative abundance of OTU 7, OTU 25 and OTU 8585, respectively. The results further showed that high relative abundances of OTU25 were related to high dispersal rates of this OTU from the inlet, while other factors were more strongly related to the high abundances of OTUs 8585 and 7 (Fig. 5). Thus, the strongest evidence for mass effect from the inlet on the population level was for OTU25. This OTU showed in general much higher proportion of reads in the inlet compared with the epilimnion (up to 21.9% of the reads, Supporting Information Table S3; Fig. S6), explaining why dispersal could be of importance also in low water flow seasons.

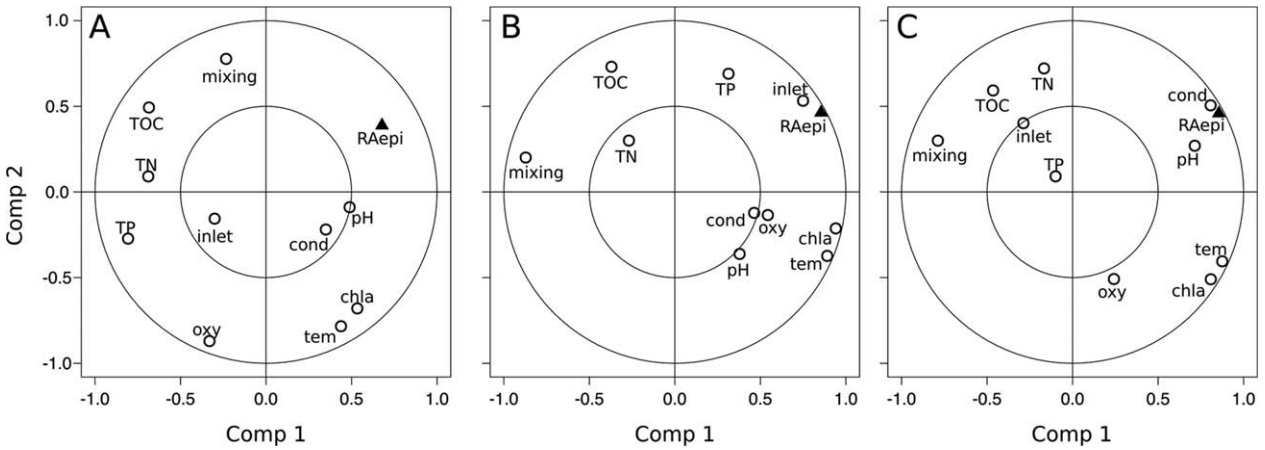

Fig. 5. Variables representation for PLS regression for OTU 7 (A), OTU25 (B) and OTU 8585 (C) in the epilimnion of Lake S. Independent (X) variables are represented by open circles and correspond to conductivity (cond), total concentrations of phosphorus (TP), nitrogen (TN) and organic carbon (TOC), Chlorophyll *a* (chla), oxygen concentration (oxy), temperature (tem) and pH. Inlet refers to the estimated flow rates of the selected OTUs from the inlet (see Experimental procedures for explanation) and mixing corresponds to a dummy variable representing a situation with or without mixing. The dependent (Y) variable is represented by black triangle and refers to the relative abundance (RA) of the selected OTUs in the lake S epilimnion. The inner and outer radii correspond to 50% and 100% explained variance respectively.

For Lake R the probabilities of the inlet being a source for the selected OTUs were generally lower than in Lake S (Fig. 4). The OTUs showing highest probabilities were primarily Betaproteobacteria (betI lineage with several OTUs) but also the Actinobacterium Luna1 (OTU25) showed a high probability at several occasions. However, no OTU showed a high probability of dispersing from the inlet to Lake R over the entire season since the inlet was dried out in summer, and probabilities were, therefore, zero by definition. Still, the same OTUs that showed higher relative abundances in the inlet compared with in the epilimnion of Lake S also did so in lake R (Supporting Information Fig. S6), indicating that they were important dispersers also here. We, however, did no PLS analysis of the individual OTUs in Lake R because of the lower number of data points from the inlet (6).

Discussion

Our analyses showed that for both lakes the most important source of bacteria was the resident community, while all investigated dispersal sources appeared to have limited immediate effect on the community level. Thus, the importance of mass effect on the community level appeared small, and we found little support for our first hypothesis. In detail, the influence of precipitation appeared low and we also found little evidence of an importance of resuspension of sediment bacteria to water column communities and populations. Mixing appeared to change community composition, but this change appeared more related to physico-chemical conditions than to the dispersal of cells associated with the mixing. The role of inlet bacterial dispersal for community composition seemed to be of moderate importance. In most cases, the estimated dispersal rates from the inlet were much lower than necessary for mass effect, i.e. 2.5–40% of standing stock per day (Lindström and Östman, 2011; Severin *et al.*, 2013; Souffreau *et al.*, 2014). Consequently, in summary, our results corroborate earlier findings of a limited importance of dispersal from air (Jones *et al.*, 2008; Jones and McMahon, 2009) and inlets for mass effects on the community level (Logue and Lindström, 2010; Adams *et al.*, 2014).

However, we also hypothesized (hypothesis #3) that the importance of mass effects could differ among taxa. Following the potential effect of dispersal on individual populations by SourceTracker analysis, several OTUs were identified as possible dispersers to the epilimnion from the inlet of both lakes. Surprisingly, this effect was also detected in low water flow seasons. The effect of dispersal of aquatic bacteria on community composition is always hard to distinguish from the effects of species sorting since bacterial cells are travelling with the media and, thus, great dispersal events due to high water flow are coinciding with great changes in environmental conditions (Lindström *et al.*, 2006). In line with this, as mentioned above, we also saw that the mixing effect we discovered by SourceTracker analysis could be explained to a great extent by a change in environmental conditions, i.e. species sorting, rather than by dispersal of cells by mixing in itself. This is a limitation of the SourceTracker analysis since it is only calculating probabilities based on abundances in sources and sinks and does not simultaneously evaluate other explanatory models, such as the importance of species sorting. However, in low flow seasons high abundances of certain OTUs in inlets can have an

effect on abundances in lakes, while environmental conditions are little affected. By PLS analysis we also demonstrated an importance of inlet dispersal for OTU25 even in relation to other potentially important factors. Thus, our results show some support for the existence of mass effects from the inlet on the population level.

Previous studies have shown that lake bacteria are in fact to a great extent a subset of soil bacteria (Crump *et al.*, 2012; Ruiz-González *et al.*, 2015) and that diverse soil and head water communities are successively depleted of typical soil taxa along a lake or river chain, while typical lake taxa are enriched (Crump *et al.*, 2012; Ruiz-González *et al.*, 2015; Savio *et al.*, 2015; Niño-García *et al.*, 2016). In this context, Ruiz-González and coworkers (Ruiz-González *et al.*, 2015) coined the term 'seed' bacteria for bacteria that originate from soils or headwater streams in low abundances but increase to their highest abundances in the lakes. Among those named as 'seed' bacteria (Ruiz-González *et al.*, 2015) are Betaproteobacteria and Actinobacteria, which were identified as potential inlet dispersers also in our lakes (Fig. 4).

However, in contrast to the idea of 'seed' bacteria (Ruiz-González *et al.*, 2015) our most obvious inlet disperser, the Actinobacterium *Luna1;Luna1-A;Luna1-A2* sensu Newton and coworkers (2011) (OTU25), showed another dispersal strategy. These bacteria appeared to have their abundance sustained in the lakes during long time periods by a high to extremely high abundance in the inlets (Figs 4 and 5, Supporting Information Table S3; Fig. S6). The substantial decrease in abundances from inlets to lakes (Supporting Information Fig. S6) indicates that this OTU was selected against by the lake environments, possibly due to differences in the preference for organic substrates (Crump *et al.*, 2012; Shabarova *et al.*, 2013; Adams *et al.*, 2014) or due to vulnerability to grazing (e.g. Simek *et al.*, 2014). The possibility of a vulnerability to grazing by dispersed bacteria is especially interesting considering that such taxa may serve as a subsidy rather than a source, contributing to higher trophic levels even if they never establish in the lake.

Another possible explanation for differences in immigration success among taxa is the resistance to invasions provided by the resident community by outcompeting some taxa dispersing into the lake, i.e. that there was an importance of priority effects (e.g. Hanson *et al.*, 2012). Priority effects have been demonstrated in bacteria in relatively simple laboratory systems (e.g. Fukami *et al.*, 2007; Tan *et al.*, 2012; Tucker and Fukami, 2014) but are so far little investigated in natural communities although indications of such effects exist (Andersson *et al.*, 2014; Ruiz-González *et al.*, 2015). The great importance of the 'past' community in our studies indicates that priority effects may have been important. A third category of OTUs might have contributed to this effect, i.e. those that are transported from the sources only to a very small extent, but still range among the most abundant OTUs in the lakes, i.e. being sustained by internal growth or mixing. For instance, the Alphaproteobacterium LD12 was only to a small extent introduced by dispersal from the inlet (Fig. 4). LD12 was among the most abundant OTUs in the lakes, but of considerably lower abundance in the inlets, especially in Lake R (Supporting Information Fig. S6). This 'tribe' was previously assigned as being limited to, but widespread and highly abundant, in freshwaters (e.g. Newton *et al.*, 2011), and might be a candidate for causing priority effects.

One hypothesis of our study was that the two lakes would differ in the importance of different sources of dispersal (hypothesis #2). Our results did, however, not show a great difference between them regarding the importance of the inlet, possibly because the water retention times of the lakes were rather similar (Supporting Information Table S1). The effect of the difference in lake area would also go unrecorded since dispersal by precipitation was so low. However, there appeared to be a slight difference between the lakes regarding the importance of sediment resuspension, since water samples from the shallower Lake R were more similar to Lake R sediment as opposed to Lake S. Still, we found little convincing evidence of a frequent migration of bacteria from sediment to water, and communities, in general, were extremely different. As mentioned in the results section, some hypolimnion samples in Lake R appeared to have been contaminated by sediments during sampling, and we noticed that the hypolimnion may rather be a source and the sediment the sink (Supporting Information Text S2). Thus, the contribution of the sediment as a source, as presented in Fig. 2, must be considered an overestimation.

The small effect of sediment resuspension is surprising considering that our lakes are relatively shallow and estimated resuspension rates were sometimes very high. However, since resuspension traps mostly contained lake bacteria rather than sediment bacteria, the estimated dispersal rates from the sediment were probably not accurate. A limited exchange of cells between lake sediment and water has been reported previously (Youngblut *et al.*, 2014) but it seems poorly investigated in general. For further studies, it should also be noted that we investigated the role of resuspension from the deepest point of the lakes and that our results do not exclude that resuspension could have occurred in shallower areas such as the littoral zone. We also expect that the role of sediment resuspension may be greater in more turbulent waters such as rivers (Staley *et al.*, 2015; 2016). Further, sampling of sediment bacteria might also be subject for methods development, for instance representative dispersers may be present only in the upper millimetres rather than centimetres as sampled here.

To summarize, we found that dispersal from sediments, inlets, the other strata of the lake and precipitation did not appear to have a great effect on community composition in the studied lakes. In contrast, the resident communities were more important. However, on the population level, the potential importance differed among bacterial groups, and may also not be coinciding with the highest total dispersal event from, for instance, the inlet. We therefore expect that for a complete understanding of biogeography as well as seasonal succession of bacterial communities in nature, research on dispersal on a population rather than community level would be informative, but is generally lacking (Martiny, 2015). Other under-explored aspects of dispersal for future studies are 1) the possible effects of dispersal over multiple temporal scales with some dispersing cells capable of immediate growth in the recipient habitat, while in other cases the effect on community composition is delayed (Comte *et al*., 2014), 2) the functional traits of the colonisers and the implication of the traits of the dispersed bacteria for community functioning (Severin *et al*., 2013) and 3) the effects of dispersal on higher trophic levels if dispersed bacteria are sensitive to grazing (e.g. Simek *et al*., 2014). Further, we expect that the inclusion of more lake types could have affected the results and conclusions slightly. We here focused on two relatively small and shallow forest lakes with lakes upstream in the catchment area. Possibly, including head water lakes could have resulted in different results (Nelson *et al*., 2009). In addition, being relatively undisturbed perhaps the communities of the lakes can be considered as relatively mature and, thus, able to resist colonization, while more disturbed lakes could have been subject to more immediate changes as a consequence of dispersal event. A greater effect of dispersal on community composition has, for instance, been shown in communities with depleted diversity (Zha *et al*., 2016).

Experimental procedures

Study sites, sampling, environmental variables and cell dispersal rate estimates

The studied lakes Siggeforasjön (Lake S) and Ramsjön (Lake R) were sampled monthly from 25th February 2011 to 24th November 2011 (on 10 occasions). Both lakes are small (<1 km^2) forest lakes, though Lake S is deeper, has a greater surface area, and a slightly longer water retention time than Lake R (Supporting Information Table S1). Both lakes have one main inlet draining $>50\%$ of the drainage area. The composition of the drainage area is similar for the two lakes (Supporting Information Table S1); for instance, none of them are first order lakes. The distance to the closest lake upstream is about 900 and 1000 meters away from the inlet of Lake S and Lake R, respectively. They are situated within 20 km from each other but do not lie within the same drainage area (Brunberg and Blomqvist, 1998). Based on total phosphorus values Lake S can be classified as relatively oligotrophic (<20 μg L^{-1}) and Lake R as meso-eutrophic (27-126 μg L^{-1}) during the study period. The pH varied between 5.6–6.9 and 6.0–7.5 in Lake S and Lake R, respectively, and TOC varied between 16.7–20.5 and 21.4–28.1 mg L^{-1}. Thus, both lakes can be classified as circumneutral brown water lakes.

On each sampling occasion, lake water column profiles were determined measuring water temperature (tem) and dissolved oxygen (oxy) concentration in situ using a combined temperature and dissolved oxygen probe (Oxi 340i; WTW, Weilheim, Germany). The temperature profiles were used to determine if lakes were mixed or not, i.e. if the temperature change was less than one degree per meter (vertically) the lake was considered to be unstratified. We do not have any information on whether lakes were mixed or not in between samplings. Water samples were collected from the epilimnion (at a depth of 1 meter below the water surface) and the hypolimnion (1 meter above the sediment). The collected water was used for characterizing bacterial community composition (see below) and water chemistry including total phosphorus (TP), nitrogen (TN) and organic carbon (TOC) as well as conductivity, pH and chlorophyll-a concentration (Chla). Details of the procedures are described in Comte *et al*. (2014). Two sediment cores were collected at the lakes' deepest points, using a gravity corer and from which the first 2 cm were sliced for further analysis. Bacterial communities in the inlet of each lake were sampled from shore approximately in the middle of the stream. No inlet samples could be collected in winter (February and March) because they were frozen, and in summer (June and July) the inlet of Lake R dried out, and, thus, no sample was obtained. Bacterial communities from precipitation and from sediment resuspension traps (see Supporting Information Text S1) were also collected.

Bacterial community composition (BCC)

Cells in water samples (lake, inlet, sediment resuspension traps, and precipitation) were collected by filtration onto 0.2 μm filter (Supor Membrane Disc Filters, 47 mm; Pall Corporation, East Hills, NY, USA) and stored at -80°C until further processing. Approximately 1g of lake sediment was placed into a sterile Eppendorf tube and kept frozen (–80° C). DNA extractions were performed using the PowerSoil DNA Isolation Kit (MO BIO Laboratories Inc., Carlsbad, CA, USA), following the manufacturer's instructions. The extraction of DNA from precipitation samples was generally difficult and successful in only one case (Lake S, May), which is why precipitation samples were removed from the dataset.

The V3-V4 regions of the 16S rRNA gene were amplified from the DNA extracts using the forward primer 341F and reverse primer 805R (Herlemann *et al*., 2011). Three technical replicates were PCR-amplified per sample, pooled after amplification, and purified using the Agencourt AMPure XP PCR purification kit (Beckman Coulter Inc., Brea, USA) (see Logue *et al*., 2012 for an in-detail description of primer modifications, PCR reactions, as well as conditions). Nucleic acids were eluted in 1X TE buffer and quantified using the Quant-iT^{TM} PicoGreen ds DNA quantification kit (Invitrogen, Carlsbad, USA). After pooling the 140 PCR amplicons in equimolar proportions, the final pooled amplicon was sequenced unidirectionally (Lib-L chemistry) on a 454 GS-FLX+ system

(454 Life Sciences, Branford, USA) at the SNP & SEQ Technology Platform (Uppsala University, Sweden; http://www.molmed.medsci.uu.se/SNP + SEQ + Technology + Platform/), using GS-FLX Titanium reagents.

Sequence processing

Sequence processing began with submitting the 454 pyrosequencing reads to quality filtering using AmpliconNoise (v1.24; Quince *et al.*, 2011). Thereby, sequences were removed that were shorter than 400 bp, did not match barcode and/or primer sequences, and identified as chimeric (Perseus; Quince *et al.*, 2011). After cleaning, a total of 1,648,334 reads were retrieved with on average 14,303 reads per sample in Lake S and its sources (min 4,948; max 19,532) and 13,376 reads in Lake R and its sources (min 4,322; max 19,829). Using QIIME (v1.6; Caporaso *et al.*, 2010), denoised sequences were clustered into operational taxonomic units (OTUs) using UCLUST (Edgar, 2010) at a 97% sequence similarity cut-off, whereupon singletons (OTUs with only one read in the dataset) were further removed. The remaining sequences were then aligned using PyNAST (Caporaso *et al.*, 2010) with the pre-aligned Greengenes 16S rRNA core set as template (DeSantis *et al.*, 2006). OTU representatives were taxonomically classified using UCLUST consensus taxonomy assigner (Edgar, 2010) as implemented in QIIME with 16S rRNA gene ARB sequence database as reference dataset (Newton *et al.*, 2011). As a whole, 22,053 OTUs were detected. The 454 sequences have been deposited in the NCBI Sequence Read Archive under accession number SRP049116.

The representative sequences of selected OTUs (see details below) were aligned using MUSCLE v3.8.425 (Edgar, 2004) and the resulting alignment was manually curated and trimmed. A maximum-likelihood phylogenetic tree (Fig. 4) was reconstructed using a general time reversible nucleotide substitution model with gamma-distributed rate variation across variable sites (GTR + G model) in RAXML v.8.1.24 (Stamatakis, 2014). Node support was computed from 1000 bootstrap replications.

To construct a reference phylogenetic tree (Supporting Information Fig. S3), we first performed BLASTn searches of the representative 'short' reads of 57 selected OTUs (see below) in order to identify the best hits (similarity $>$ 97%), retrieving a set of corresponding 'long' reference 16S rRNA gene sequences ($>$ 1300 bp) from Genbank. We also added a group of outgroup sequences. These references and outgroup reads (108 sequences in total) were aligned using MUSCLE v3.8.425 (Edgar, 2004) and the resulting alignment was inspected and trimmed. A reference tree was reconstructed from aligned long sequences using RAXML v.8.1.24 (Stamatakis, 2014). Specifically, we used GTR + G model to identify the best phylogenetic tree. Node support values of this reference tree were computed from 1000 bootstrap trees. Phylogenetic placement of the 'short' representative sequences of the selected OTUs onto the reference phylogenetic tree was performed using the RAXML v.8.1.24 (Stamatakis, 2014) evolutionary placement algorithm (Berger and Stamatakis, 2011) (Supporting Information Fig. S3). Prior to phylogenetic placements, reference, outgroup and short reads were aligned with MUSCLE v3.8.425 (Edgar, 2004) and the resulting alignment was inspected and trimmed using SeaView v.4 (Gouy *et al.*, 2010). Only common alignment positions between reference sequences and short reads were used for phylogenetic placements.

Calculation of bacterial dispersal rates

Cell dispersal rates (#cells/48 h) from the three sources (inlet inflow, precipitation, and sediment resuspension) were estimated for each sampling occasion as described in supplementary information (Supporting Information Text S1). In short, inlet dispersal rates were calculated based on bacterial numbers in the inlets (see below for method) multiplied by the estimated water discharge, while dispersal rates from precipitation and sediment resuspension was based on collected bacteria in resuspension and precipitation traps. These numbers were thereafter recalculated to %/48 h related to the total number of bacterial cells in the whole lake or in epilimnion and hypolimnion (in case of stratification) separately. For these calculations, when lakes were stratified, precipitation and surface inflow was assumed to disperse only into the epilimnion. The total numbers of bacterial cells in each stratum were calculated by multiplication of the bacterial abundance (cells L^{-1}, see below for method) with the lake volume or the volume of each stratum (L).

Total bacterial abundance in all water fractions was determined by use of flow cytometry (CyFlow space; Partec, Münster, Germany) of formaldehyde (final concentration 3.7% w/v) preserved samples after SYTO13 (Invitrogen) staining (del Giorgio *et al.*, 1996).

Community composition analyses

Community level results were analysed by use of Morisita-Horn (MH) distances as well as SourceTracker analysis (see details below). MH distances between samples were calculated on the full BCC dataset by use of the function vegdist in the R-package 'vegan' (Oksanen *et al.*, 2015) and presented in a two-dimensional NMDS ordination plot. In order to evaluate the relationship between calculated dispersal rates and similarities between dispersal sources and sinks (epilimnia and hypolimnion of the lakes) Spearman rank correlation analyses between dispersal rates and MH distances analyses were run. Mann-Whitney U-test was used to test for differences between lakes regarding MH distances between sources and sinks. Wilcoxon's paired test was used to investigate the importance of sedimentation trap location for MH distance between traps and sediment communities. All tests were run in R version 3.0.3 (R Core Team, 2014).

SourceTracker analyses

SourceTracker analysis (Bayesian based algorithm; Knights *et al.*, 2011) was used to identify the different dispersal sources and estimate their contribution (i.e. as proportions) to the bacterial community composition and population dynamics of the lakes. Independent SourceTracker analyses were carried out for each sink and sampling occasion; Lake S hypolimnion, Lake S epilimnion, Lake R hypolimnion and Lake R epilimnion at ten sampling events (i.e. forty in total). Two distinct models were run. Both models included as sources sediments and

inlets as well as the samples from the other stratum of the lake (i.e. hypolimnion for an epilimnion sink and vice-versa) from the respective sampling occasion. Model 1, however, also comprised sink communities from the preceding sampling occasion ('past') as a source. Thus, Model 1 had more focus on internal processes as compared with Model 2. The outputs of both models were used to evaluate the importance of the different sources for the communities in the different sinks over time (Fig. 2). 5000 sequences per sample were used; however for one sample with lower sequence numbers all the sequences were used.

SourceTracker was furthermore used to compute the probability by which individual OTUs in the sink communities dispersed from a specific source (population level analysis). For this analysis, we used the output of Model 2. Based on the results of the community level analysis (see results below) and to somewhat restrict the amount of information presented here, we – in the population level analysis – chose to present and scrutinize the results of the two epilimnia as sinks and hypolimnia as well as inlets as sources. OTUs were selected for an in-depth analysis according to the following criteria: 1) OTUs should make up at least 1% of the reads in any of the sinks on at least one occasion, 2) OTUs were identified (by SourceTracker Model 2) as having the inlet as a source on at least one occasion, 3) OTUs should have been identified by SourceTracker analysis (Model 2) to have the hypolimnion or the inlets as a source at least at 50% of the sampling occasions; that is 5 times for the hypolimnion and 3 (Lake R) or 4 (Lake S) times for the inlet. Via this procedure 57 OTUs were selected for which the probability of dispersal from the hypolimnion and inlet to the respective epilimnion over the season will be discussed and analysed in detail below.

Permutational manova (PERMANOVA, 1000 permutations) were performed, using the 'adonis' function as implemented in the R-package 'vegan' (Oksanen *et al*., 2015), to test for significant differences in the probabilities of dispersal of the selected OTUs from the hypolimnion or the inlet to the epilimnion of the lakes over the sampled season. In this analysis, we used the argument strata = lake ID in order to constrain permutations within each block (lake). Season was entered as Winter (February and March), Spring (April and May), Summer (June, July and August) and Fall (September, October and November). Probability values were arcsine-transformed prior to calculation of Morisita-Horn distances among the samples.

Multivariate analyses

Distance-based redundancy analysis (db-RDA), using the Morisita-Horn distances, was performed to test and identify the influence of physico-chemical variables as well as mixing events on community composition. The environmental matrix was composed of 8 variables including TP, TN and TOC concentrations, temperature, pH, conductivity, chlorophyll-a (Chl a) concentration and mixing events were recoded into binary data with 1 indicating mixing and 0 no mixing. Data were log-transformed with the exception of pH and mixing. Variable clustering analysis was applied to assess the redundancy of the environmental variables. Only environmental variables with Spearman correlation Rho values < 0.60 to each other were selected. TN was therefore removed from the matrix of environmental variables because of a high correlation with TP (Spearman Rho = 0.71). db-RDA was then performed using capscale command in the vegan R package and the significance of the model was tested after 1000 permutations. The significance of the explanatory variables was tested and the variance inflation factors (VIFs) were calculated for each variable. A forward selection procedure confirmed that the best db-RDA model included all included environmental variables. In addition, none of the VIF exceeded 10, which suggest that our model is unlikely affected by multicollinearity (Borcard *et al*., 2011).

To identify the variables that were most likely related to the dynamics of specific bacterial populations, we conducted partial least square (PLS) analyses (Wold *et al*., 2001). PLS fits linear models based on linear combinations of the explanatory variables (X). PLS is particularly robust towards imbalanced dataset (more X variables than observations) and collinearity among X variables. PLS analyses were performed using the SIMPLS algorithm and a leave-one-out cross-validation was performed as implemented in the pls 2.4 R-package (Mevik *et al*., 2013). Specifically, these analyses were aimed to model the dynamics of three particular OTUs (OTU7, OTU25 and OTU8585) in the epilimnion of Lake S given that they presented clear patterns that suggest that they were subject to dispersal from the inlet (see results). Therefore, three separate PLS models were run, the Y variable being the relative abundance the OTU of interest in the epilimnion of the lake and X variables were a set of dispersal related (mixing events, flow of these OTUs from the inlet) and environmental variables (dissolved oxygen, TP, TN and TOC concentrations, temperature, Chla, conductivity, pH). Mixing was used as described for the db-RDA. Flow of OTUs from the inlet to the lake was estimated at each sampling occasion as the product between the relative number reads of the particular OTU in the inlet by the percentages of total transported cells from the inlet in relation to the number of cells in the epilimnion over 48 h (see above for method). Missing values in X (conductivity) were replaced by model prediction using the NIPALS algorithm as implemented in the mixOmics package in R (Dejean *et al*., 2014). Prior to analysis, environmental data were log transformed (except for pH) to attain a normal distribution and the Y variable was arcsin transformed. Outcomes from PLS were visualized from a loading plot that presents the relationship among the variables with high correlation among variables that cluster together.

Acknowledgments

We thank Yinghua Zha for assistance with DNA extraction and preparation for sequencing. We are grateful to three anonymous reviewers and the Editor for their valuable comments on a previous version of the manuscript. This study was supported by grants from the Swedish research council to E.S. Lindström (project number 2009-5172), from the Wenner Gren foundation to E.S. Lindström and J. Comte, and from the Tryggers foundation to E.S. Lindström and Ö. Östman. 454 pyrosequencing was enabled by an instrument grant from the K&A Wallenberg Foundation and the SciLifeLab SNP&SEQ Technology Platform hosted by Uppsala University. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project b2010008.

Support by D. Lundin from BILS (Bioinformatics Infrastructure for Life Sciences) is gratefully acknowledged. The authors declare no conflict of interest.

References

Adams, H.E., Crump, B.C., and Kling, G.W. (2014) Metacommunity dynamics of bacteria in an arctic lake: the impact of species sorting and mass effects on bacterial production and biogeography. *Front Microbiol* **5:** 82.

Andersson, M.G.I., Berga, M., Lindström, E.S., and Langenheder, S. (2014) The spatial structure of bacterial communities is influenced by historical environmental conditions. *Ecology* **95:** 1134–1140.

Barberán, A., and Casamayor, E.O. (2010) Global phylogenetic community structure and beta-diversity patterns in surface bacterioplankton metacommunities. *Aquat Microb Ecol* **59:** 1–10.

Berger, S.A., and Stamatakis, A. (2011) Aligning short reads to reference alignments and trees. *Bioinformatics* **27:** 2068–2075.

Bissett, A., Richardson, A.E., Baker, G., Wakelin, S., and Thrall, P.H. (2010) Life history determines biogeographical patterns of soil bacterial communities over multiple spatial scales. *Mol Ecol* **19:** 4315–4327.

Borcard, D., Gillet, F., and Legendre, P. (2011) *Numerical Ecology with R*. New York: Springer.

Brunberg, A.-K., and Blomqvist, P. (1998) *Vatten i Uppsala Län 1997*. Upplandsstiftelsen, Rapport nr 8/1998 p. 944.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., *et al*. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7:** 335–336.

Comte, J., Lindström, E.S., Eiler, A., and Langenheder, S. (2014) Can marine bacteria be recruited from freshwater sources and the air? *ISME J* **8:** 2423–2430.

Crump, B.C., Amaral-Zettler, L.A., and Kling, G.W. (2012) Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soil. *ISME J* **6:** 1629–1639.

Dejean, S., Gonzalez, I., and Le Cao, K.-A. (2014) mixOmics: Omics Data Integration Project. R package version 5.0-3. http://CRAN.R-project.org/package=mixOmics

del Giorgio, P.A., Bird, D.F., Prairie, Y.T., and Planas, D. (1996) Flow cytometric determination of bacterial abundance in lake plankton with the green nucleic acid stain SYTO 13. *Limnol Oceanogr* **41:** 783–789.

DeLeon-Rodriguez, N., Lathem, T.L., Rodriguez-R, L.M., Barazesh, J.M., Anderson, B.E., Beyersdorf, A.J., *et al*. (2013) Microbiome of the upper troposphere: species composition and prevalence, effect of tropical storms, and atmospheric implications. *Proc Natl Acad Sci USA* **110:** 2575–2580.

de Oliveira, L.F.V., and Margis, R. (2015) The source of the river as a nursery for microbial diversity. *PLoS One* **10:** e0120608.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., *et al*. (2006) Greengenes, a chimera-checked 16S rRNA gene data base and workbench compatible with ARB. *Appl Environ Microbiol* **72:** 5069.

Edgar, C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32:** 1792–1997.

Edgar, C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26:** 2460–2461.

Fahlgren, C., Hagström, Å., Nilsson, D., and Zweifel, U.L. (2010) Annual variations in the diversity, viability, and origin of airborne bacteria. *Appl Environ Microbiol* **76:** 3015–3025.

Fahlgren, C., Gómez-Consarnau, L., Zábori, J., Lindh, M.V., Krejci, R., Mårtensson, E.M., *et al*. (2015) Seawater mesocosm experiments in the Arctic uncover differential transfer of marine bacteria to aerosols. *Environ Microbiol Rep* **7:** 460–470.

Fukami, T., Beaumont, H.J.E., Zhang, X.X., and Rainey, P.B. (2007) Immigration history controls diversification in experimental adaptive radiation. *Nature* **446:** 436–439.

Gouy, M., Guindon, S., and Gasquel, O. (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27:** 221–224.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B.H. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10:** 497–506.

Herlemann, D.P.R., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J.J., and Andersson, A.F. (2011) Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J* **5:** 1571–1579.

Jones, S.E., and McMahon, K.D. (2009) Species-sorting may explain an apparent minimal effect of immigration on freshwater bacterial community dynamics. *Environ Microbiol* **11:** 905–913.

Jones, S.E., Newton, R.J., and McMahon, K.D. (2008) Potential for atmospheric deposition of bacteria to influence bacterioplankton communities. *FEMS Microbiol Ecol* **64:** 388–394.

Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., *et al*. (2011) Bayesian community-wide culture-independent microbial source tracking. *Nat Methods* **8:** 761–763.

Lear, G., Bellamy, J., Case, B.S., Lee, J.E., and Buckley, H. (2014) Fine-scale spatial patterns in bacterial community composition and function within freshwater ponds. *ISME J* **8:** 1751–1726.

Lindström, E.S., and Östman, Ö. (2011) The importance of dispersal for bacterial community composition and functioning. *PLoSOne* **6:** e25883.

Lindström, E.S., and Langenheder, S. (2012) Local and regional factors influencing bacterial community assembly. *Environm Microbiol Rep* **4:** 1–9.

Lindström, E.S., Forslund, M., Algesten, G., and Bergström, A.-K. (2006) External control of bacterial community structure in lakes. *Limnol Oceanogr* **51:** 339–342.

Logue, J.B., and Lindström, E.S. (2010) Species sorting affects bacterioplankton community composition as determined by 16S rDNA and 16S rRNA fingerprints. *ISME J* **4:** 729–738.

Logue, J.B., Langenheder, S., Andersson, A.F., Bertilsson, S., Drakare, S., Lanzén, A., and Lindström, E.S. (2012) Freshwater bacterioplankton richness in oligotrophic lakes depends on nutrient availability rather than on species-area relationships. *ISME J* **6:** 1127–1136.

Martiny, J.B.H. (2015) Dispersal and the microbiome. *Microbe* **10:** 191–196.

Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., *et al*. (2006) Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4:** 102–112.

Mevik, B.-H., Wehrens, R., and Liland, K.H. (2013) P Bjørn-Helge Mevik, Ron Wehrens and Kristian Hovde Liland (2013). pls: Partial Least Squares and Principal Component regression. R package version 2.4-3. http://CRAN.R-project.org/package=pls

Nelson, C.E., Sadro, S., and Melack, J.M. (2009) Contrasting the influences of stream inputs and landscape position on bacterioplankton community structure and dissolved organic matter composition in high-elevation lake chains. *Limnol Oceanogr* **54:** 1292–1305.

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. *Microbiol Mol Biol Rev* **75:** 14–49.

Newton, R.J., Bootsma, M.J., Morrison, H.G., Sogin, M.L., and McLellan, S.L. (2013) A microbial signature approach to identify fecal pollution in the waters off an urbanized coast of Lake Michigan. *Microbial Ecol* **65:** 1011–1023.

Niño-García, J.P., Ruiz-González, C., and del Giorgio, P.A. (2016) Interactions between hydrology and water chemistry shape bacterioplankton biogeography across boreal freshwater networks. *ISME J* **10:** 1755–1766.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., *et al*. (2015) vegan: Community Ecology Package. R package version 2.2-1. http://CRAN.R-project.org/package=vegan

Peter, H., and Sommaruga, R. (2016) Shifts in diversity and function of lake bacterial communities upon glacier retreat. *ISME J* **10:** 1545–1554.

Quince, C., Lanzén, A., Davenport, R.J., and Turnbaugh, P.J. (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* **12:** 38.

R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Read, D.S., Gweon, H.S., Bowes, M.J., Newbold, L.K., Field, D., Bailey, M., and Griffiths, R.I. (2015) Catchment-scale biogeography of riverine bacterioplankton. *ISME J* **9:** 516–526.

Ruiz-González, C., Niño-García, J.P., and del Giorgio, P.A. (2015) Terrestrial origin of bacterial communities in complex boreal freshwater networks. *Ecol Lett* **18:** 1198–1206.

Savio, D., Sinclair, L., Ijaz, U.Z., Parajka, J., Reischer, G.H., Stadler, P., *et al*. (2015) Bacterial diversity along a 2600 km river continuum. *Environ Microbiol* **17:** 4994–5007.

Severin, I., Östman, Ö., and Lindström, E.S. (2013) Variable effects of dispersal on productivity of bacterial communities due to changes in functional trait composition. *PLoSOne* **8:** e80825.

Shabarova, T., Widmer, F., and Pernthaler, J. (2013) Mass effects meet species sorting: transformations of microbial assemblages in epiphreatic subsurface karst water pools. *Environ Microbiol* **15:** 2476–2488.

Shade, A., Kent, A.D., Jones, S.E., Newton, R.J., Triplett, E.W., and McMahon, K.D. (2007) Interannual dynamics and phenology of bacterial communities in a eutrophic lake. *Limnol Oceanogr* **52:** 487–494.

Simek, K., Nedoma, J., Znachor, P., Kasalicky, V., Jezbera, J., Hornak, K., and Sed'a, J. (2014) A finely tuned symphony of factors modulates the microbial food web of a freshwater reservoir in spring. *Limnol Oceanogr* **59:** 1477–1492.

Souffreau, C., Pecceu, B., Denis, C., Rummens, K., and De Meester, L. (2014) An experimental analysis of species sorting and mass effects in freshwater bacterioplankton. *Freshw Biol* **59:** 2081–2095.

Staley, C., Gould, T.J., Wang, P., Phillips, J., Cotner, J.B., and Sadowsky, M.J. (2015) Species sorting and seasonal dynamics primarily shape bacterial communities in the upper Mississippi river. *Sci Tot Environ* **505:** 435–445.

Staley, C., Gould, T.J., Wang, P., Phillips, J., Cotner, J.B., and Sadowsky, M.J. (2016) Sediments and soils act as reservoirs for taxonomic and functional bacterial diversity in the upper Mississippi river. *Microb Ecol* **71:** 814–824.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30:** 1312–1313.

Székely, A.J., and Langenheder, S. (2013) The importance of species sorting differs between habitat generalists and specialists in bacterial communities. *FEMS Microbiol Ecol* **87:** 102–112.

Tan, J., Pu, Z., Ryberg, W.A., and Jiang, L. (2012) Species phylogenetic relatedness, priority effects, and ecosystem functioning. *Ecology* **93:** 1164–1172.

Tucker, C.M., and Fukami, T. (2014) Environmental heterogeneity counteracts priority effects to facilitate species coexistence: evidence from nectar microbes. *Proc R Soc B* **281:** 20132637.

Wold, S., Sjöström, M., and Eriksson, L. (2001) PLS-regression: A basic tool of chemometrics. *Chemometr Intell Lab Syst* **58:** 109–130.

Youngblut, N.D., Dell'aringa, M., and Whitaker, R.J. (2014) Differentiation between sediment and hypolimnion methanogen communities in humic lakes. *Environ Microbiol* **16:** 1411–1423.

Zha, Y., Berga, M., Comte, J., and Langenheder, S. (2016) Effects of dispersal and initial diversity on the composition and functional performance of bacterial communities. *PLoS One* **11:** e0155239.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Table S1. Characteristics of the two lakes sampled. Data is from (Brunberg and Blomqvist, 1998)

Table S2. Estimated rates of dispersal from the different sources (columns) to the different sinks (rows) during the period of study (% of standing stock in the respective stratum over 48 h). The range gives the maximum and minimum numbers and the average in brackets. n.c. = not calculated.

Table S3. Total number of OTUs and taxonomic assignment with ARB (Newton *et al*., 2011) or † Genbank databases of the five most abundant OTUs detected in the investigated dispersal sources and sinks for each of the two sampled lakes. *Data were collected for 6 (Lake R) or 8 (Lake S) of the 10 months of the sampling season. Maximum relative number of reads are shown in brackets (%).

Table S4. Results from PERMANOVA analyses testing whether the origin of the selected 57 OTUs (see figure 4) was driven only by the source of dispersal (hypolimnion, inlet), or varied according season or by the seasonal variability in the dispersal sources. The analysis was performed on the probabilities of these to disperse from two sources (hypolimnion, inlet) to the epilimnion throughout the study period in both lakes. Season refers to the sampling season (winter, spring, summer and fall). P values underwent bonferroni correction for multiple comparisons. Significant p-values (<0.05) are shown in bold.

Fig. S1. Surface water flow to the two lakes during the year of study (2011). Black diamonds: daily averages for all days. White squares: daily averages at the days of sampling. Letters along the x-axes denote the month of the year. Estimates of water flow rates (m^3/s) for Lake S were calculated using data of daily averages of outflow rates modelled by the Swedish Meteorological and Hydrological Institute (SMHI; the S-HYPE 2012 model, www.smhi.se, downloaded on 26th November 2013). Knowing the size of the drainage area for the lake, the specific runoff was calculated ($m^3*s^{-1}*km^{-2}$). Multiplying this runoff with the size of the area draining to the lake (Brunberg and Blomqvist 1998), inflow rates were calculated (Figure S1). Modelled flow rates data were not available for Lake R, instead outflow rates from the nearby Lake Strandsjön (within 4 km from Lake R) were used to calculate specific runoff and inflow rate was calculated in the same manner as for Lake S.

Fig. S2. Relative number of reads of OTU #15 (*Chlorobium*) in the hypolimnion and sediment of Lake R during the study period. Letters refer to month of sampling.

Fig. S3. Phylogenetic mapping of 57 selected OTUs (see text). Rooted reference phylogenetic tree (left phylogram) was reconstructed using maximum-likelihood, and node statistical supports were computed using 1000 rapid bootstrap replicates (only support values of >70 are shown). OTU reads were mapped onto the reference tree using RAxML evolutionary placement algorithm (right cladogram). Phylogenetic placements are indicated by red branches and the corresponding likelihood weight is indicated in brackets. The corresponding taxonomic assignment using ARB database (Newton *et al*., 2011) is given. Outgroup (not shown) included Spirochaeta representatives (HE962136).

Fig. S4. Boxplot of the ratio between the relative number of reads in inlet and in the epilimnion of the two lakes for a selected number of OTUs (see Table SI3, Figure 4 and Figure SI3 for more information about the OTUs). This ratio was calculated for each month (different colours in the plot). When the number of reads was 0 in the epilimnion the lowest relative abundance in the data set was used. The figure shows the value of log ($x+1$). The dashed line corresponds to a ratio of 1, i.e. data points above the line represent a higher abundance in the inlet than in the epilimnion and data points below the line represents a higher abundance in the epilimnion than in the inlet.