

Importance of space and the local environment for linking local and regional abundances of microbes

Örjan Östman^{1,*}, Stina Drakare², Emma S. Kritzberg³, Silke Langenheder¹,
Jürg B. Logue¹, Eva S. Lindström¹

¹Department of Ecology and Genetics, Uppsala University, Norbyvägen 18D, 75236 Uppsala, Sweden

²Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences (SLU), PO Box 7050, 75007 Uppsala, Sweden

³Limnology, Department of Biology, Ecology Building, Lund University, 22362 Lund, Sweden

ABSTRACT: It is frequently observed that the local relative abundances of aquatic microbial taxa are correlated with their average relative abundance at the regional scale, which results in the composition of different communities being more similar than expected by chance or invariant. The degree to which communities within a region match the regional average community is variable and likely depends on several different mechanisms that control the process of microbial community assembly. Here, we show that environmental variables were associated with the community specific degree of regional invariance in 9 of 10 datasets of microbial communities in aquatic systems, being the main set of variables explaining differences in regional invariance in 5 of them. This indicates that variation in local environmental conditions across a region reduces the degree of regional invariance amongst communities. Spatial distances between communities were not related to the degrees of regional invariance, but in 7 of the datasets, regional invariance differed among different parts of the regions, particularly for phytoplankton communities. This suggests an influence of spatial or historical processes on the community specific degree of regional invariance. We conclude that both local environmental conditions and spatial/historical processes cause between-site differences in the degree of invariance between local and regional abundances in aquatic microbial metacommunities. We argue that studies of regional invariance can be an important complement to other statistical methods due to its propensity to detect variation in stochastic processes along gradients.

KEY WORDS: Metacommunities · Regional abundances · Invariance · Species sorting · Composition · Stochasticity

— Resale or republication not permitted without written consent of the publisher —

INTRODUCTION

The importance of local environmental species-sorting processes in relation to regional processes (e.g. dispersal, history, and large scale environmental properties) for community assembly is a complex and partly unresolved issue in microbial ecology (reviewed by Martiny et al. 2006, Logue & Lindström 2008, Lindström & Langenheder 2012). In general, local environmental conditions tend to explain more between-site variation in microbial assemblages (β -

diversity) than regional factors in statistical models (Beisner et al. 2006, Van der Gucht et al. 2007, Verleyen et al. 2009, Lindström et al. 2010). Yet, regional processes may be important as well, as some studies show that microbial communities display patterns related to spatial differences (Crump et al. 2007, Soininen et al. 2007, 2011, Vyverman et al. 2007) or landscape scale environmental properties (e.g. average phosphorus levels in a region; Ptacnik et al. 2008, 2010) and that variation in dispersal rates may affect community composition (Lindström & Östman 2011).

*Email: orjan.ostman@ebc.uu.se

Many studies aim to identify the extent to which local and regional factors explain variation in community composition among sites, i.e. β -diversity. Most of these studies rely on statistical analyses (e.g. RDA, PCoA, or Mantel-tests) that are designed to reveal consistent (structural) changes in community composition along local environmental gradients or in relation to regional factors, for example, if nearby communities are more similar. However, these methods may miss stochastic processes that can be important for community assembly and β -diversity, which instead cause noise around correlations along gradients. For example, Chase (2010) showed the importance of stochastic community assembly, and that priority effects, giving rise to multiple equilibria, were more intense at higher nutrient levels, leading to an overall increase in β -diversity. Moreover, in accordance with long-term neutral local dynamics in combination with immigration from a common source pool, a ubiquitous pattern is that locally abundant taxa are more frequently detected at the regional scale (Sloan et al. 2006, Drakare & Liess 2010, Östman et al. 2010).

The phenomenon that abundant taxa also occur frequently across regions results in microbial communities that are more similar to one another than is expected by chance, i.e. communities are 'invariant' (sensu Östman et al. 2010). The average relative abundances of taxa at the regional scale (regional relative abundance) therefore, to a smaller or larger degree, explain variation in local abundances, which was termed 'regional invariance' by Östman et al. (2010). When this 'regional invariance' is high, the average relative abundance of taxa at the regional scale explains large parts of the variation in local abundances. Regional invariance among microbial communities can arise because dispersal homogenises communities but can also arise because of species-sorting processes if local environmental conditions are similar amongst communities or be due to dominance of habitat generalists at the regional level (Östman et al. 2010). Previous work showed that the average degree of regional invariance in a region is negatively related to the degree of environmental heterogeneity in that region. In other words, local relative abundances of taxa in regions with high environmental heterogeneity were on average less predictable by their regional relative abundances (Östman et al. 2010).

The degree of regional invariance may also differ substantially among communities within regions (see Table S1 in the supplement at www.int-res.com/articles/suppl/a067p035_supp.pdf). That is, within

the same region, local relative abundances of taxa can be explained by their respective regional relative abundances in some communities, whilst in other communities, this is not the case (see Table S1 in the supplement). Currently, we lack knowledge about the mechanism(s) determining these differences in the degree of regional invariance within regions. Whilst Östman et al. (2010) addressed differences in regional invariance across different regions, we here study which mechanisms may be related to the degree of regional invariance within regions.

There are several processes that may cause these within-region differences in the degree of regional invariance. First, dispersal rates between communities or the history of communities may differ depending on the location or geographic distances between communities. For example, nearby communities or communities connected by surface water could be more invariant than more distant or isolated communities. Second, the degree of regional invariance could be influenced by the local environmental conditions. Habitats that have environmental conditions typical for the region might be highly invariant, whereas habitats that deviate in their environmental conditions will be less invariant. Alternatively, there can be differences in the importance of stochastic processes (neutral dynamics or priority effects) along environmental gradients such that (deterministic) species-sorting processes dominate at one end of the gradient (e.g. only a few taxa tolerating some environments, like low pH, making communities more invariant) and random stochastic processes at the other end (neutral dynamics or priority effects and alternative stable equilibria under more benign or productive conditions), making them less invariant.

Here, we use the same 10 data sets as Östman et al. (2010) to test the following hypotheses: (1) Local abundances will be more invariant in nearby communities than in a random subsample of communities in a region. (2) The degree of regional invariance is related to the spatial location of communities or surface water connections between communities. (3) The degree of regional invariance changes along environmental gradients. (4) Regional invariance is related to changes in community composition (β -diversity) along local environmental gradients or spatial location of communities. The last hypothesis is tested by comparing the variables related to regional invariance with the variables explaining variation in community composition from redundancy analyses (RDA). If these variables are the same, this would suggest that dif-

ferences in regional invariance among communities are related to the same mechanism causing structural changes in community composition, e.g. species sorting to local environment or local dispersal. In contrast, if they differ, this would suggest that differences in regional invariance are mainly linked to variation in other community processes, e.g. variation in importance of neutral dynamics or priority effects along gradients.

MATERIALS AND METHODS

We used 5 data sets on local abundances for both bacterioplankton and phytoplankton. All data sets originated from freshwater/brackish habitats in southern and central Sweden (Fig. 1). The number of communities in each data set (n) varied between 9 and 40, and the spatial scale ranged from hundreds of meters to hundreds of kilometres. Three data sets were derived from bacterioplankton communities from lakes in the provinces of Jämtland ($n = 9$, scale ca. 30 km), Uppland ($n = 13$, scale ca. 65 km), and Småland ($n = 40$, scale ca. 130 km), one data set stemmed from stream bacterioplankton communities in Småland ($n = 18$, scale ca. 120 km), and one originated from rock pool bacterioplankton communities from the coast of Uppland ($n = 35$, scale ca. 0.5 km). The 5 data sets of lake phytoplankton communities were from the provinces of Jämtland ($n = 9$, scale ca. 30 km), Uppland ($n = 13$, scale ca. 65 km), Götaland (the province of Småland and parts of adjacent provinces; $n = 17$, scale ca. 200 km), lakes in Mälardalen ($n = 19$, scale ca. 200 km), and lakes along the catchment of the Kolbäckån river ($n = 19$, scale ca. 90 km). All communities within data sets were sampled within a couple of days, except for phytoplankton in Götaland (24 d between first and last sample) and in Mälardalen (14 d). A local community was defined as 1 lake, stream, or rock pool in a data set. Regions were delineated as the lakes, streams, or rock pools within a data set. Because of practical reasons (it was impossible to combine different terminal-restriction fragment length polymorphism [t-RFLP] data sets) and differences in sampling occasions, data sets were not combined. The lakes in the provinces of Jämtland and Uppland and along Kolbäckån were mainly selected because of their location along water systems. Rock pools, lakes, and streams in the province of Småland were chosen more or less randomly. Lakes in Mälardalen and Götaland are long-term or reference lakes in

monitoring programs and selected to cover a broad range of environmental conditions. The compilation of data sets is described in more detail by Östman et al. (2010) and references therein. In short, bacterioplankton community composition was assessed as relative abundances from peak heights of t-RFLP data, whereas morphological identification was employed for calculating relative biomass of phytoplankton.

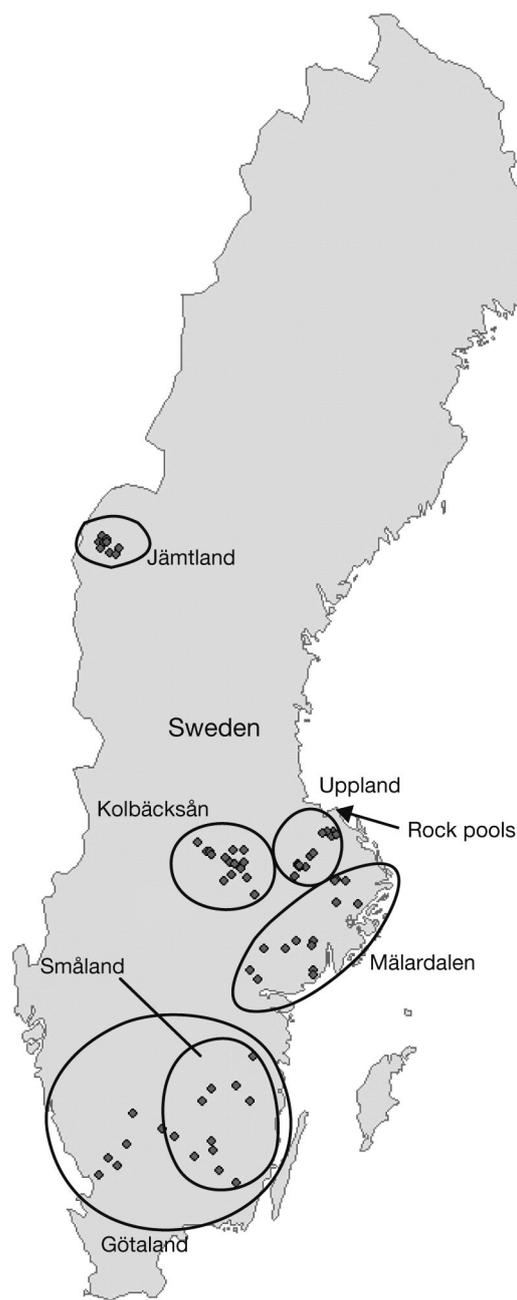


Fig. 1. Sampling communities and regions in Sweden. Note that Götaland (phytoplankton) and the 2 bacterial data sets from Småland are overlapping

Calculation of invariance

For each community, we calculated 2 community specific estimators of invariance, or similarity, between local and regional abundances. These estimators are LRsite and BCsite.

LRsite is the co-variation between local relative abundances and average relative abundances in all other communities in the region of all taxa detected in the region. That is, LRsite describes how much of the variation in local relative abundances can be explained by the average regional relative abundance. The regional relative abundances were not weighed with total abundance or size of communities. LRsite was calculated as the r^2 value from a Pearson correlation between the local relative abundances of taxa at Site x and the average regional relative abundances of taxa from all other sites in that data set excluding Site x . All taxa occurring in the region, even if not detected at Site x , were included in the calculation of the r^2 value. This is the identical measure to 'regional invariance' averaged over all communities in a region by Östman et al. (2010). LRsite is, thus, a measurement of how invariant local relative abundances of taxa in a community are in comparison to the regional relative abundances of the same taxa. If the relative abundance of each taxon at Site x exactly matches its average regional relative abundance, LRsite = 1, whereas LRsite = 0 if there is a random association (no invariance) (see Fig. S1a in the supplement at www.int-res.com/articles/suppl/a067p035_supp.pdf). If taxa in all communities have similar relative abundances, LRsite would be close to 1, and this value would be close to 0 if no association exists. In contrast, if a few communities have different relative taxa abundances than most other communities, most of the communities would have a high LRsite, and a few communities would have a low LRsite. If r -values are negative, this indicates local abundances differing from regional abundances more than expected by random chance. But LRsite is sensitive to the number of taxa included. For example, inclusion of many locally and regionally rare taxa will increase LRsite. Also, if relative abundances are very unevenly distributed among taxa (i.e. a few very dominant taxa), LRsite becomes highly dependent on the co-variation between the local and regional abundance of the most abundant taxa (Fig. S1b in the supplement).

Therefore, we also calculated BCsite, the Bray-Curtis similarity between the local relative abundance at Site x and the regional relative abundance of taxa from all other sites in that data set excluding Site x , calculated as 1 minus the Bray-Curtis dissimi-

larity index. BCsite ranges between 0 and 1, with 0 indicating that communities share no taxa with other communities in the region, whereas 1 means that they are identical. Bray-Curtis similarities are commonly used in spatial ecology, and BCsite is, compared to LRsite, rather insensitive to the number of taxa included and unaffected by a few deviating observations. However, its disadvantage compared to LRsite is that BCsite cannot quantify the deviation expected from chance. A script executable in R for calculating LRsite and BCsite is available in the supplement at www.int-res.com/articles/suppl/a067p035_supp.pdf.

Statistical procedure

To study whether LRsite and BCsite differed depending on the extent of the spatial scale used for calculating mean relative regional abundances, LRsite and BCsite were calculated for each community using relative abundances only from communities within a certain geographic distance between communities (smaller than the entire region). We chose distances in such a manner that on average approximately 20, 40, 60, and 80% of the communities within a region were included in the calculations. However, this would in some cases have resulted in rather small differences in distance between 2 different spatial scales. In such cases, we used intervals between spatial distances, increasing the radius by ~50%. Hence, the actual division into different distance intervals was somewhat arbitrary. For rock pools that occurred in 5 distinct clusters along the shoreline (Langenheder & Ragnarsson 2007), we applied the number of neighbouring rock pool clusters that was included for the calculation of the regional relative abundance instead of using geographic distances. That is, first, the rock pools within the same cluster were chosen, then rock pools within the same cluster and the nearest cluster, and so on until rock pools from all 5 clusters were included. To assess if the observed LRsite and BCsite differed from a random subset of communities, we performed non-parametric permutation tests comparing the observed values with randomised bootstrapped values using Matlab R2011a (MathWorks 2011). This procedure is described in detail in the supplement (www.int-res.com/articles/suppl/a067p035_supp.pdf).

To study whether LRsite and BCsite changed along spatial, hydrological, or environmental gradients within a region, we did multiple linear regressions in SAS 9.2 (SAS Institute 2008). Either LRsite

or BCsite was used as the response variable (both were arcsine square-root transformed prior to the analyses). Spatial variables (S), hydrological connectivity variables (H), and local environmental variables (E) were predictor variables. By entering both local environmental variables and spatial/hydrological variables as explanatory variables in a multiple regression, it is possible to separate the variation that can be related to environmental factors and spatial factors, with the latter being interpreted as spatial (regional effects) differences not caused by local environmental effects.

Site specific (local) environmental variables included in the analysis were community size (km² of water surface for lakes, km² of the drainage area for streams, m³ for rock pools), pH, total phosphorus (TP), total nitrogen (TN), the concentration of organic carbon (C_{org}, measured as dissolved organic carbon [DOC] in rock pools and lakes and streams in Småland and as total organic carbon [TOC] in the other data sets), and finally absorbance (Abs, measured between 420 and 436 nm in the different data sets; in the Uppland and Jämtland data sets, 250 nm was further included). Chlorophyll *a* (chl *a*) and salinity were additionally used in the rock pool data set. We selected the final model by forward selection, that is, the variable having the lowest p-value was added to the model until no variable could be entered with $p < 0.05$.

Spatial variables were the eigenvectors associated with positive eigenvalues generated from the (Euclidean) distance matrix between sites in a data set using the PCNM-algorithm from Griffith & Peres-Neto (2006). Hydrological connectivity (of surface water) variables (H) were site scores from principal components of a site-flow matrix (Blanchet et al. 2008a), which we calculated for the Jämtland, Kolbäcksån, and Uppland data sets, where at least some lakes were physically connected by streams or rivers. In the site-flow matrix, the presence of a hydrological connection of surface water to an upstream community is indicated by a '1', whereas the lack of a connection is indicated by a '0'.

We used published results of RDA on the bacterial data from rock pools (Langenheder & Ragnarsson 2007) and lakes and streams from Småland (Lindström et al. 2010) to study variation explained by spatial, hydrological, and local environmental variables on spatial turnover of taxa (β -diversity). For the other data sets, we performed RDA between the relative abundance of taxa and the spatial, hydrological, and environmental variables already described, following the procedure suggested by Blanchet et al.

(2008b) and using the package 'Packfor' in R 2.10.0 (R-project 2009). Spatial, hydrological, and environmental variables for all communities are available in Table S1 in the supplement.

RESULTS

There was no support for our first hypothesis that local abundances will be more invariant to communities close by than to a random subsample of communities in a region. LRsite and BCsite generally did not differ from what could be expected from a random subsample of the region at any of the spatial distances used for calculating regional abundances (Figs. 2 & 3). The only exception to this pattern was that the observed LRsite among streams in Småland was higher compared to random values at spatial scales smaller than the whole region (Fig. 2).

Although altering the size of the region did not seem to change LRsite or BCsite, we found support for the hypothesis that the degree of regional invariance is related to the spatial location of communities or surface water connections among communities. In 4 of the 5 phytoplankton data sets and in the bacterioplankton data set from Jämtland, LRsite showed a spatial structure (S in Table 1). Hence, in parts of each region, local relative abundances were more invariant than in other parts of the region. In 4 of the data sets, communities in the western parts showed a higher degree of regional invariance compared to more easterly situated communities (bacteria, Jämtland: $r = -0.71$; phytoplankton, Jämtland: $r = -0.58$, phytoplankton, Mälardalen: $r = -0.54$; phytoplankton, Götaland: $r = -0.41$). However, in the Jämtland bacterial data set, this association depended largely on the 2 easternmost communities showing low degrees of regional invariance and, thus, must be considered as uncertain ($r = 0.04$ if these were removed). Spatial structures in BCsite were only evident in the 2 data sets in Jämtland (Table 1). LRsite and BCsite were associated with hydrological connectivity variables in the Uppland lakes for both bacterio- and phytoplankton (Table 1). BCsite for phytoplankton also showed an association with the hydrological connectivity among lakes along the Kolbäcksån river (Table 1). However, these associations were strongly dependent on the relative local abundances in 2 lakes (which were hydrologically isolated from the others) that were considerably less regionally invariant. No associations with hydrological variables were evident for bacterio- and phytoplankton in the Jämtland data sets (Table 1).

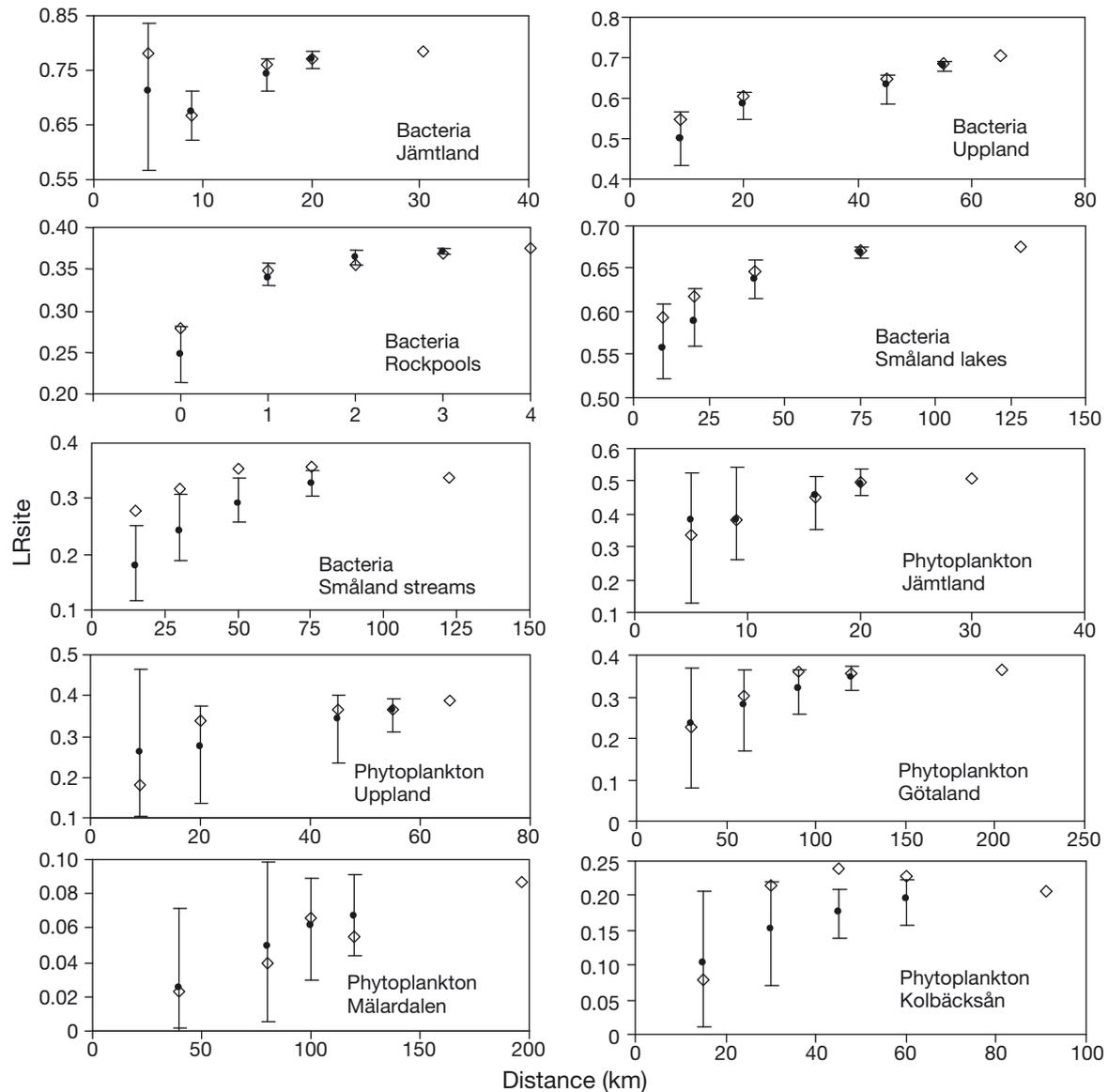


Fig. 2. Observed average LRsite (open diamonds, the average degree to which regional abundances can explain variation in local abundances) calculated at different spatial scales generally do not differ from random subsamples of communities from the whole region in the 10 data sets. Filled circles and error bars are the overall averages and 95% confidence intervals from 1000 randomisations, calculated from an average LRsite of the communities for each randomisation. At the largest scale, all communities in the data set were included. Note that the increasing trends in both observed and expected LRsite with increasing spatial scale is an artefact of more taxa being included in the regional abundance at the larger spatial scales. For rock-pool data, we did not use Euclidean distances but the number of neighbouring rock pool clusters: '0' indicates that only rock pools within a cluster were used, '1' indicates inclusion of nearest neighbouring cluster, and so on for 5 clusters (see map in Langenheder & Ragnarsson 2007)

There was also support for the hypothesis that the degree of regional invariance changes along environmental gradients. In all data sets except one, LRsite and BCsite were associated with local environmental conditions. This indicates that variation in local environmental conditions resulted in deviating degrees of regional invariance among communi-

ties. For bacterioplankton, LRsite and BCsite were positively associated with TP or TN in 3 of the 5 data sets but negatively correlated to these variables in the rock pools (Table 1, Fig. 4a,b). In the 2 Småland bacterial data sets with more acidic conditions, LRsite and BCsite showed opposite associations with pH (Fig. 4c,d): positive among lake communi-

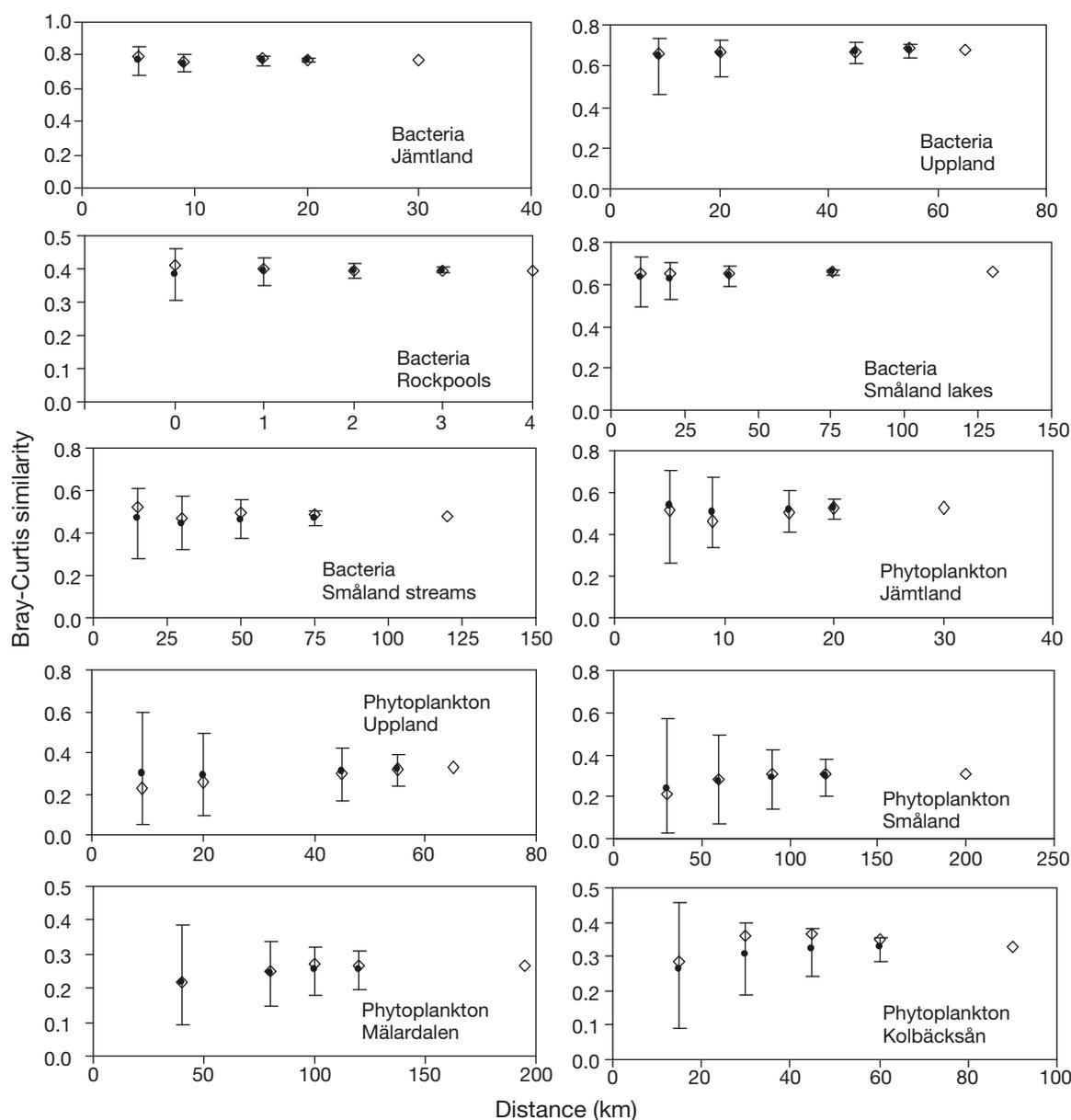


Fig. 3. Observed and randomised average BCsite (the average similarity between local and regional abundances) at different spatial scales used for calculating regional abundances in the 10 data sets. See Fig. 2 for explanations of symbols and treatment of rock pool data

ties and negative among stream communities. For phytoplankton, LRsite and BCsite were most frequently associated with Abs or C_{org} , for which both positive and negative associations were observed (Table 1, Fig. 4e,f), and in a few data sets also with TP and TN.

The results from the RDA are presented in Table 2. For bacterial data sets, LRsite and BCsite were correlated to the same variables that explained most of the β -diversity in the RDA. The exception was the rock pool data set, where the RDA showed that commu-

nity composition was mainly related to space and salinity, whereas LRsite and BCsite were related to chl *a* or TP. Also, in the phytoplankton data sets, there was good coherence between significant predictor variables for β -diversity and variation in LRsite and BCsite, the exception being the Mälardalen data set. Here, spatial variation in community composition was mainly related to Abs and pH, whereas LRsite and BCsite were more related to productivity (TP or TN) or space (an east-westward relationship was found for LRsite).

Table 1. Multiple linear regression analyses, using forward selection with LRsite (variation in local abundances predicted by regional abundances) or BCsite (Bray-Curtis similarity between local and regional abundances) as response variables. Different environmental, spatial (S) and hydrological connectivity (H) variables were used as predictors. 'Abs' and 'C_{org}' (absorbance and concentration of organic carbon) have been combined into 1 variable here because of strong colinearity in all data sets, as both are mainly estimates of humic substances in the water. The sign within the bracket indicates the direction of the relationship. ‡ indicates that a variable was correlated ($p < 0.05$) to LRsite or BCsite but excluded from the final model because of colinearity with other variables. † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ in the final model. TP: total phosphorus; TN: total nitrogen. Empty cells represent variables removed from the final model

Organism	Data set	Variable	TP	TN	Chl a	pH	Abs/C _{org}	S	H
Bacteria	Jämtland	LRsite	$F_{1,6} = 6.7^*(+)$					$F_{1,6} = 6.8^*$	
		BCsite	$F_{1,6} = 5.1†(+)$					$F_{1,6} = 6.7^*$	
	Uppland	LRsite					$F_{1,10} = 6.3^*(-)$		$F_{1,10} = 37^{***}$
		BCsite					$F_{1,10} = 3.6†(-)$		$F_{1,10} = 14^{**}$
	Rock pools	LRsite	‡	‡	$F_{1,33} = 8.0^{**}(-)$				
		BCsite		$F_{1,33} = 4.6^*(+)$					
Småland Lakes	LRsite	‡	$F_{1,37} = 8.8^*(+)$		$F_{1,37} = 25^{***}(+)$		‡		
	BCsite	‡	$F_{1,37} = 9.3^*(+)$		$F_{1,37} = 80^{***}(+)$		‡		
Småland Streams	LRsite	‡	$F_{1,15} = 14^{**}(+)$			‡	$F_{1,15} = 17^{***}(+)$	‡	
	BCsite	‡	‡		$F_{1,16} = 11^{**}(-)$		‡		
Phyto-plankton	Jämtland	LRsite					‡	$F_{1,7} = 11^*$	
		BCsite					$F_{1,6} = 3.9†(-)$	$F_{1,6} = 10^*$	
	Uppland	LRsite					$F_{1,10} = 29^{***}(+)$		$F_{1,10} = 7.8^*$
		BCsite					$F_{1,10} = 4.9^*(+)$		‡
	Göteborg	LRsite	$F_{1,14} = 28^{***}(+)$	‡	‡		‡	$F_{1,14} = 6.1^*$	
		BCsite		$F_{1,14} = 4.2(-)†$			$F_{1,14} = 6.3(+)$		
Mälardalen	LRsite	‡	‡				$F_{1,17} = 6.2^*$		
	BCsite		$F_{1,17} = 4.3(-)^*$						
Kolbäckån	LRsite							$F_{1,9} = 5.0^*$	
	BCsite						$F_{1,8} = 3.7†(-)$		$F_{1,8} = 5.8^*$

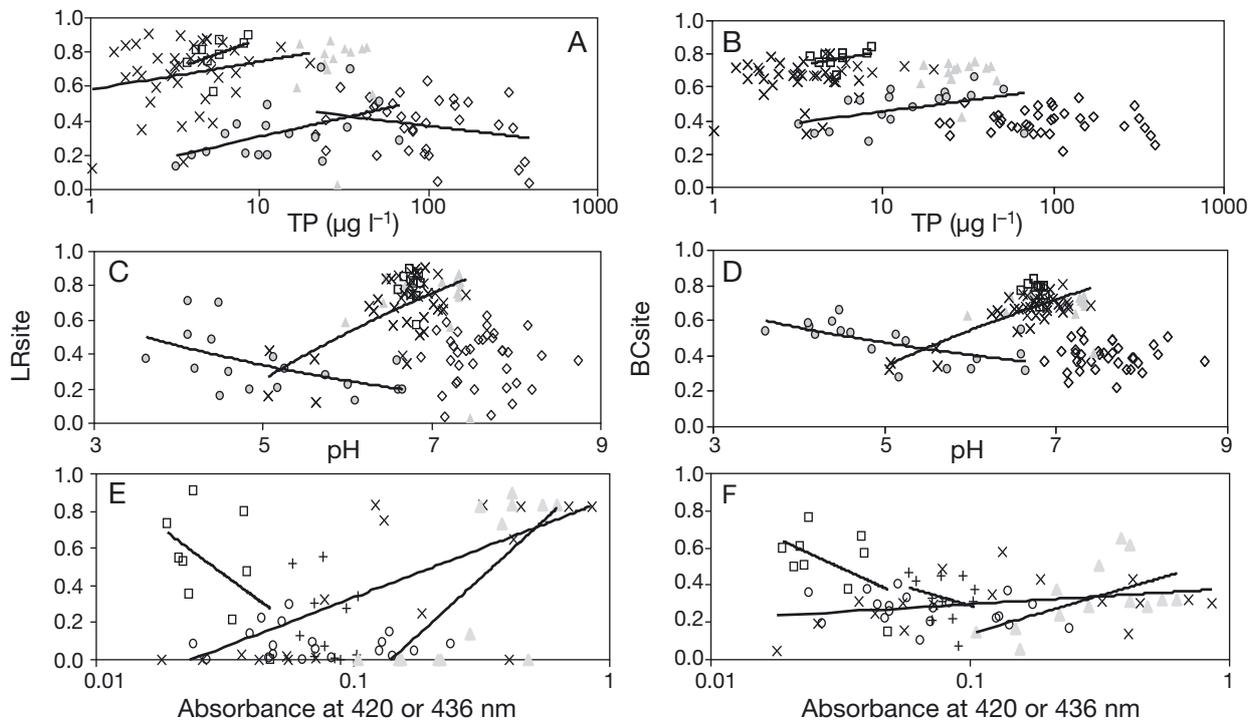


Fig. 4. Relationships between LRsite (the degree to which regional abundances can explain variation in local abundances) and BCsite (the similarity between local and average regional relative abundance) and environmental variables. For bacterioplankton (A–D): □ = Jämtland, ◇ = rock pools, × = Småland lakes, ○ = Småland streams, ▲ = Uppland. (A,B) LRsite and BCsite, respectively, along gradients of total phosphorus content (TP) in all the different data sets. (C,D) LRsite and BCsite, respectively, along pH gradients among streams (○) and lakes (×) in Småland. For phytoplankton, (E,F) show LRsite and BCsite, respectively, along gradients of absorbance at 420 or 436 nm. □ = Jämtland, + = Kolbäckån, × = Göteborg (which largely overlaps with Småland), ○ = Mälardalen, ▲ = Uppland. Only significant associations between LRsite and BCsite are shown by a regression line

Table 2. The amount of variation in community composition explained by spatial, hydrological and environmental variables for the different data sets using redundancy analysis (RDA). 'Variation explained' is the percentage explained by the final model (numbers in **bold** indicate significant models, $p < 0.05$). 'Predictor' shows which variable explained most of the variation in the final model. S: spatial variables; Abs: absorbance at 420 or 436 nm; Abs250: absorbance at 250 nm; cond.: conductivity; TP: total phosphorus; H: hydrological connectivity variables; –: No data entered into the final model

Organism	Data	Variation explained (%)	Predictor
Bacteria	Jämtland	21	S
	Uppland	20	Abs250
	Rock pools	25	S (10%), Cond. (8%), Abs (4%), Chl <i>a</i> (4%)
	Småland lakes	23	pH, TP
	Småland streams	17	pH
Phyto-plankton	Jämtland	26	H
	Uppland	39	Abs
	Götaland	59	TP
	Mälardalen	14	Abs (8%), pH (6%)
	Kolbäcksån	–	–

DISCUSSION

The general pattern emerging from the present study is that the extent to which local relative abundances of taxa match their relative abundances in a region is associated with both local environmental variables and large-scale spatial factors. Thus, relative abundances of taxa in local communities differ from what would be expected based on their relative abundances in the region due to deviations in local environmental conditions, spatial isolation, or both.

In most data sets, the results were qualitatively similar using LRsite and BCsite as measures of regional invariance (Table 1). The exceptions were LRsite being related to spatial variables in the phytoplankton data sets from Götaland and Mälardalen but not BCsite (S in Table 1). LRsite is more sensitive to co-variation in local and regional abundances of the most abundant taxa than BCsite, so this difference between LRsite and BCsite is probably caused by the most abundant taxa (*Gonyostomum semen* and *Botryococcus* spp.) being more strongly correlated to the spatial variables than the other taxa. However, as LRsite and BCsite produced similar results otherwise, we refer to both estimators when using the term 'degree of regional invariance'.

For bacterioplankton, the degree of regional invariance was associated with TP (Fig. 4a,b) or TN (Table 1) within regions except for the Uppland lakes data set. The degree of regional invariance increased

with nutrient levels in the nutrient-poorer data sets, whereas a negative relationship was found in the nutrient-richer rock pools (Fig. 4a,b). In the fifth bacterioplankton data set, Uppland, where all lakes had moderate nutrient levels, there was no evident association between nutrient levels and LRsite or BCsite. This is in line with both our previous observation that the average levels of regional invariance in a region decrease with increasing variation in phosphorus levels in that region (Östman et al. 2010) and that β -diversity of bacterioplankton in rock pools was positively related to average phosphorus levels among rock pools over time (Langenheder et al. 2012). Thus, it seems that communities deviating in local nutrient levels, lower or higher, from other communities in the region have a lower degree of invariance between local and regional abundances.

It is empirically well established that pH has a structuring effect on the community composition of both bacterio- and phytoplankton communities in lakes (Arnott et al. 2001, Logue & Lindström 2008, Newton et al. 2011), and a high variation in pH among communities in a region is associated with lower degree of regional invariance (Östman et al. 2010). The greatest ranges in pH in our data sets were found in the Småland lakes and streams (5.1 to 7.4 and 3.6 to 6.7, respectively). In the Småland lakes, the degree of regional invariance of bacterial communities increased with increasing pH (Fig. 4 c,d), that is, local relative abundances in lakes with higher pH (pH > 6) were more invariant to regional abundances than the local relative abundances in acidic lakes. In contrast, the degree of regional invariance showed negative relationships to pH among the streams in Småland, i.e. regionally abundant taxa were relatively more abundant in the acidic streams. The opposite pattern between data sets is likely caused by differences in the regional average of pH. Only 4 of the 19 streams had pH > 6, whereas only 4 of 40 lakes had pH < 6 (Table S1 in the supplement). We conclude that large differences in local pH and regional average pH seem to act as a filter between local and regional abundances causing low degrees of regional invariance in deviating environments.

Absorbance and C_{org} , which both relate to the amount of humic and coloured substances in the water (Wetzel 2001, Mattsson et al. 2005), were the variables most frequently associated with the degree of regional invariance in the phytoplankton data sets (Table 1, Fig. 4e,f). The sign of the relationship depended on the concentration of humic substances. In regions with humic waters (Götaland and Uppland), the regionally dominant phytoplankton taxa were

usually the species that were dominant in humic waters, especially *Gonyostomum semen* and to some extent *Botryococcus braunii*. Yet, these species were absent or rare in less humic lakes, resulting in a low degree of regional invariance in these lakes. In contrast, in the more oligotrophic Jämtland lakes and lakes along the Kolbäckån river, the regionally dominant species (species of *Ochromonas*, *Malomonas*, and *Rhodomonas*) were most abundant in the clearest lakes. Although these species occurred in the less clear lakes, they were not as abundant there, resulting in a negative relationship between humic matter content and degree of regional invariance (Fig. 4e,f).

In most cases, LRsite and BCsite calculated from communities close to each other did not differ from a random subsample of communities from the region (data set) (Figs. 2 & 3), i.e. variation in distances to the communities used for calculating regional abundances had no evident effect on the degree of community specific regional invariance. Although the actual distances between communities used here were set somewhat arbitrarily, the results were robust across the actual distances employed (see Figs. 2 & 3), and we believe that the results obtained would not have changed if other intervals had been used.

We did, however, record some spatial structure related to the location of communities or hydrological connectivity between them, especially in the LRsite of phytoplankton data sets. This suggests that the most regionally abundant taxa were more dominant in some parts of the region than in others. In the phytoplankton data sets from Mälardalen and Götaland, LRsite decreased in an eastward direction. This finding coincides with an increase in genus richness of phytoplankton in an eastward direction from the Atlantic coast toward Sweden and Finland, which could be related to changes in average phosphorus concentrations over larger spatial scales (>100 km) (Ptacnik et al. 2008, 2010). Whether these 2 patterns are linked and caused by the same mechanisms should be addressed in future studies. Although the results of spatial predictors are not consistent across data sets, we conclude that dispersal pathways, large scale environmental properties, or different histories of different parts of regions seem to be more important in determining regional invariance than spatial distances or the size of the region per se. This is also what Crump et al. (2007) and Soininen et al. (2011) concluded from studies on β -diversity of microbial plankton.

In most of the data sets studied here, the variable(s) explaining differences in the degree of regional invariance were either the same as or co-varied with

those explaining β -diversity based on the RDA. This suggests that, at least partly, the differences in regional invariance observed among local communities could be attributed to deterministic species-sorting processes in relation to local environmental condition or spatial differences (e.g., dispersal pathways or history). In the bacterial rock pool data set, variation in nutrient loads had no clear structuring effect on community composition (Table 2; Langenheder & Ragnarsson 2007) but nevertheless showed an association to the degree of regional invariance (Table 1, Fig. 4). Similarly, the decrease in regional invariance eastward in the Mälardalen phytoplankton data set, where lakes also had higher nutrient levels, was not reflected in the RDA. As the changes in degree of regional invariance are not evidently related to structural changes in community composition along these gradients, this suggests that processes other than species sorting may be driving community dynamics in more nutrient rich communities. For example, stronger priority effects at high nutrient loads could cause communities in nutrient rich environments to be less similar and not differentiate in relation to a local condition, which was shown experimentally by Chase (2010). Alternatively, more taxa are, or become, more equivalent in terms of growth rates, resulting in more neutral community dynamics in nutrient richer environments. Studies of regional invariance may serve as a useful complement to previous statistical methods for analysing community assembly processes.

In summary, our results suggest that both local environmental variation and regional factors cause a deviation in the degree of regional invariance among communities. For bacterioplankton especially, differences in nutrient levels and pH rendered communities with different degrees of invariance between local and regional abundances. For pH, this seems to depend on structural changes in community composition, whereas variation in priority effects or neutral dynamics may be more important for the differences in regional invariance with different nutrient levels. Deviations in water colour and nutrient levels were associated with the degree of regional invariance in phytoplankton communities, which also explained variation in community composition. In particular for phytoplankton, differences in the degree of regional invariance among communities were also related to spatial factors. We also conclude that differences in the degrees of regional invariance seem to be related to community processes causing structural variation in community composition along gradients or locations in many of the data sets and thereby detected

by other methods (e.g. RDA). However, in some cases, in particular along nutrient gradients, estimates of regional invariance hinted at the existence of other important community assembly processes, suggesting estimates of regional invariance to be a useful complement to other methods. Although we cannot identify the actual mechanism(s) causing aquatic microbial communities to be regionally invariant in the first place, we have provided some evidence for considering both local environmental conditions and spatial factors if we want to understand when and why communities show high or low degrees of regional invariance.

Acknowledgements. We thank 2 anonymous reviewers for improvements to an earlier version of the manuscript. Financial support for the present study was provided by the Swedish Research Council (to Ö.Ö., E.S.L.), Formas (S.L.), and the Crafoord Foundation (E.S.K.). The Olsson-Borgh's foundation financed sampling of microbial communities in Jämtland and Uppland, and the Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, provided and compiled 3 phytoplankton data sets from Mälardalen, Götaland, and Kolbäckån.

LITERATURE CITED

- Arnott SE, Yan N, Keller W, Nicholls K (2001) The influence of drought-induced acidification on the recovery of plankton in Swan Lake (Canada). *Ecol Appl* 11:747–763
- Beisner BE, Peres-Neto PR, Lindström ES, Barnett A, Longhi ML (2006) The role of environmental and spatial processes in structuring lake communities from bacteria to fish. *Ecology* 87:2985–2991
- Blanchet FG, Legendre P, Borcard D (2008a) Modelling directional spatial processes in ecological data. *Ecol Modell* 215:325–336
- Blanchet FG, Guillaume F, Legendre P, Borcard D (2008b) Forward selection of explanatory variables. *Ecology* 89:2623–2632
- Chase JM (2010) Stochastic community assembly causes higher biodiversity in more productive environments. *Science* 328:1388–1391
- Crump BC, Adams HE, Hobbie JE, Kling GW (2007) Biogeography of bacterioplankton in lakes and streams of an arctic tundra catchment. *Ecology* 88:1365–1378
- Drakare S, Liess A (2010) Local factors control the community composition of cyanobacteria in lakes while heterotrophic bacteria follow a neutral model. *Freshw Biol* 55:2447–2457
- Griffith DA, Peres-Neto PR (2006) Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology* 87:2603–2613
- Langenheder S, Ragnarsson H (2007) The role of environmental and spatial factors for the composition of aquatic bacterial communities. *Ecology* 88:2154–2161
- Langenheder S, Berga M, Östman Ö, Székely AJ (2012) Temporal variation of beta diversity and assembly mechanisms in a bacterial metacommunity. *ISME J* 6:1107–1114
- Lindström ES, Langenheder S (2012) Local and regional factors influencing bacterial community assembly. *Environ Microbiol Rep* 4:1–9
- Lindström ES, Östman Ö (2011) The importance of dispersal for bacterial community composition and functioning. *PLoS ONE* 6:e25883
- Lindström ES, Feng XM, Graneli W, Kritzberg ES (2010) The interplay between bacterial community composition and the environment determining function of inland water bacteria. *Limnol Oceanogr* 55:2052–2060
- Logue JB, Lindström ES (2008) Biogeography of bacterioplankton in inland waters. *Freshw Rev* 1:99–114
- Martiny JBH, Bohannan BJM, Brown JH, Colwell RK and others (2006) Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* 4:102–112
- MathWorks (2011) MatLabR2011a. Natick, MA
- Mattsson T, Kortelainen P, Räike A (2005) Export of DOM from boreal catchments: impacts of land use cover and climate. *Biogeochemistry* 76:373–394
- 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
- Östman Ö, Drakare S, Kritzberg ES, Langenheder S, Logue JB, Lindström ES (2010) Regional invariance among microbial communities. *Ecol Lett* 13:118–127
- Ptácnik R, Solimini AG, Andersen T, Tamminen T and others (2008) Diversity predicts stability and resource use efficiency in natural phytoplankton communities. *Proc Natl Acad Sci USA* 105:5134–5138
- Ptácnik R, Andersen T, Brettum P, Lepistö L, Willen E (2010) Regional species pools control community saturation in lake phytoplankton. *Proc Biol Sci* 277:3755–3764
- SAS Institute (2008) SAS version 9.2. SAS Institute, Cary, NC
- Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP (2006) Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol* 8:732–740
- Soiminen J, McDonald R, Hillebrand H (2007) The distance decay of similarity in ecological communities. *Ecography* 30:3–12
- Soiminen J, Korhonen JJ, Karhu J, Vetterli A (2011) Disentangling the spatial patterns in community composition of prokaryotic and eukaryotic lake plankton. *Limnol Oceanogr* 56:508–520
- Van der Gucht K, Cottenie K, Muylaert K, Vloemans N and others (2007) The power of species sorting: local factors drive bacterial community composition over a wide range of spatial scales. *Proc Natl Acad Sci USA* 104:20404–20409
- Verleyen E, Vyverman W, Sterken M, Hodgson DA and others (2009) The importance of dispersal related and local factors in shaping the taxonomic structure of diatom metacommunities. *Oikos* 118:1239–1249
- Vyverman W, Verleyen E, Sabbe K, Vanhoutte K and others (2007) Historical processes constrain patterns in global diatom diversity. *Ecology* 88:1924–1931
- Wetzel RG (2001) *Limnology* 3rd edn. Academic Press, San Diego, CA