

ORIGINAL ARTICLE

Coherent dynamics and association networks among lake bacterioplankton taxa

Alexander Eiler, Friederike Heinrich and Stefan Bertilsson

Department of Ecology and Genetics, Limnology, Uppsala University, Uppsala, Sweden

Bacteria have important roles in freshwater food webs and in the cycling of elements in the ecosystem. Yet specific ecological features of individual phylogenetic groups and interactions among these are largely unknown. We used 454 pyrosequencing of 16S rRNA genes to study associations of different bacterioplankton groups to environmental characteristics and their co-occurrence patterns over an annual cycle in a dimictic lake. Clear seasonal succession of the bacterioplankton community was observed. After binning of sequences into previously described and highly resolved phylogenetic groups (tribes), their temporal dynamics revealed extensive synchrony and associations with seasonal events such as ice coverage, ice-off, mixing and phytoplankton blooms. Coupling between closely and distantly related tribes was resolved by time-dependent rank correlations, suggesting ecological coherence that was often dependent on taxonomic relatedness. Association networks with the abundant freshwater *Actinobacteria* and *Proteobacteria* in focus revealed complex interdependencies within bacterioplankton communities and contrasting linkages to environmental conditions. Accordingly, unique ecological features can be inferred for each tribe and reveal the natural history of abundant cultured and uncultured freshwater bacteria.

The ISME Journal (2012) 6, 330–342; doi:10.1038/ismej.2011.113; published online 1 September 2011

Subject Category: microbial population and community ecology

Keywords: freshwater; 454 pyrosequencing; 16S rRNA; phylogenetic tribes; bacteria

Introduction

Exploration of microbial diversity with molecular methods has enabled the identification of abundant and characteristic freshwater bacterioplankton groups (for example, Bahr *et al.*, 1996; Zwart *et al.*, 2002; Eiler and Bertilsson, 2004; Newton *et al.*, 2011). Such approaches applied to amplicons from the 16S rRNA gene have also revealed that bacterial community composition varies among and within lakes over contrasting temporal and spatial scales (for example, Simek *et al.*, 1999; Lindström, 2000; Van der Gucht *et al.*, 2001; Kent *et al.*, 2004; Newton *et al.*, 2006; Shade *et al.*, 2007; Hutalle-Schmelzer and Grossart, 2009; Jones *et al.*, 2009). Bacterial communities and many of the abundant bacterial groups have been suggested to hold central roles in freshwater food webs (Pernthaler, 2005) and the biogeochemical processes in these ecosystems (Cotner and Biddanda, 2002). It has been suggested that these functional traits are products of multiple interacting populations within these communities, rather than those of single populations (Little *et al.*,

2008; Strom, 2008), but there are little empirical data supporting such concepts. Cultivation approaches (Hahn, 2009; Jezbera *et al.*, 2009; Hahn *et al.*, 2010a,b), application of targeted quantitative methods (Pernthaler *et al.*, 1998, 2004; Simek *et al.*, 1999; Eiler and Bertilsson, 2007; Salcher *et al.*, 2008, 2010) and metagenomic characterization (Rusch *et al.*, 2007; Debroas *et al.*, 2009) have improved our current understanding about the ecology of abundant freshwater bacterioplankton groups. Recently, association network approaches have been proposed for exploring the broad range of interactions that are likely to take place among bacterial populations that jointly make up the typically very complex bacterioplankton communities characteristic for aquatic systems (Ruan *et al.*, 2006; Fuhrman and Steele, 2008; Fuhrman, 2009; Steele *et al.*, 2011).

Inspired by these studies, we assessed bacterial interdependencies by examining co-occurrence patterns among bacterial groups and their correlations to environmental properties in a temporal survey of freshwater bacterioplankton. Epilimnetic bacteria from a productive dimictic lake located in south-east Sweden (Lake Erken) were studied over an annual cycle including periods of ice cover, stratification, mixing and phytoplankton blooms. The V4 region of bacterial small subunit rRNA genes were amplified from mixed community DNA and sequenced *en masse* by multiplexed 454 pyrosequencing. This methodology enabled a detailed

Correspondence: S Bertilsson, Department of Ecology and Genetics, Limnology, Uppsala University, Norbyvägen 18D, Uppsala SE-75236, Sweden.

E-mail: stebe@ebc.uu.se

Received 28 February 2011; revised 11 July 2011; accepted 12 July 2011; published online 1 September 2011

account of the temporal dynamics of combined bacterial communities in ways comparable to earlier studies based on fingerprinting techniques (for example, Kent *et al.*, 2004; Lindström *et al.*, 2005; Yannarell and Triplett, 2005; Judd *et al.*, 2006; Newton *et al.*, 2006; Bertilsson *et al.*, 2007; Nelson, 2009). However, it also enabled us to study the temporal dynamics of individual bacterial populations annotated into taxonomic groups. Clustering based on temporal dynamics was used to relate the representation of taxonomic groups (tribes) to various seasonal events. Local similarity analysis (LSA; Ruan *et al.*, 2006) revealed contemporaneous and time-lagged correlation patterns among populations in the bacterioplankton community members and associations to environmental variables. We then analyzed and visualized correlation patterns as association networks and sub-networks that provided valuable insights about the natural history of some abundant freshwater bacterioplankton groups.

Materials and methods

Sampling and contextual environmental variables

Surface water samples were obtained from dimictic Lake Erken (59°51'N, 18°36'E) at monthly to weekly intervals spanning a total period from February 2007 to January 2008 (33 occasions). This study period included three phytoplankton blooms, summer stratification, fall turnover of the lake and periods of ice cover. The lake is mesotrophic and typically experiences extensive periods of ice coverage and a summer stratification period that starts to develop in late May (for example, Bell *et al.*, 1998). Previous work has demonstrated that the composition and activity of the planktonic community, including bacteria, vary over season (Bell *et al.*, 1998; Bertilsson *et al.*, 2007; Eiler and Bertilsson, 2007). Water samples from discrete depths (1 m depth intervals) were collected at the deepest point of the lake. Equal amounts of subsamples were combined either from the entire 20-m-deep water column or from the upper mixed layer (epilimnion) during summer stratification (3 June to 6 August). Temperature and oxygen concentrations were analyzed for each depth by using a portable Oxi 340i oxygen meter equipped with a Cellox 325-20WTW probe. The pooled water samples were further characterized for chemical characteristics at the Lake Erken Field Station certified water chemistry laboratory by following SS-ISO standard methods (Supplementary Table 1). The parameters analyzed included pH, conductivity, nitrate, nitrite, alkalinity, turbidity, suspended matter, phosphate, ammonium, total and particulate phosphorus and nitrogen, chlorophyll-*a*, loss on ignition, silicate, water color and absorbance at 420 nm. The loss on ignition gives a crude measure of the organic content of the particulate matter. Bacterial cells for DNA-based community analyses were collected from 0.5 l of

lake water by gentle vacuum filtration onto 0.2 µm membrane filters (Supor-200 Membrane Disc Filters, 47 mm; Pall Corporation, East Hills, NY, USA). Individual filters were stored frozen until further processing. Samples for bacterial abundance were preserved in 50% ethanol and stored at -20 °C followed by counting of cells by flow cytometry after Syto13 staining (del Giorgio *et al.*, 1996). A comparison of samples preserved with either formaldehyde or ethanol gave highly similar results (not shown). The metadata are given in MIMARKS format (Supplementary Table 2).

DNA extraction, PCR amplification, pyrosequencing and sequence quality control

DNA was extracted by bead-beating and solid-phase extraction by using the Ultra clean Soil DNA extraction kit as recommended by the manufacturer (MoBio, Laboratories, Solana Beach, CA, USA). The quality and amount of extracted DNA were determined by agarose gel electrophoresis (1% agarose, 0.5 × TBE) followed by detection by ethidium bromide staining, UV transillumination and image analysis (Gel-Pro Analyzer, Version 3.1; Media Cybernetics Inc., Bethesda, MD, USA). Bands were compared against a DNA ladder (High Mass ladder; Invitrogen, Carlsbad, CA, USA). The extracts contained 1–20 ng DNA per microliter. Bacterial 16S rRNA genes (*Escherichia coli* position 341–805) were amplified by using general bacterial primers 341F (CCTACGGGNGGCWGCAG) and 805R (GACTACHVGGGTATCTAATCC) (Herlemann *et al.*, 2011). This primer pair matches approximately 90% of all good-quality bacterial sequences (>1200 bp) and covers all phyla in the Ribosomal Database Project release 10.25. Primer 341F carried a 454FLX adaptor B at the 5' end and primer 805R carried a 5-bp molecular barcode specific for each sample (Supplementary Table 2) followed by a 454FLX adaptor A at the 5' end. Each DNA sample was individually PCR-amplified in triplicate 20-µl reactions by initial denaturation at 95 °C for 5 min followed by 25 cycles of 40 s at 95 °C, 40 s at 53 °C and 60 s at 72 °C. At the end of the amplification, the amplicons were subjected to a final 7-min extension at 72 °C. Each reaction contained between 1 and 10 ng of target DNA, 1 nM Phusion HF buffer (Finnzymes, Espoo, Finland), 0.02 nM Phusion DNA polymerase (Finnzymes), 0.2 mM dNTPs and 0.5 µM of each primer. Replicate PCRs were pooled and the amplicon concentration for each sample was assessed by electrophoretic separation on 1% agarose gel followed by ethidium bromide staining and UV transillumination using a cooled CCD camera and comparison to a Low DNA Mass Ladder (Invitrogen). Equal amounts of uniquely barcoded amplicons from each of the samples were combined and gel-purified by using the Qiaquick gel purification kit (Qiagen, Hilden, Germany) as recommended by the manufacturer. The purified amplicons were

again sized and quantified (see above), followed by high-throughput pyrosequencing from adaptor A using the 454 GS-FLX system (454 Life Sciences, Branford, CT, USA), where the resulting reads carried the sample-specific molecular barcode and covered the entire V4 region of the 16S rRNA gene as well as more conserved flanking regions. Sequencing was performed at the Centre for Metagenomic Sequence Analysis hosted by the Royal Institute of Technology, Stockholm, Sweden. The sequence run included also samples that were not part of this study. From a total of almost 200 000 sequence reads, 120 564 sequence reads were assigned to samples used in this study based on their barcode sequence. Ambiguous sequences were removed from the data set, including reads with <200 bases and quality scores <25. Also sequences that did not carry the exact primer sequence; sequences containing ambiguous bases (N) and sequences with homopolymer stretches longer than eight bases were removed from the data set. After implementation of these quality-control criteria, 91 187 sequences were retained for further analysis. The 454 sequence run has been deposited in the NCBI Short Read Archive under accession number SRA029594.1.

OTU assignments and community analysis

The `unique.seqs.` command implemented in MOTHUR version 1.13.0 (Schloss *et al.*, 2009) was used to obtain a non-redundant set of sequences from the high-quality reads. The resulting 18 241 unique sequences were aligned against a local freshwater bacterial sequence database including almost 12 000 sequence entries (Newton *et al.*, 2011). Aligned sequences were then filtered to remove columns that corresponded to '.' or '-' (gaps) in all sequences. To reduce pyrosequencing noise, a pre-clustering step was performed by using the `pre.cluster` command in MOTHUR. This step bins sequences that are less abundant and differ from the dominant sequence type by less than 1%. Again, aligned sequences were filtered to remove columns that corresponded to '.' or '-' in all sequences. These alignments were then used to generate an uncorrected pairwise distance matrix by using the `dist.seqs` command in MOTHUR. To assign sequences into operational taxonomic units (OTUs), single linkage clustering was applied with a 97% sequence similarity cut-off as described previously (Huse *et al.*, 2010). All OTUs defined at the 0.03 cut-off were classified by combining the `classify.seqs` and `classify.otus` commands in MOTHUR (for more details on the workflow see Supplementary Figure 1), and a matrix of the OTU abundances for each sample was generated.

In order to compare and perform statistics across samples, we used the Perl script `daisychopper.pl` (available at <http://www.genomics.ceh.ac.uk/GeneSwytch/Tools.html>; Gilbert *et al.*, 2009). `Daisychopper` identifies the minimum number of

reads (in our case the minimum number of reads was 1612 per sample) for any individual sample in the OTU abundance matrix and then randomly re-samples all other samples to this minimum. Non-metric multidimensional scaling (NMDS) plots were used to visualize seasonal dynamics in community structure (β -diversity) by using the OTU abundance matrix from the single linkage clustering. NMDS plots were generated from Bray–Curtis similarity index matrices of the 33 different samples by using the function `metaMDS` in the `Vegan` library program implemented in R (<http://www.r-project.org/>).

Taxonomic placement

All 18 241 unique sequences were attributed to taxonomic analyses by using a naïve Bayesian approach (Wang *et al.*, 2007) implemented in the `classify.seqs` command in MOTHUR in combination with two template sequence databases (a freshwater sequence database and the Silva 102 database) and their corresponding taxonomic hierarchy frameworks (a freshwater taxonomy as suggested by Newton *et al.*, 2011 and the `rdp` taxonomic framework). We usually refer to the placements based on the freshwater bacterioplankton taxonomy framework described by Newton *et al.* (2011) and secondarily to the `rdp` taxonomy framework when sequences could not be affiliated with a freshwater taxon (see Supplementary Tables 3 and 4 for comparison). The former uses a hierarchical naming structure (phylum/lineage/clade/tribe) similar to the Linnean taxonomy while it is based on phylogenetics. The most refined taxonomic group is the tribe, consisting of a group of full-length sequences clustering as a monophyletic branch of a phylogenetic tree, with each sequence having $\geq 97\%$ sequence identity to another sequence of that branch. This classification system was designed to maintain the phylogenetic context by which freshwater bacterial gene sequences historically have been identified, clustered and named. Thus, such a unified freshwater taxonomic framework allowed the comparison of ecological observations among previously published freshwater surveys (reviewed by Newton *et al.*, 2011) and our study.

Statistical analyses of tribe dynamics

Reads were binned into well-defined, highly resolved phylogenetic groups as described above. In order to summarize seasonal patterns in the dynamic representation of amplicons from freshwater tribes, we performed k-means clustering using the Hartigan–Wong algorithm (Hartigan and Wong, 1979) after statistical re-sampling for equal sample size (using `daisychopper`; Gilbert *et al.*, 2009). As normal distribution was assumed, standard normal distribution and its related z-score could be used for standardization. z-score is convenient in comparing dynamics among populations as both the mean and

standard deviation are used for the standardization (Legendre and Legendre, 1998; Yannarell and Triplett, 2005). The z-score is $x_{ij}-\mu/\sigma$, where x is the number of reads in each sample i for each tribe j and μ is the mean number of reads of each tribe j among all samples i , and σ is the corresponding standard deviation.

Using the re-sampled (not z-score-standardized) data matrix, significant correlations ($P>0.01$) with a false discovery rate below 0.05 among tribes and between tribes and environmental variables were identified by using LSA (Ruan *et al.*, 2006) as implemented in R (<http://www.r-project.org/>). Other parameters for the LSA were set to 1000 permutations and a maximum delay of 3. As sampling points are not evenly spaced, delays (time lags) can correspond to periods ranging from a week to several months. The resulting correlation matrix was translated into an association network by using Cytoscape 2.6.3 (Shannon *et al.*, 2003). These statistical methods have been used in previous studies on both marine and freshwater bacterioplankton communities (see for example, Fuhrman and Steele, 2008; Fuhrman, 2009; Shade *et al.*, 2010; Steele *et al.*, 2011). The input file used is shown in Supplementary Table 3. Cytoscape depicts data sets as nodes (environmental variables and tribes) connected by lines that denote the character (positive or negative correlation) and the time lag of the relationship.

Results

β -Diversity

Overall, sequencing yielded a total of 120 564 reads whereof 91 187 reads were of high quality. On average 2763 sequences were obtained for each of the 33 samples (range 1652–5102 sequences). Using a 97% similarity cut-off, these sequences were clustered by using a single linkage pre-clustering approach (slp; Huse *et al.*, 2010). A total of 1853 OTUs were obtained whereof 28 OTUs were detected in all 33 samples and 137 were detected in more than 10 samples. More than half of the OTUs in the entire data set (980 OTUs) were singletons. After the quality control of alignments and re-sampling step by using daisychopper, 1612 sequences remained for each sample. These sequences clustered into on average 135 OTUs per sample (range from 94 to 188 OTUs). In total, 1267 OTUs remained in the entire data set after the re-sampling, and detailed seasonal dynamics are provided for 198 OTUs in Supplementary Figure 2 (z-score-standardized). These OTUs were represented by at least 10 reads in the entire re-sampled data set. Taxonomic placements of the sequence reads were also used to obtain the highest taxonomic resolution for these 198 OTUs (Supplementary Figure 2).

To visualize the temporal dynamics of the community (β -diversity), the 1267 OTUs remaining after

re-sampling by daisychopper were used to compute a Bray–Curtis similarity matrix that was subsequently ordinated into two dimensions by using NMDS (Figure 1). Samples were grouped according to season, with samples of adjacent sampling occasions being more similar than samples that were further apart over season. This was confirmed by a Mantel test revealing a strong and significant correlation between community dissimilarity and cyclic (seasonal) temporal distance ($R=0.60$, $P<0.0001$).

Taxonomic placement

Using the silva/rdp framework, only 0.7% of the 91 167 reads could not be resolved to a taxonomic level of phylum. This pool of unresolved sequences was slightly larger with the freshwater database (1.2%), but considering the size difference of the two sequence databases this is not surprising. Using the freshwater taxonomic framework, 38% of the reads could be binned into tribes, 59% into clades and 76% into lineages. For comparison, 70% could be binned into genera and 97% into bacterial orders by using the rdp taxonomic framework. It should be mentioned that the taxonomic levels of the two frameworks are not directly comparable; still genera and clades correspond in many cases (see Supplementary Tables 3 and 4 for comparison).

Both analyses revealed that most reads in the data set were affiliated with the phylum *Actinobacteria* (38% when the rdp classification was used). This phylum contributed the majority of the reads in most samples with a few notable exceptions: (i) most reads in March were classified as *Verrucomicrobia* (more than 30% of the total reads on two occasions); (ii) most reads in May were classified as *Bacteroidetes* (26% of the total reads) and (iii) toward the end of the sampling period (the last three samples in

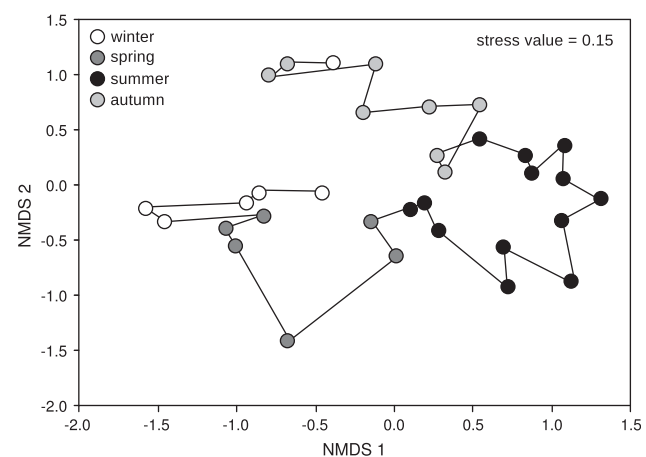


Figure 1 NMDS of a Bray–Curtis resemblance matrix among 33 pelagic samples obtained from Lake Erken throughout a year. This analysis was based on re-sampled abundances of 1267 bacterial OTUs. Samples are grouped by season (see legend) and connected over time by the solid line.

late fall/winter), *Proteobacteria* was the most abundant phylum, contributing from 36% to 44% to the total number of reads (Supplementary Figure 3).

In addition to the members of previously described tribes and clades, a significant fraction (38.3% of the total 16S rRNA sequences) could not be placed within any of these previously defined freshwater groups based on sequence similarity using a bootstrap cut-off of 80%. These sequences do not represent any novel taxonomic groups as 97% of the sequences could be assigned at least to order level when the rdp taxonomic framework was used. Still, there was a single exception for an unclassified bacterial OTU, which could not even be taxonomically placed into a phylum by using the rdp framework. This OTU was most abundant in spring and was represented by approximately 300 reads.

Tribe dynamics

For each individual sample, we extracted the number of reads that binned into the most narrowly defined phylogenetic groups, so called tribes and clades. This tribe/clade abundance matrix was first re-sampled by daisychopper (to re-sample all samples to smallest integer) and standardized by estimating the z-scores for each tribe in each sample. Positive z-scores reflect abundances above the average, whereas a negative z-score indicates abundance below the average. k-means clustering was then used to bin freshwater groups (tribes and clades) according to their seasonal dynamics (Figures 2a–j). This analysis grouped tribes and clades into broad groups of putative ecological coherence based on their temporal distribution patterns. For example, tribes listed in Figures 2a and b showed highest abundances during spring, coinciding with a spring diatom bloom and enhanced snowmelt discharge. The β -proteobacterial tribes Lhab-A4, LD28 and others peaked in winter (Figure 2j). By contrast, the tribes in Figure 2i peaked in early autumn coinciding with high nitrification as suggested by the decrease in ammonia and particulate nitrogen paralleled by an increase in nitrate (Supplementary Figure 4). The latter include the freshwater α -proteobacterial tribe LD12 (the freshwater sibling group of the marine SAR11), which was the most abundant tribe in the combined data set (Table 1). LD12 was detected over the entire annual cycle and represented almost 7% of the total number of bacterial 16S rRNA sequences.

Other abundant tribes were Iluma-A2 and Iluma-A1, recruiting 4.4% and 3.0% of the reads in the data set, respectively (Table 1). Both of these actinobacterial tribes were detected in all samples, with tribe Iluma-A2 making up between 1.7% and 8.6% of the reads in each sample, whereas tribe Iluma-A1 varied from 0.7% to 7.0%. These closely related tribes follow divergent temporal dynamics, with Iluma-A2 being most abundant in between

phytoplankton blooms (2C), whereas reads belonging to tribe Iluma-A1 were more abundant in autumn.

The phylum *Bacteroidetes* represented the third largest phylum in our data set (Supplementary Figure 3). However, only a minor portion of the reads could be resolved to clade and tribe level by using the freshwater database for classification. The RDP taxonomy resolved most of these reads to the family level, with *Chitinophagaceae* (on average 3.5%), *Crymorphaceae* (on average 2.5%) and *Flavobacteriaceae* (on average 2.5%) representing the most abundant *Bacteroidetes* groups. Tribes within the gammaproteobacteria, *Verrucomicrobia*, OP10 and *Fibrobacteres* were detected infrequently but occasionally contributed significantly to the total number of reads (Table 1). For example, reads affiliated with clade verI-B (*Xiphinematobacter*) contributed more than 4.5% to the total reads on three sampling occasions in spring and one in autumn.

Groups within the *Cyanobacteria* showed pronounced variations over the annual cycle, where tribe *Cyanothece* and tribe *Microcystis* made the largest contribution in May (up to 7.2% of the reads) and September (up to 7.0% of the reads), respectively. Tribe *Anabaena* peaked during a chlorophyll-*a* peak in summer (corroborated by microscopy observations; unpublished data), suggesting that these filamentous cyanobacteria represented a major component of this bloom event.

Network analysis

The synchrony of tribes already identified by k-means clustering and the progression of tribes over time was also analyzed in association networks. These networks also enabled the detection of time-lagged interdependencies and provide further resolution to the succession patterns. For example, the k-means clustering and plotting of dynamics revealed that acI tribes had similar dynamics, and LSA permitted the uncovering of the specific time lag between these tribes; for example, the correlations of acI-B1 with acI-A3 and acI-A4 were in both cases significantly delayed (Figure 3a). The successive shift from LD12 to Iluma-C1 and acI-C1 could be resolved by LSA as well as the progression from the *Acetobacteraceae* and acTH1-A1 to *Anabaena* and further to A0904 (Figure 4a). These delayed transitions were not obvious from k-means clustering (Figure 2).

The networks revealed not only the occurrence patterns among tribes, but also associations among members of the bacterioplankton community and between community members (tribes) and environmental variables (Figures 3 and 4). In general, the number of correlations among tribes and those between bacterial tribes and environmental variables were approximately the same despite an unequal amount of total correlations tested among

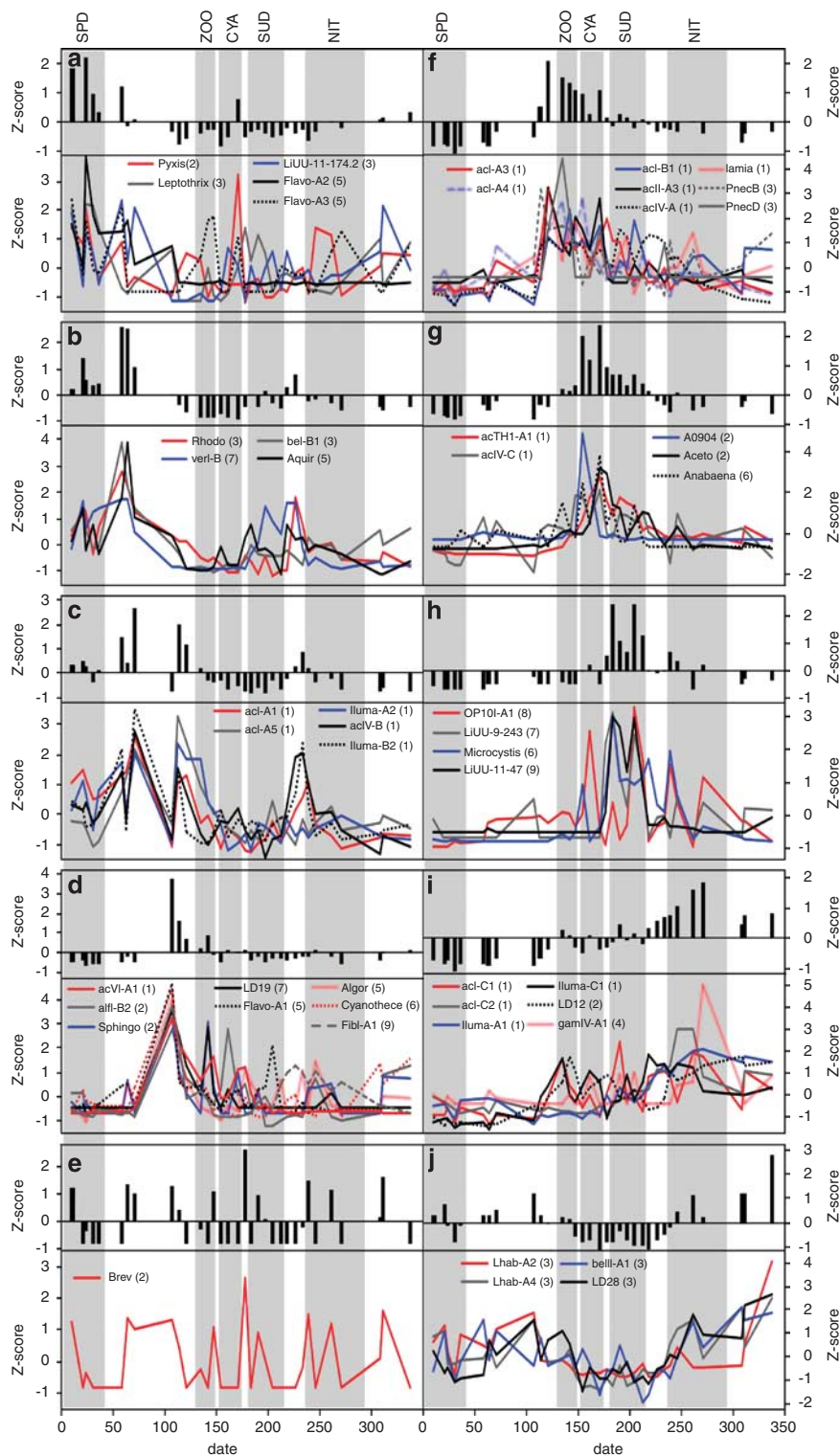


Figure 2 Standardized abundance profiles (z-score) for identified tribes. Tribes with similar abundance profiles are plotted together according to their seasonality determined by k-means clustering. Each panel (a–j) includes two plots where the upper (stack) plot represents the average standardized abundance of the tribes in each cluster, whereas the lower plot represents the individual abundance profiles of each tribe. Panels are sorted according to the seasonal progression (succession). The numbers in parentheses following the tribe name correspond to phylum or sub-phylum names: (1) *Actinobacteria*, (2) *α-proteobacteria*, (3) *β-proteobacteria*, (4) *γ-proteobacteria*, (5) *Bacteroidetes*, (6) *Cyanobacteria*, (7) *Verrucomicrobia*, (8) OP10 and (9) *Fibrobacteres*. Different seasonal events are indicated by shading: (SPD) Spring diatom bloom, (ZOO) zooplankton bloom, (CYA) cyanobacterial bloom, (SUD) summer diatom bloom and (NIT) nitrification.

Table 1 Relative contribution of reads to the 20 most abundant tribes (clades) in the 454 pyrosequencing data set

Phylum	Lineage/clade	Tribe	Average	Min.	Max.
Actinobacteria	acI; acI-A	acI-A3	0.010	0.001	0.036
	acI; acI-A	acI-A4	0.012	0.003	0.029
	acI; acI-B	acI-B1	0.025	0.005	0.056
	acIV; acIV-A	Iluma-A1	0.029	0.007	0.066
	acIV; acIV-A	Iluma-A2	0.043	0.016	0.085
	acIV; acIV-B	Iluma-B2	0.010	0.002	0.035
	acIV; acIV-C	Iluma-C1	0.012	0.002	0.023
	AcTH1; acTH1-A	acTH1-A1	0.011	0.000	0.045
Bacteroidetes	B0		0.021	0.003	0.040
	baII; baII-A	Flavo-A2	0.008	0.000	0.055
Cyanobacteria		Cyanothece	0.012	0.000	0.070
		Microcystis	0.011	0.000	0.069
		LD12	0.063	0.008	0.139
Proteobacteria	alfIV; alfIV-A	Lhab-A4	0.011	0.003	0.027
	beI; beI-A	Rhodo	0.016	0.003	0.042
	beI; beI-B	beI-B1	0.008	0.000	0.037
	beII	PnecB	0.011	0.003	0.033
	beIII; beIII-A	beIII-A1	0.010	0.001	0.017
	beIV; beIV-A	LD28	0.015	0.003	0.035
	verI; verI-B		0.018	0.000	0.048
Verrucomicrobia					

The average, maximum (max.) and minimum (min.) proportions from 33 samples are shown.

these two groups (Table 2). There was also a minor difference between the number of positive and negative correlations. A total of 161 time-lagged correlations among tribes was detected compared with the 83 contemporaneous correlations (Table 2). For individual tribes, on average 14.1 significant correlations (associated nodes in the networks) were observed (see Supplementary Table 5 for LSA results). The actinobacterial tribe acTH1-A1 had the highest degree of association (degree = 33) in our network, with correlation coefficients ranging from 0.41 to 0.65. Other groups with a high degree of linkages were the α -proteobacterial family *Acetobacteraceae* (Figure 4a, degree = 28) and clade verI-B (degree = 28), as well as the cyanobacterial genera (tribe) *Microcystis* (degree = 24) and *Cyanothece* (degree = 24).

Discussion

A clear succession of the bacterioplankton community over the annual cycle was revealed by NMDS ordination (Figure 1) and k-means clustering (Figure 2). Such temporal trajectories over annual cycles have been reported in several studies examining the temporal dynamics of bacterioplankton communities in lakes by using fingerprinting techniques such as terminal restriction fragment length polymorphism, amplified ribosomal intergenic spacer analysis and denaturing gradient gel electrophoresis to separate 16S rRNA genotypes (see for example Yannarell *et al.*, 2003; Kent *et al.*, 2004). More recent studies based on high-frequency, multi-year data sets have revealed that seasonal patterns in bacterioplankton community structure are reoccurring in freshwater systems (Crump and

Hobbie, 2005; Kent *et al.*, 2007; Shade *et al.*, 2007; Crump *et al.*, 2009). It is likely that such reoccurring seasonal patterns or phenologies also occur in Lake Erken, but this could not be tested in this single-year study.

For comparing dynamics in α and β -diversity, the units of diversity are usually based on reproducible criteria, like a clustering algorithm and a 97% cut-off. However, when making inferences about the ecology of bacterial populations, a taxonomic framework with predefined phylogenetic groups (tribes) is better suited for comparisons among studies to build a natural history guide for bacterial tribes (Newton *et al.*, 2011). As indicated in Figures 2c and h, multiple tribes from a single phylum increased simultaneously during certain times of the year and there is hence indications of shared ecological features at this higher taxonomic level (Phillippot *et al.*, 2010). Examples for this are tribes within the acI lineage. The closely related acI-C1 and acI-C2 had highest relative abundance in autumn (Figure 2i), whereas acI-A1 and acI-A5 made the largest contribution to the community in spring (Figure 2c). Tribes acI-A3, acI-A4 and acI-B1 had yet another temporal pattern with highest contributions to the community in summer (Figure 2f). To further explore if ecological coherence is more common among broader taxonomic groups, we calculated the ratio between tribes with and without coherent dynamics within phyla to 0.19 (45 coherent versus 236 non-synchronized). This ratio was higher compared with a similar ratio for tribes from different phyla, that is, 0.075 (74 coherent versus 989 non-synchronized), indicating that tribes from the same phyla appear to share ecological properties more often than tribes from more distantly related OTUs. This also suggests that closely related



b

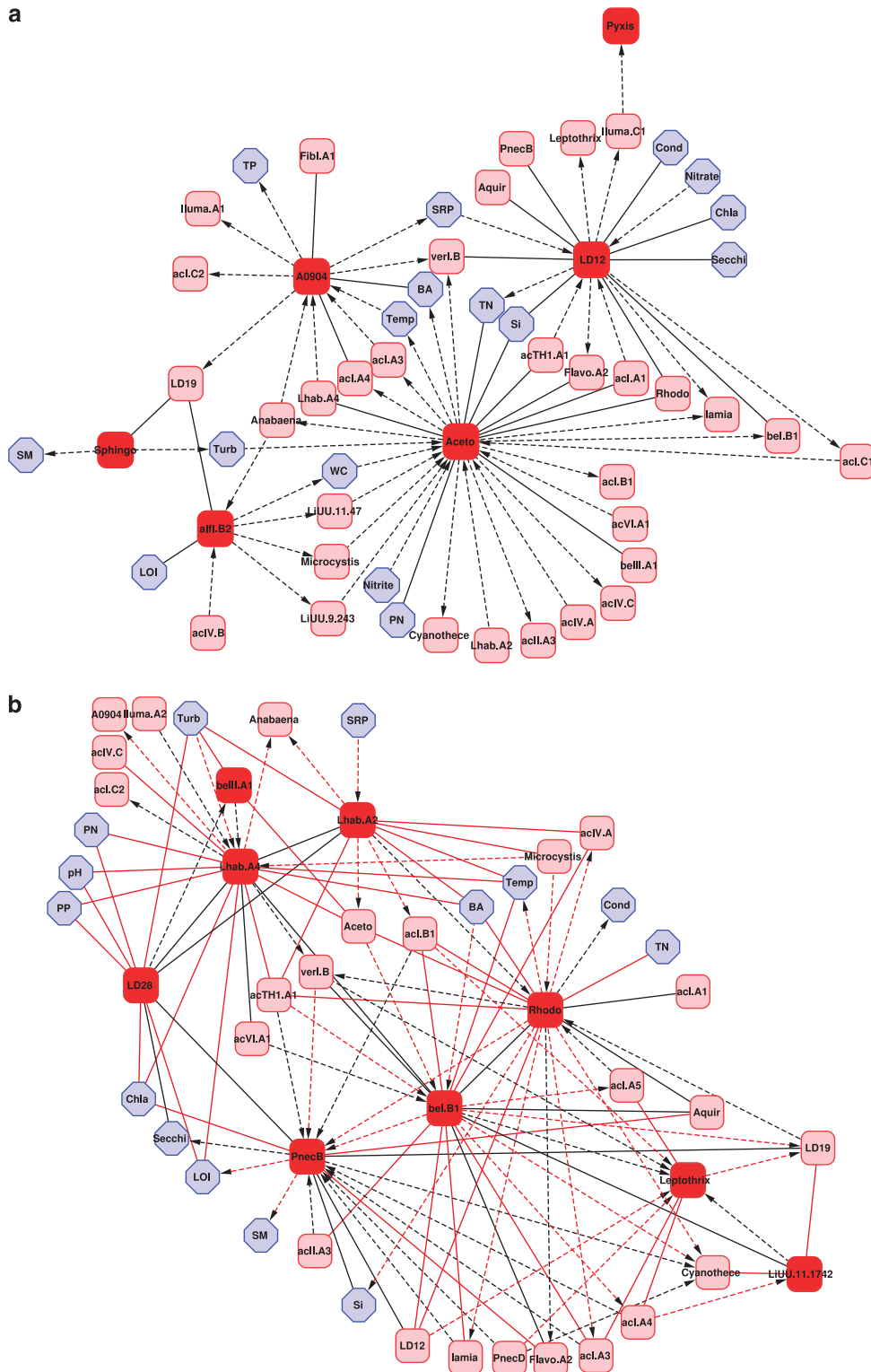


Figure 4 Sub-network organized around α -proteobacterial (a) and β -proteobacterial (b) tribes. Sub-networks were extracted from the entire association network (see Supplementary Figure 5), including only the edges (correlations) between the target groups (α -proteobacterial tribes in panel a and β -proteobacterial in panel b) and their associated groups and environmental variables (edges between non-targeted variables are not shown). The red rectangles represent freshwater bacterial tribes, with abbreviation according to Newton *et al.* (2011), where the strong red colored rectangles represent the target groups. The blue hexagons represent environmental variables. The black lines indicate positive correlations; the red lines indicate negative correlations; the dashed lines indicate a time shift in the correlations (a lag of 1–3) and the arrows indicate the direction of the time shift (the arrow points to the variable lagging behind). The abbreviations for the environmental variables are translated as follows: BA, bacterial abundance; chl *a*, chlorophyll-*a*; Cond, conductivity; LOI, loss on ignition; PN, particulate nitrogen; PP, particulate phosphorus; secchi, secchi depth; Si, silicate concentration; SM, suspended matter; SRP, soluble reactive phosphorus; Temp, temperature; TN, total nitrogen; TP, total phosphorus; Turb, turbidity; WC, water color.

Table 2 Number of correlations as inferred by the LSA

	<i>Tribe × tribe</i>	<i>Tribe × Env</i>	<i>Env × Env</i>	<i>Total</i>
No. of pairwise correlations	2926	1771	253	4950
No. of significant correlations	244 (161)	176 (112)	51 (14)	471 (287)
No. of significant positive correlations	130 (87)	98 (69)	35 (11)	263 (167)
No. of significant negative correlations	114 (74)	78 (43)	16 (3)	208 (120)

Abbreviation: LSA, local similarity analysis.

The total number of pairwise correlations as well as significant correlations among tribes (tribe × tribe), among environmental variables (env × env), and between tribes and environmental variables (tribe × env) are shown. The numbers in parentheses indicate when significant correlations with a time lag had a higher correlation coefficient compared with contemporaneous correlations. Detailed results from the LSA are given in supplementary Table 5.

phylogenetic groups with coherent dynamics could either have high functional redundancy and/or serve very similar ecological roles in the environment. Phylogenetically constrained groups seem to feature a common seasonal distribution pattern, providing some first indications of a coupling between phylogeny and ecology in freshwater bacterioplankton similar to what has been observed previously in brackish bacterioplankton communities (Andersson *et al.*, 2010). Despite coherent temporal patterns among closely related actinobacterial tribes within the acI lineage, we also observed patterns indicative for temporal niche partitioning. This is highlighted in the *Actinobacteria* Iluma-A2 and Iluma-A1, which follow divergent temporal dynamics, Iluma-A2 being most abundant in between phytoplankton blooms (Figure 2c), whereas reads belonging to tribe Iluma-A1 were more abundant in autumn (Figure 2i). Also the three most abundant cyanobacterial groups (*Cyanothece*, *Microcystis* and *Anabaena*) feature divergent seasonal patterns corresponding to previously described patterns of phytoplankton dynamics in Lake Erken (Elliott *et al.*, 2007).

Such synchrony (or temporal coherence) of tribes already identified by k-means clustering and the succession of tribes was often further resolved in the association networks. The visualization of LSA in association networks reveals not only the occurrence patterns among tribes, but also their associations to environmental variables, which provide first indications of environmental conditions that are conducive or inhibitory for particular tribes. Accordingly, these association networks can be used to formulate relevant and data-informed hypotheses about the ecology and specific ecological roles of specific tribes and other taxonomic groups (Fuhrman and Steele, 2008; Fuhrman, 2009). For example, the closely related Iluma-A1 and Iluma-C1 tribes correlate positively to both total phosphorus and soluble reactive phosphorus (Figure 3b), suggesting that these tribes compete successfully when phosphorus levels are elevated as a result of lake mixing or during events of external nutrient inputs. In agreement with this observation from a single system, a recent synoptic study based on the accumulated rRNA genes from lakes suggested a

strong negative correlation between these two actinobacterial tribes and the ratio of dissolved organic carbon to total phosphorus (Newton *et al.*, 2011). Furthermore, high temperatures seem to be conducive for some of the most abundant tribes within the *Actinobacteria* (acI-A3, acI-A4, acI-B1, acIV-A and acIV-C), suggesting that they compete successfully under the warm, stratified period of the yearly cycle (Figures 3a and b). Still, k-means clustering and network analysis did not always provide coherent results with regard to the progression of tribes over time. For example, tribe Iluma-C1, clades acIV-A and acIV-C within the actinobacterial lineage acIV were positively correlated without a time lag (Figure 3b) while being classified into different k-means clusters (Figure 2). Similar to the k-means clustering, the LSA networks suggest contrasting levels of ecological coherence within different phyla. For example, the most abundant α -proteobacteria tribes group separate in the k-means clustering (Figure 2) and also show limited coupling in the LSA (Figure 4a). The lack of either significant positive or negative correlations between these tribes and their linkages to alternative environmental parameters led us to hypothesize that they hold different ecological roles in the environment with limited niche overlap. Conversely, many tribes within the β -proteobacteria (LD28, beIII.A1, Lhab.A2 and A4) are strongly coupled in the LSA, with overall negative correlations to chlorophyll-*a*, particulate nitrogen and phosphorus, turbidity and other drivers characteristic for periods of high phytoplankton biomass and primary production (Figure 4b). These shared features along with some apparent time lags may suggest that these tribes have overlapping niches and while responding to the same environmental drivers, gradually replace each other over time in a successive manner.

Also strong correlations between distantly related tribes are apparent in the LSA. The β -proteobacterial tribe PnecB and the α -proteobacterial tribe LD12 both respond favorably to low chlorophyll-*a* levels (Figure 4). Still, these tribes feature non-synchronized dynamics (Figure 2) and have a diverse array of associations in the network (Figure 4), corroborating findings from a global comparison of distribution patterns where these tribes have contrasting

correlations to both nutrient status and geographic location, for example, latitude (Newton *et al.*, 2011). An interesting feature of the β -proteobacterial LD28 tribe was its high proportion of connections to environmental properties and a lesser degree of associations to other bacterial taxa and none to groups outside of the β -proteobacteria (Figure 4b). This may point to a unique role in the ecosystem with few competitors or functional interdependencies.

Overall, a total of 161 time-lagged correlations among tribes were detected compared with the 83 contemporaneous correlations (Table 2), possibly indicating that time-shifted interdependencies prevail in bacterioplankton communities, and further emphasizing the importance of highly resolved temporal data in disentangling the ecological features of abundant bacterioplankton groups. As samples were collected at weekly to monthly intervals, time lags cannot be assigned to a consistent number of days. Hence, one should be cautious when making ecological interpretations as interdependencies with a time lag can span from a single week to several months.

We also detected numerous tribes with a high number of associations, which may represent groups with particularly strong interdependencies with other taxa and environmental driver variables. Accordingly they can be regarded as 'hubs' in the complex networks that lake bacterioplankton communities form. These 'hubs' may represent 'keystones' as predicted from network theory (Montoya *et al.*, 2006) and as the resident freshwater microbial communities are highly complex, there may be hundreds of such 'keystones'. Food web models suggest that the loss of these most connected taxa could break the networks into many disconnected sub-networks (Dunne *et al.*, 2002). Still, the effects of organism loss also depend on the specific position of the organisms in the network and the specific nature of their interactions with the other organisms. An interesting simulation could be to remove specific groups of organisms from a temporal cycle and to follow the propagating effects through networks. Still, we lack an empirical understanding of how species loss propagates through complex natural microbial communities and if this can be detected by analyzing the dynamic nature of ecological networks.

To conclude, association networks provide a highly compressed and simplified version of the typically complex ecological interactions that shape bacterial communities. Some of the co-occurrences may represent functional guilds of organisms performing similar or complementary functions, whereas negative correlations may reflect direct interactions such as competition or may be the result of differences in resistance to losses by grazing or viral lysis. Positive correlations without delay most likely indicate bacteria with functional interdependencies or groups that thrive under similar environmental conditions, whereas delayed

positive correlations are indicative for succession. Negative correlations on the other hand are indicative for competition or differences in growth constraints among taxa. Delayed negative correlations can be the result of allelopathy (secretion of antimicrobial substances) or of communication through quorum sensing.

We still need to be aware that co-occurrence patterns and observed correlation need to be interpreted carefully as the nature of occurrences are difficult if not impossible to assess purely from shifts in abundances. Still, our results facilitate the formulation of concepts and hypotheses about freshwater bacterial communities that can be experimentally corroborated or rejected in future studies.

Acknowledgements

We thank the staff at the Lake Erken Limnological field station for logistic support, sampling assistance and access to water quality monitoring data. We also thank Anders Andersson for advice and help with an initial analysis of the 454 pyrosequencing data, and three anonymous reviewers for helpful suggestions and comments. The study was funded by the Swedish Research Council (grants to SB). 454 sequencing was partially funded by the Knut and Alice Wallenberg foundation through a grant to the Centre for Metagenomic Sequence Analysis at the Swedish Royal Institute of Technology.

References

- Andersson AF, Riemann L, Bertilsson S. (2010). Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J* **4**: 171–181.
- Bahr M, Hobbie JE, Sogin ML. (1996). Bacterial diversity in an arctic lake: a freshwater SAR11 cluster. *Aquat Microb Ecol* **11**: 271–277.
- Bell RT, Stensdotter U, Pettersson K, Istanovics V, Pierson DC. (1998). Microbial dynamics and phosphorus turnover in Lake Erken. *Adv Limnol* **51**: 1–20.
- Bertilsson S, Eiler A, Nordqvist A, Jorgensen NOG. (2007). Links between bacterial production, amino-acid utilization and community composition in productive lakes. *ISME J* **1**: 532–544.
- Cotner JB, Biddanda BA. (2002). Small players, large role: microbial influence on biogeochemical processes in pelagic aquatic ecosystems. *Ecosystems* **5**: 105–121.
- Crump BC, Hobbie JE. (2005). Synchrony and seasonality in bacterioplankton communities of two temperate rivers. *Limnol Oceanogr* **50**: 1718–1729.
- Crump BC, Peterson BJ, Raymond PA, Amon RMW, Rinehart A, McClelland JW *et al.* (2009). Circumpolar synchrony in big river bacterioplankton. *Proc Natl Acad Sci USA* **106**: 21208–21212.
- Debroas D, Humbert JF, Enault F, Bronner G, Faubladiere M, Cornillot E *et al.* (2009). Metagenomic approach studying the taxonomic and functional diversity of the bacterial community in a mesotrophic lake (Lac du Bourget—France). *Environ Microbiol* **11**: 2412–2424.

- Del Giorgio PA, Bird AF, Prairie YT, Planas D. (1996). The flow cytometric determination of bacterial abundance in lake plankton with the green acid stain SYTO 13. *Limnol Oceanogr* **41**: 783–789.
- Dunne JA, Williams RJ, Martinez ND. (2002). Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecol Lett* **5**: 558–567.
- Eiler A, Bertilsson S. (2004). Composition of freshwater bacterial communities associated with cyanobacterial blooms in four Swedish lakes. *Environ Microbiol* **6**: 1228–1243.
- Eiler A, Bertilsson S. (2007). Flavobacteria blooms in four eutrophic lakes: linking population dynamics of freshwater bacterioplankton to resource availability. *Appl Environ Microbiol* **73**: 3511–3518.
- Elliott JA, Persson I, Thackeray SJ, Blenckner T. (2007). Phytoplankton modelling of Lake Erken, Sweden by linking models PROBE and PROTECH. *Ecol Model* **202**: 421–426.
- Fuhrman JA. (2009). Microbial community structure and its functional implications. *Nature* **459**: 193–199.
- Fuhrman JA, Steele JA. (2008). Community structure of marine bacterioplankton: patterns, networks, and relationships to function. *Aquat Microb Ecol* **53**: 69–81.
- Gilbert JA, Field D, Swift P, Newbold L, Oliver A, Smyth T *et al.* (2009). The seasonal structure of microbial communities in the Western English Channel. *Environ Microbiol* **11**: 3132–3139.
- Hahn MW. (2009). Description of seven candidate species affiliated with the phylum *Actinobacteria*, representing planktonic freshwater bacteria. *Int J Syst Evol Microbiol* **59**: 112–117.
- Hahn MW, Kasalicky V, Jezbera J, Brandt U, Jezberova J, Simek K. (2010a). *Limnolobus curvus* gen. nov., sp. nov., a planktonic bacterium isolated from freshwater lakes. *Int J Syst Evol Microbiol* **60**: 1358–1365.
- Hahn MW, Lang E, Brandt U, Lunsdorf H, Wu QL, Stackebrandt E. (2010b). *Polynucleobacter cosmopolitanus* sp. nov., free-living planktonic bacteria inhabiting freshwater lakes and rivers. *Int J Syst Evol Microbiol* **60**: 166–173.
- Hartigan JA, Wong MA. (1979). A K-means clustering algorithm. *Appl Stat* **28**: 100–108.
- Herlemann DPR, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J*; e-pub ahead of print 7 April 2011; doi:10.1038/ismej.2011.41.
- Huse SM, Welch DM, Morrison HG, Sogin ML. (2010). Ironing out the wrinkles in the rare biosphere through improved clustering. *Environ Microbiol* **12**: 1889–1898.
- Hutalle-Schmelzer KML, Grossart HP. (2009). Changes in the bacterioplankton community of oligotrophic Lake Stechlin (northeastern Germany) after humic matter addition. *Aquat Microb Ecol* **55**: 155–167.
- Jezbera J, Sharma AK, Brandt U, Doolittle WF, Hahn MW. (2009). ‘*Candidatus Planktophila limnetica*’, an actinobacterium representing one of the most numerically important taxa in freshwater bacterioplankton. *Int J Syst Evol Microbiol* **59**: 2864–2869.
- Jones SE, Newton RJ, McMahon KD. (2009). Evidence for structuring of bacterial community composition by organic carbon source in temperate lakes. *Environ Microbiol* **11**: 2463–2472.
- Judd KE, Crump BC, Kling GW. (2006). Variation in dissolved organic matter controls bacterial production and community composition. *Ecology* **87**: 2068–2079.
- Kent AD, Jones SE, Yannarell AC, Graham JM, Lauster GH, Kratz TK *et al.* (2004). Annual patterns in bacterioplankton community variability in a humic lake. *Microb Ecol* **48**: 550–560.
- Kent AD, Yannarell AC, Rusak JA, Triplett EW, McMahon KD. (2007). Synchrony in aquatic microbial community dynamics. *ISME J* **1**: 38–47.
- Legendre P, Legendre L. (1998). *Numerical Ecology*, 2nd English edn. Elsevier Science BV: Amsterdam, The Netherlands.
- Lindstrom ES. (2000). Bacterioplankton community composition in five lakes differing in trophic status and humic content. *Microb Ecol* **40**: 104–113.
- Lindstrom ES, Kamst-Van Agterveld MP, Zwart G. (2005). Distribution of typical freshwater bacterial groups is associated with pH, temperature, and lake water retention time. *Appl Environ Microbiol* **71**: 8201–8206.
- Little AEF, Robinson CJ, Peterson SB, Raffa KE, Handelsman J. (2008). Rules of engagement: interspecies interactions that regulate microbial communities. *Annu Rev Microbiol* **62**: 375–401.
- Montoya JM, Pimm SL, Solé RV. (2006). Ecological networks and their fragility. *Nature* **442**: 259–264.
- Nelson CE. (2009). Phenology of high-elevation pelagic bacteria: the roles of meteorologic variability, catchment inputs and thermal stratification in structuring communities. *ISME J* **3**: 13–30.
- Newton RJ, Kent AD, Triplett EW, McMahon KD. (2006). Microbial community dynamics in a humic lake: differential persistence of common freshwater phylotypes. *Environ Microbiol* **8**: 956–970.
- Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. (2011). A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol Rev* **75**: 14–49.
- Pernthaler J. (2005). Predation on prokaryotes in the water column and its ecological implications. *Nat Rev Microbiol* **3**: 537–546.
- Pernthaler J, Glockner FO, Unterholzner S, Alfreider A, Psenner R, Amann R. (1998). Seasonal community and population dynamics of pelagic bacteria and archaea in a high mountain lake. *Appl Environ Microbiol* **64**: 4299–4306.
- Pernthaler J, Zollner E, Warnecke F, Jurgens K. (2004). Bloom of filamentous bacteria in a mesotrophic lake: identity and potential controlling mechanism. *Appl Environ Microbiol* **70**: 6272–6281.
- Phillippot L, Andersson SGE, Battin TJ, Prosser JJ, Schimel JP, Whitman WB *et al.* (2010). The ecological coherence of high bacterial taxonomic ranks. *Nat Rev Microbiol* **8**: 523–529.
- Ruan QS, Dutta D, Schwalbach MS, Steele JA, Fuhrman JA, Sun F. (2006). Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics* **22**: 2532–2538.
- Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S *et al.* (2007). The sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical pacific. *PLoS Biol* **5**: 398–431.
- Salcher MM, Pernthaler J, Posch T. (2010). Spatiotemporal distribution and activity patterns of bacteria from three phylogenetic groups in an oligomesotrophic lake. *Limnol Oceanogr* **55**: 846–856.

- Salcher MM, Pernthaler J, Zeder M, Psenner R, Posch T. (2008). Spatio temporal niche separation of planktonic betaproteobacteria in an oligo-mesotrophic lake. *Environ Microbiol* **10**: 2074–2086.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.
- Shade AL, Chiu CY, McMahon KD. (2010). Differential bacterial dynamics promote emergent community robustness to lake mixing: an epilimnion to hypolimnion transplant experiment. *Environ Microbiol* **12**: 455–466.
- Shade AL, Kent AD, Jones SE, Newton RJ, Triplett EW, McMahon KD. (2007). Inter-annual dynamics and phenology of bacterial communities in a eutrophic lake. *Limnol Oceanogr* **52**: 487–494.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D *et al.* (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**: 2498–2503.
- Simek K, Kojecka P, Nedoma J, Hartman P, Vrba J, Dolan JR. (1999). Shifts in bacterial community composition associated with different microzooplankton size fractions in a eutrophic reservoir. *Limnol Oceanogr* **44**: 1634–1644.
- Steele JA, Countway PD, Xia L, Vigil PD, Beman JM, Kim DY *et al.* (2011). Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *ISME J*; e-pub ahead of print 24 March 2011; doi:10.1038/ismej.2011.24.
- Strom SL. (2008). Microbial ecology of ocean biogeochemistry: a community perspective. *Science* **320**: 1043.
- Van der Gucht K, Sabbe K, De Meester L, Vloemans N, Zwart G, Gillis M *et al.* (2001). Contrasting bacterioplankton community composition and seasonal dynamics in two neighbouring hypertrophic freshwater lakes. *Environ Microbiol* **3**: 680–690.
- Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.
- Yannarell AC, Kent AD, Lauster GH, Kratz TK, Triplett EW. (2003). Temporal patterns in bacterial communities in three temperate lakes of different trophic status. *Microb Ecol* **46**: 391–405.
- Yannarell AC, Triplett EW. (2005). Geographic and environmental sources of variation in lake bacterial community composition. *Appl Environ Microbiol* **71**: 227–239.
- Zwart G, Crump BC, Agterveld MPKV, Hagen F, Han SK. (2002). Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquat Microb Ecol* **28**: 141–155.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)