Neutral community model explains the bacterial community assembly in freshwater lakes

Adélaïde Roguet, Grégory Laigle, Claire Therial, Adèle Bressy, Frédéric Soulignac, Arnaud Catherine, Gerard Lacroix, Ludwig Jardillier, Céline Bonhomme, Thomas Lerch, et al.

Neutral community model explains the bacterial community assembly in freshwater lakes

Adélaïde Roguet[1,*], Grégory S. Laigle[1], Claire Therial[1], Adèle Bressy[1], Frédéric Soulignac[1], Arnaud Catherine[2], Gérard Lacroix[3,4], Ludwig Jardillier[5], Céline Bonhomme[1], Thomas Z. Lerch[6] and Françoise S. Lucas[1]

[1]Laboratoire Eau Environnement et Systèmes Urbains (UMR MA 102), Université Paris-Est, AgroParisTech, Faculté des Sciences et Technologie, 61 avenue du Général de Gaulle, FR 94000 Créteil, France.

[2]Unité Molécules de Communication et Adaptation des Micro-organismes (UMR 7245), Sorbonne Université, Muséum National d'Histoire Naturelle, Case 39, 57 rue Cuvier, FR 75005 Paris, France.

[3]Institute of Ecology and Environmental Sciences of Paris (UMR 7618 [UPMC, UPEC, Paris Diderot, CNRS, IRD, INRA]), Université Pierre et Marie Curie, Bâtiment A, 7 quai St Bernard, FR 75005 Paris, France.

[4]CEREEP – Ecotron Ile De France (UMS 3194 [CNRS, ENS]), Ecole Normale Supérieure, 78 rue du Château, 77140 St-Pierre-lès-Nemours, France.

[5]Écologie Systématique Évolution, Univ. Paris-Sud, CNRS, AgroParisTech, Université ParisSaclay, 91400, Orsay, France.

[6]Institute of Ecology and Environmental Sciences of Paris (UMR 7618 [UPMC, UPEC, Paris Diderot, CNRS, IRD, INRA]), Université Paris-Est Créteil, Faculté des Sciences et Technologie, 61 avenue du Général de Gaulle, FR 94000 Créteil, France.

*__Corresponding author:__ Laboratoire Eau Environnement et Systèmes Urbains (UMR MA 102), Université Paris-Est, AgroParisTech, Faculté des Sciences et Technologie, 61 avenue du Général de Gaulle, FR 94000 Créteil, France. Tel: +33145171630; E-mail: adelaide.roguet@gmail.com

One sentence summary: At both regional and local scales, the neutral community model appears to be relevant in explaining the bacterial assemblage structure in shallow freshwater lakes located in the Paris area.

Keywords: bacterioplankton; community structure; deterministic factors; neutral community model; lake; T-RFLP

Abstract

Over the past decade, neutral theory has gained attention and recognition for its capacity to explain bacterial community structure (BCS) in addition to deterministic processes. However, no clear consensus has been drawn so far on their relative importance. In a metacommunity analysis, we explored at the regional and local scale the effects of these processes on the bacterial community assembly within the water column of 49 freshwater lakes. The BCS was assessed using terminal restriction fragment length polymorphism (T-RFLP) of the 16S rRNA genes. At the regional scales, results indicated that the neutral community model well predicted the spatial community structure (R^2_{mean} = 76%) compared with the deterministic factors – which explained only a small fraction of the BCS total variance (less than 14%). This suggests that the bacterial compartment was notably driven by stochastic processes, through loss and gain of taxa. At the local scale, the bacterial community appeared to be spatially structured by stochastic processes (R^2_{mean} = 65%) and temporally governed by the water temperature, a deterministic factor, even if some bacterial taxa were driven by neutral dynamics. Therefore, at both regional and local scales the neutral community model appeared to be relevant in explaining the bacterial assemblage structure.

Introduction

One of the key issues in microbial ecology is to identify and quantify ecological processes that drive bacterial community assembly in aquatic environments. Freshwater bacterial communities appear to be shaped by environmental conditions or dispersal-related processes (*e.g.*, Martiny *et al.* 2006; Lindström and Langenheder 2012). Taxa selection by local environmental conditions was suggested to be the main mechanism controlling aquatic bacterial biogeography (*e.g.*, Beisner *et al.* 2006; Langenheder and Ragnarsson 2007; van der Gucht *et al.* 2007). Environmental factors, such as water temperature, osmotic conditions or nutrient availability, are known to shape the bacterial community at the regional and local scale (*e.g.*, Muylaert *et al.* 2002; Yannarell and Triplett 2004, 2005; Kent *et al.* 2007; Lindström, Kamst-Van Agterveld and Zwart 2005; Shade *et al.* 2007; Jones, Newton and McMahon 2009). Inversely, dispersal related mechanisms were less often identified as having an influence on aquatic bacterial metacommunities (*e.g.*, Logue and Lindström 2010; Östman *et al.* 2010; Soininen *et al.* 2011), which is likely due to an underestimation or a difficulty in taking into account these processes (Lindström and Langenheder 2012). Dispersal-related mechanisms are represented by (i) dispersal limitation, *i.e.* the extent to which taxa reach another location (Martiny *et al.* 2006), (ii) the mass effect that corresponds to a massive supply of an exogenous taxa that disturbs the composition of a local community (Hubbell 2001; Leibold *et al.* 2004) and (iii) the neutral model that describes the stochastic balance between the immigration, speciation, emigration and extinction of organisms (Hubbell 2001; Leibold *et al.* 2004).

To date, only a few studies using statistical approaches, such as variation partitioning analysis (Borcard, Legendre and Drapeau 1992), have conjointly quantified the relative importance of local environmental factors and processes involved in the spatial dispersion of the bacterial taxa (*e.g.*, Langenheder and Ragnarsson 2007; van der Gucht *et al.* 2007). The neutral assembly theory has seldom been included in these studies (*e.g.*, Drakare and Liess 2010; Langenheder and Székely 2011), although it can correctly explain on its own the bacterial community structure (BCS) in diverse aquatic environments at the regional and local scales (*e.g.*, Sloan *et al.* 2006; Woodcock *et al.* 2007; Ofiţeru *et al.* 2010).

In this study, we investigated the spatial distribution of freshwater bacterioplankton in a set of 49 shallow and artificial lakes located in the same hydrographical basin around Paris (France) for three consecutive years in summer. These ecosystems constitute a useful model to determine the relative importance of local environmental factors and spatial and neutral processes since they display large environmental gradients and strong variations in variables that could impact the distribution of bacteria *(e.g.*, river or watershed connections) (Catherine *et al.* 2008, 2010). Our study had three specific aims. First, we sought to determine to what extent local environmental characteristics and spatial factors shaped the regional distribution of the bacterial community. Then, in an attempt to confirm or refute the role of environmental and spatial factors in explaining the observed BCS, we evaluated whether stochastic dynamics could accurately predict the metacommunity structure. Finally, because the different processes could also shape the BCS at the local scale, the importance of environmental, spatial and neutral processes was evaluated in a single lake by monthly monitoring over 2 years. Bacterioplankton community structure was assessed using terminal restriction fragment length polymorphism (T-RFLP) targeting a fragment of the 16S rRNA gene that allows screening of a large set of samples.

Materiel and methods

Study area and sampling

This study was conducted in the Paris area (Fig. 1), which is the most populated area in France (with 18% of the metropolitan French population) and covers about 12 000 km^2 (INSEE 2013). This region displays a large gradient of land use. In spite of large industrial towns and residential suburbs, half of the Paris area territory is used for agricultural purposes, while 26% is still covered by forests (INSEE 2012). Among the 248 water bodies larger than 5 ha referenced in the hydrological database Carthage 3.09® (IGN, Paris, France), 49 lakes (Fig. 1 and Supplementary Table S1) were selected using a random and stratified sampling strategy (Catherine *et al.* 2008). Brie y, the 49 lakes were chosen to represent an unbiased set of water

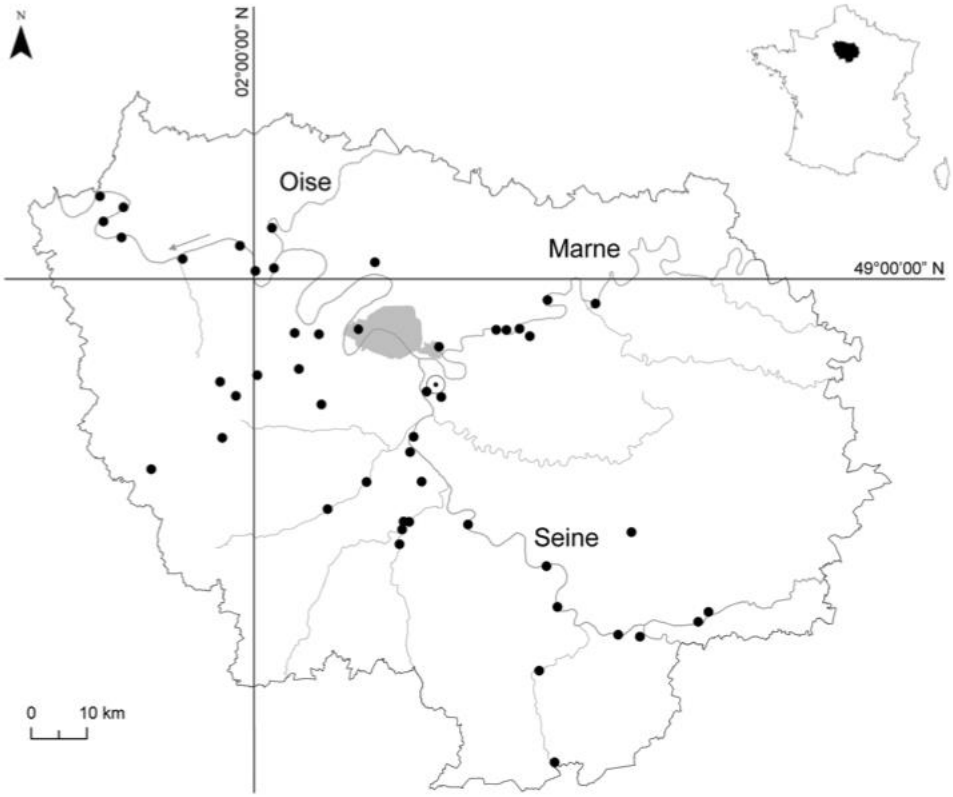

Figure 1. Position of the 49 lakes in the Paris area (France). Créteil Lake is indicated by an open circle.

bodies that reflects the whole range of environmental conditions (*e.g.*, land use, hydrology, altitude and surface) found within the Paris area. All the 49 lakes have an artificial anthropogenic origin and can be considered as shallow according to the definition of Scheffer (2004). The lakes were sampled yearly in late July from 2011 to 2013. Each sampling campaign was conducted in less than 15 days to reduce the variability caused by short-term changes in meteorological conditions and nutrient inputs.

For each lake, three equidistant sampling stations were selected. At each station, water samples were collected at three discrete depths (depending on the depth of the water column) using a Niskin bottle (General Oceanics Inc., Miami (FL), USA). All samples were then pooled to obtain an integrated sample.

Immediately after sampling, water was filtered through a 0.22 μm pore-size Sterivex GP filter (Millipore, Billerica (MA), USA) after pre-filtration through 50 μm pore-size nylon mesh. Sterivex filters were kept at 4°C during transport and then stored at −20°C.

Among the 49 lakes, Créteil Lake (Figs 1 and 2) was monitored monthly at three different stations from December 2011 to December 2013 (25 sampling dates) in order to identify spatial and temporal variability in the BCS. This is a mesotrophic lake (Table 1) covering 40 ha in an urbanized area (Val-de-Marne, France). It is a former sandpit mainly supplied by alluvial groundwater. A storm sewer outlet drains an impermeable surface of 100 ha and releases its effluent into the lake. The water residence time of this lake is greater than 180 days.

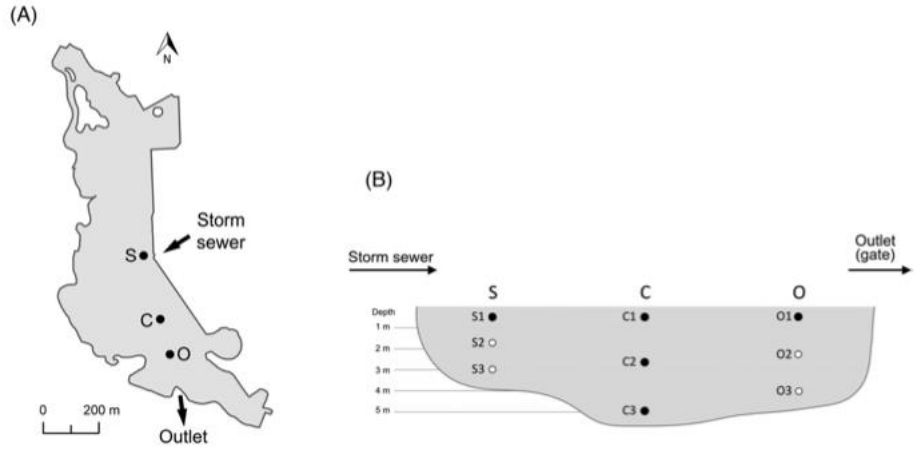

Figure 2. Sampling strategies in the Créteil Lake. (A) Location of the three stations. (B) Representation of the two transects: the horizontal transect (S1, C1 and O1) and the vertical transect (C1, C2 and C3).

A horizontal and a vertical transect were studied (Fig. 2). Subsurface samples were obtained at three stations along the inlet– outlet axis of the lake (Storm outlet (S1), center of the lake (C1), and lake out ow (O1)). At the central station, the vertical axis of the lake was sampled at three depths (C1, C2 and C3). In addition, for all sampling dates, a pooled sample (M) was obtained as previously described to estimate the BCS variation in the integrated sample obtained in summer during the 49 lake campaigns. The average point M was the mix of the three stations at the three depths (S1, S2, S3, C1, C2, C3, O1, O2 and O3). All samples were stored at 4°C in the dark until preparation (3–4h after collection). For each sample, 1 L of lake water was filtered through a Sterivex GP filter cartridge as previously described and then stored at −20°C prior to analysis.

Table 1. Characteristics of Créteil Lake. Values are the means (standard deviations) of data obtained from December 2011 to December 2013.

Parameter	Values
Location	48°46.50'N 2°27.10'E
Surface area (km^2)	0.40
Mean depth (m)	4.5
Max depth (m)	6.0
Trophic status	Mesotrophic
Total phosphorous ($\mu gP\ L^{-1}$)	64.7 (41.8)
Chlorophyll a ($\mu g\ L^{-1}$)	6.3 (4.3)
Secchi depth (m)	2.4 (0.9)
COD ($mgC\ L^{-1}$)	6.1 (0.6)
COP ($mgC\ L^{-1}$)	0.8 (0.3)
TSS ($mg\ L^{-1}$)	4.1 (1.5)
pH	7.8 (0.5)
Conductivity ($\mu S\ cm^{-1}$)	1509 (63)

DOC, dissolved organic carbon; POC, particulate organic carbon.

DNA extraction

For each sample, the membrane inside the Sterivex units was extracted under sterile conditions and cut into small pieces of approximately 1 mm^2. All membrane pieces were pooled in a sterile tube and DNA was extracted using the FastDNA® SPIN Kit (QBiogene, Carlsbad (CA), USA) according to the manufacturer's instructions. Two modifications of this protocol were applied: bacterial cells were lysed in a FastPrep bead beater three times for 30 s at 4.0 m s^{-1} and an additional wash was performed on the SPIN filters.

T-RFLP analysis

A 1326 bp fragment of the 16S rRNA gene was amplified by PCR using the primer set 63F (5′-CAGGCCTAACACATGCAAGTC-3′ ; labeled with 6-carboxy fluorescein) and 1392R (5′ -ACG GGCGGTGTGTACAAG-3′) targeting the bacterial domain (Osborn, Moore and Timmis 2000). Each 20 µL reaction contained 20 ng of template DNA, 0.2 µM of each primer (Microsynth, Balgach, Switzerland), 120 µM of each deoxynucleoside triphosphates (Promega, Madison (WI), USA), 1 mM of $MgCl_2$, 0.1 mg mL^{-1} of bovine serum albumin (BSA), 1× GoTaq® Colorless buffer and 0.5 U of GoTaq DNA polymerase (Promega, Madison (WI), USA). Reactions were carried out in a T1 thermocycler (Biometra, Göttingen,

Germany) with the following cycle: initial denaturation at 94°C for 2 min, 20 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min, extension at 72°C for 2 min, and a final extension at 72°C for 10 min. The amplification and the size of the amplicons were checked on a 1% agarose gel. Amplicons were digested using the restriction enzyme AluI (Promega, Madison (WI), USA) (Osborn, Moore and Timmis 2000) with 10 U for 3 h at 37°C. Restriction digests were desalted by ethanol precipitation in the presence of glycogen and subsequently resuspended in deionized formamide and size standard GeneScan 500 LIZ (Life Technologies, Carlsbad (CA), USA).

The terminal restriction fragments (T-RFs) were separated using the automated sequencer ABI Prism 3130 (Life Technologies, Carlsbad (CA), USA) at the IMRB facility (Institut Mondor de Recherche Biomédicale at University Paris-Est Créteil). The resolution was of ∼1 bp for fragments up to 500 bp. Raw data files containing peak information were tabulated in the Peak Scanner™ software v 1.0 (Life Technologies, Carlsbad (CA), USA). Only peaks between 60 and 500 bp were selected. TRFLP data were then processed and analysed with the on-line tool T-REX (Culman *et al.* 2009). Data were subjected to quality control procedures: noise filtering (peak area, standard deviation multiplier = 2) and T-RF alignment (clustering threshold = 0.5). T-RFs detected in only one sample were not taken into account. Relative abundances were calculated using peak areas.

Bacterial enumeration

Bacterial counts (heterotrophic bacteria and cyanobacteria) from the Créteil Lake samples were assessed by ow cytometry using a Becton Dickinson FACScan (BD Biosciences, Oxford, UK) instrument equipped with a 15 mW 488 nm laser. Pigmented microorganisms were identified based on the auto fluorescence of the chlorophyll *a* thanks to the orange fluorescence (FL2, 585/42 bandpass filter) vs red fluorescence (FL3, 650 nm longpass filter). Cells containing chlorophyll *a* were distinguished using side scatter (SSC) vs FL3 and distinct size classes were clustered to discriminate cyanobacteria from small and large picoeukaryotes. A subsample of each sample was stained with the nucleic acid stain SYBr Green I (Marie *et al.* 1997) in order to visualize populations of heterotrophic organisms using SSC vs green fluorescence (530/30 bandpass filter). Chlorophyll *a*-containing organisms were gated off plots of SSC vs FL1, having been identified in plots of FL3 vs FL1. Populations were enumerated using a syringe-pump calibrated to 0.5 and 1 μm microspheres (Polysciences, Eppelheim, Germany) according to Zubkov and Burkill (2006).

Environmental and meteorological parameters

For the 49 lakes, the total nitrogen and phosphorus concentrations were assayed by colorimetry using a Cary 50 Scan spectrophotometer (Varian Inc., Palo Alto (CA), USA) respectively according to Rogora *et al.* (2006) and the French standard (AFNOR NF T 90–023).

Secchi depth and vertical profiles of physicochemical parameters were determined for each lake at the three stations (49 lakes and Créteil Lake). Chlorophyll *a* (Chl *a*) concentration was determined using a FluoroProbe *in situ* fluorometer (BBE Moldaenke GmbH, Kiel, Germany). Conductivity, temperature, pH and oxygen profiles were measured using a submersible CTD profiler SBE 19 Seacat (Sea-Bird Electronics Inc., Bellevue (WA), USA).

Concentrations of total suspended solids (TSS) were quantified after filtration of 1 L of lake water on a precombusted tarred Whatman GF/F filter. Dissolved organic carbon (DOC) concentrations were measured

using a TOC-VCSN carbon analyzer (Shimadzu, Columbia (MD), USA).

Meteorological data (temperature and precipitation) were also collected during the Créteil Lake campaigns using a weather real-time transmitter WXT520 (Vaisala Inc., Boulder (CO), USA) placed on a sensing platform LakeESP (PME Inc., Vista (CA), USA).

Watershed land use assessment

Upstream watersheds were delineated for each of the 49 lakes based on the digital elevation model (DEM) BD ALTI® 15 m (IGN, Paris, France) using ArcGIS 10.0 (ESRI Inc., Redland (CA), USA). Then, they were adjusted to take into account physical barriers that potentially modify the water flow (*e.g.*, roads or underground drainages). Land use of the adjusted watershed was classified into four categories (*i.e.* natural, agricultural, open peri-urban and dense urbanized areas) using the MOS® GIS 2012 database (IAURIF, Paris, France).

Data analysis

Drivers of bacterioplankton spatial distribution in the Paris area

Based on the 49 lakes dataset, we first checked if the intra-annual variability of the BCS among lakes differed significantly from inter-annual variability within each lake using the Mann– Whitney test based on the Bray–Curtis dissimilarity index. The beta diversity index or the dissimilarity between lakes was calculated for each sampling year based on the Sørensen index. Beta diversity was then partitioned into two components following the framework proposed by Baselga (2010) to quantify the fraction of dissimilarity explained by the T-RF replacement (turnover, based on Simpson's dissimilarity index) and from the random variations of the T-RF richness (nestedness) (Azeria *et al.* 2011). These parameters were estimated using the *beta.pair* function from the 'betapart' package (Baselga *et al.* 2013).

Then, the relative importance of environmental and spatial factors was assessed by decomposing the total bacterial community variation (Peres-Neto *et al.* 2006) using a variance partitioning analysis (VPA). Among all the environmental parameters measured, only six non-collinear variables were retained for the statistical analysis, *i.e.* water temperature, conductivity, pH, DOC concentration, trophic status and dominant land use index of the watershed. Trophic status (TS) was determined for each lake according to OECD (1982). Dominant land use index was included to assess the putative role of land use on the bacterial community. It was assessed as the land use (among the four categories) with the highest percentage of occupation. Spatial factors comprised two components, (i) variables reflecting the dispersal of bacterial taxa, *i.e.* with the presence of storm sewer outlet into the lake and the type of link to the hydrographic network, and (ii) variables integrating the community structuring at the regional scale, evaluated using the eigenvectors derived from the principal components of neighbor matrices of spatial coordinates (PCNM) (Borcard and Legendre 2002; Legendre and Gauthier 2014). The type of linkage to the hydrographic network was decomposed into three categories (which roughly characterized the lake water retention time): (i) isolated lakes, which were mainly filled by alluvial water; (ii) lakes crossed by rivers and located in the riverbeds; and (iii) connected lakes, which were linked to a river but only by a single connection (*e.g.*, small bond, pipe) and thus received water from both groundwater and river. VPA estimated the proportion of BCS variation (adjusted *R*-squared (R^2_{adj})) that can be attributed to local environmental characteristics [E] and spatial [S] components, environmental conditions without spatial component [E|S], spatial without

environmental component [S|E], variation explained by the interaction between both components [E∩S] and the unexplained variation (1 − [E + S]). The significance of the partial contribution of both components was also evaluated with a Monte Carlo permutation test (999 permutations under the reduced model). Analyses were conducted using the *varpart* function from the 'vegan' package (Oksanen *et al.* 2015). Prior to VPA, forward selections were performed for both components according to Borcard, Gillet and Legendre (2011) and Legendre and Gauthier (2014).

To assess the neutral assembly of the bacterial communities, we used the method developed by Sloan *et al.* (2006) to fit the regional relative abundance of the T-RFs and their observed detection frequency. Contrary to Hubbell's discrete model (Hubbell 2001), this continuous model is particularly suited for bacterial communities (*i.e.* with large population size) detected by fingerprinting methods (Sloan *et al.* 2006). The parameter $N_T m$ depicts the relationship between detection probability and regional relative abundance, with N_T corresponding to the size of the metacommunity and m to the immigration rate (the probability that a dead individual is replaced by an immigrant). By considering the metacommunity size to be roughly equal among regions, $N_T m$ estimates the dispersal connectivity between each lake community. This parameter was estimated using the best fit between detection frequency of T-RFs and their regional relative abundance, by minimizing the sum of squares of errors. The detection limit was fixed to 0.005, which corresponds approximately to the threshold of the T-RFLP method (*i.e.* electrophoregram peaks with more than 0.5% of the total peak area of the sample). The goodness of fit was evaluated using the determination coefficient R^2.

Spatio-temporal variability of BCS in Créteil Lake

The beta diversity as well as the turnover and the nestedness index were calculated over the 2 years of survey, considering consecutive pairwise months. Impacts of the environmental factors on the three indexes were investigated using a linear model. Data from Créteil Lake were used to test for significant variability in BCS along the horizontal and vertical transects. The presence of a space–time interaction was first evaluated using a two-way ANOVA crossed design as described by Legendre, de Cáceres and Borcard (2010). Included in the 'STI' package (Legendre, Borcard and de Cáceres 2012), this method is particularly recommended for data without replication. As no space–time interaction was identified (data not shown), the same model without interaction was then implemented (STI Model 2, STI2) to test the significance of spatial and temporal variability along the two transects. Moreover, the environmental factors (water temperature, cumulative precipitation on 1, 5 and 15 days, pH, saturated oxygen, TSS, Chl *a*, DOC and total phosphorous concentrations) shaping the BCS over the 2-year survey were then investigated on Hellinger-transformed data using redundancy analysis (RDA), considering the sampling date as covariable. The neutral community model was performed as previously described to evaluate the importance of stochastic dynamics on the spatial distribution (between S1, C1, C2, C3 and R1) of the bacterial assemblage for each of the 25 campaigns. The parameter N_T was estimated from the bacterial counts. Moreover, on the average sample M, we used the model developed by Ofiţeru *et al.* (2010) to evaluate to what extent the neutral model could explain the temporal variations of the relative abundance for each T-RF.

All statistical analyses were conducted using the statistical environment R version 3.1.1 (R Development Core Team 2014).

Results

Spatial distribution of the BCS at the regional scale

A total of 295 T-RFs were detected by T-RFLP analysis. Overall, 43% of T-RFs were removed after quality control. The 167 remaining T-RFs were used for the statistical analyses with a median of 33 T-RFs per lake. The majority (69%) of the T-RFs were found every summer in at least one lake, while 24% of the T-RFs were detected twice and 7% once. Furthermore, if only two T-RFs were present in all the 49 lakes, none was restricted to a single lake.

Per lake, the BCS dissimilarity was on average 54% between the years 2011 and 2012, and 53% between 2012 and 2013. These variations were significantly smaller than the beta diversity, *i.e.* the inter-lake variability considering the same year of sampling (Mann–Whitney test, $P < 0.001$). Indeed, the beta diversity displayed 63, 59 and 54% of dissimilarity for 2011, 2012 and 2013, respectively. Overall, by partitioning the beta diversity for the three years of campaigns, we observed that the variations between lakes were dominated by T-RF replacements that explained respectively 76, 70 and 67% of the beta diversity, whereas the random shift of T-RFs richness explained only 24, 30 and 33% of the beta diversity.

Whatever the year of sampling, environmental conditions and spatial factors explained only a small portion of the BCS variations, as evaluated by the variance partitioning analysis (VPA) (Table 2). On average over the 3 years of monitoring, the environmental variables alone (mainly the trophic status) and the spatial processes alone (*i.e.* the type of connection to the hydrological network and the spatial structuring of the bacterial community characterized by PCNM variables) and their interaction explained respectively 3, 4 and 3% of the BCS variations. Consequently, a large amount of the variance (on average 90%) remained unexplained by these variables (Table 2).

Table 2. Variation partitioning analysis of the bacterial community structure for the three years of sampling.

	Relative variance explained (adjusted R-squared)						Variables selected	
Year of sampling	Env. factors [E]	Spatial factors [S]	Env. factors alone [E|S]	Spatial factors alone [S|E]	Interaction [E∩S]	Residuals	Env. factors	Spatial factors
2011	0.05	0.06	0.03*	0.04*	0.02	0.91	TS, pH	Hydro, PCNM (no. 6)
2012	0.08	0.11	0.03*	0.06*	0.05	0.86	TS, DOC	Hydro, PCNM (no. 2,4,7,9)
2013	0.06	0.03	0.04*	0.01^{NS}	0.02	0.93	TS	Hydro, PCNM (no. 2,7,9)
Mean (sd)	0.06 (0.02)	0.06 (0.04)	0.03 (0.01)	0.04 (0.03)	0.03 (0.02)	0.90 (0.04)		

Symbols associated with adjusted R^2 correspond to the significance (* $P < 0.05$; NS, non significant) of the partial contribution of factors tested alone in the presence of the other. Abbreviations: Env, environmental; TS, trophic status; DOC, dissolved organic carbon concentration; Hydro, type of linkage to the hydrographical network; PCNM (principal components of neighbor matrices), reflect the community structuring at the spatial scale using PCNM variable; SD, standard deviation.

The neutral model explained a large fraction of the relationship between the occurrence frequency of T-RFs and their relative abundance variations (Fig. 3), with 82, 75 and 70% of explained variance for 2011, 2012 and 2013 databases. However, assuming a similar N_T in each case, large variations of the immigration rate m were observed over the 3 years of sampling. Indeed, the value of N_Tm increased from 13 in 2011 to 79 in 2013. Furthermore, since all the relationships seemed strongly constrained by two T-RFs (which were present during the three campaigns), simulations without these points were performed again. The same values of N_Tm but a decrease of about 3% of total explained variance were obtained (data not shown), suggesting that these two T-RFs did not bias the immigration rate estimates.

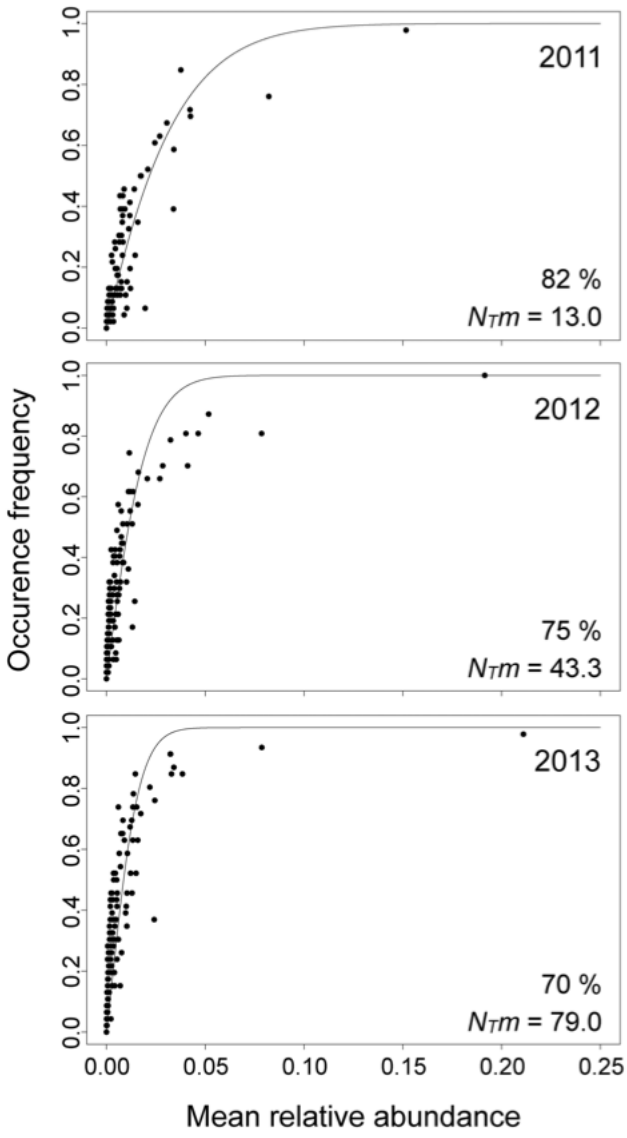

Figure 3. Relationship between the detection frequency of T-RFs and their mean relative abundance from the three years of monitoring. Each point represents a different T-RF. The lines show the best fit to the neutral model with random immigration from a common source pool as described by Sloan et al. (2006). N_Tm values represent the meta-community size multiplied by the immigration rate. The percentages represent the R^2 value of the fit.

Spatial and temporal monitoring in Créteil Lake

In Créteil Lake, the 2-year monitoring of the bacterial community by T-RFLP allowed the detection of 311 T-RFs. As for the 49 lakes, about 47% of the T-RFs were removed after quality control. A total of 166 T-RFs were used for the statistical analysis with a median value of T-RFs per month close to the one obtained per lake for the 49 lakes (*i.e.* 37 T-RFs).

There was no significant heterogeneity in the BCS along the horizontal transect over the 2 years of survey

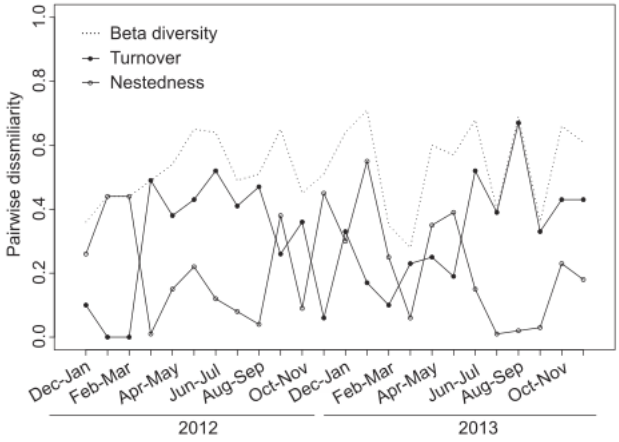

Figure 4. Temporal variations of the bacterial community beta diversity, turnover and nestedness determined by comparing pairwise successive sampling campaign (n = 24).

(STI2, $R^2 = 0.01$, $F = 0.296$, $P = 0.998$). This result suggested that there was no obvious impact of the storm sewer effluent on the BCS at the surface water, as found in the analysis of the 49 lakes. Similarly, no vertical difference in the BCS was observed at the central point (STI2, $R^2 = 0.01$, $F = 0.742$, $P = 0.713$), even when the lake was stratified (8 out of the 25 campaigns, data not shown). Since no spatial variability was encountered, we then focused on the pooled samples (M) to analyse the temporal variation in the BCS.

Redundancy analysis of T-RFLP profiles revealed ($R^2_{adj} = 14\%$) that the temporal variations in BCS were significantly explained by water temperature ($P = 0.001$). Moreover, quite important dissimilarities in BCS were observed from month to month, with values ranging from 23 to 77% (Fig. 4). No clear temporal pattern in the beta diversity fluctuations could be reported (Fig. 4). However, when decomposing the beta diversity into its turnover and nestedness components, we observed distinct temporal variations (Fig. 4) strongly linked to variations in water temperature (linear models: $R^2 = 0.46$, $F = 20.220$, $P < 0.001$ for turnover and $R^2 = 0.38$, $F = 5.825$, $P = 0.004$ for nestedness). Indeed, a positive relationship between the temperature and the turnover was observed (Supplementary Fig. S1a), indicating that T-RF replacement was higher in summer. In contrast, the negative relationship between temperature and nestedness (Supplementary Fig. S1b) suggested that the composition of T-RFs was strongly influenced by richness variations in winter.

At the local scale, the neutral community model significantly explained the relationship between the T-RFs' occurrence frequency and their relative abundance variations in 22 out of the 25 campaigns (Supplementary Fig. S2). Over the 22 campaigns, the neutral model explained on average 65% of the total variance with values ranging from 33 to 91%. Contrary to the temporal BCS variations, neither the total variance explained by the neutral model ($r_{Pearson} = -0.34$, $P = 0.116$), nor the immigration probability

appeared to be related to the water temperature ($r_{Pearson} = -0.25$, $P = 0.267$; Supplementary Fig. S2). Regarding the temporal dynamics of the T-RFs' abundance on the average sample M, the neutral time-series model proposed by Ofiţeru *et al.* (2010) could explain the relative abundance variations of only one T-RF ($R^2_{adj} = 76\%$, $P < 0.001$). The lack of relevance of this model to our data may come from the presence of numerous zeros in the dataset corresponding to the T-RFs that were not detected during campaigns. Indeed, the single T-RF that fitted the neutral model was the only one detected several times in the 2 years. Besides, when performing the neutral time-series model on the average relative abundance of the five Créteil Lake points (S1, C1, C2, C3 and R1), four out of the nine T-RFs detected over all the 25 campaigns had a relative abundance variability significantly explained by the neutral model with an average R^2_{adj} of 34% (data not shown).

Discussion

The main goal of this study was to investigate not only the relative influence of deterministic processes (*i.e.* environmental conditions and spatial factors) but also the influence of neutral processes on the bacterioplankton metacommunity within a large set of shallow lakes. One of the original aspects of this study lies in the 2-year monthly monitoring of the Créteil Lake combined with the three repeated summer samplings of 49 contrasted lakes to take into account inter-annual variability. Indeed, metacommunity surveys in continental aquatic systems integrating more than a single sampling campaign are rare (*e.g.*, van der Gucht *et al.* 2007; Logares *et al.* 2013). Our results suggested that stochastic processes were relevant at both regional and local scales to explain the bacterioplankton metacommunity assembly.

Fingerprinting methodology considerations

Factors and processes influencing BCS were explored using a fingerprinting method that may create a bias by underestimating the contribution of rare members from the sampled bacterial community (Pedrós-Alió 2006). Thus, if non-detected taxa are generalists, spatial and temporal BCS variation would be buffered. Inversely, differences would be enhanced if non-detected taxa happened to be specialists of local environmental conditions (Székely and Langenheder 2014). Although high-throughput sequencing yields a greater coverage of the bacterial diversity and may avoid this potential bias, comparable patterns in terms of diversity and structure were obtained on the same samples with fingerprinting methods including the TRFLP (*e.g.*, Castro-Carrera *et al.* 2014; Elsayed *et al.* 2014; van Dorst *et al.* 2014; Thomson *et al.* 2015) whatever the primer set used. Thus it seems reasonable to validate our conclusions regarding the relative impact on BCS of environmental conditions, spatial factors and stochastic processes.

Minor influence of deterministic processes on the BCS at the regional scale

Our results indicate that environmental conditions and spatial factors play a minor role in shaping the BCS at the regional scale, as suggested by the low amount of total variance explained by these two factors. Several reasons could lead to this result. First, according to Lindström and Langenheder (2012), the large proportion of the BCS variance that remains unexplained (90%) could be explained by the absence of

relevant variables taken into account in the statistical models. For example, van der Gucht *et al.* (2007) showed in the de Maten reserve, a well studied system of connected ponds, that 89% of the total BCS variance was explained by environmental and spatial factors. Local factors used in the de Maten analysis included classical physicochemical parameters (*e.g.*, conductivity, pH), but also biological variables (*i.e.* phytoplankton and zooplankton biomass) that could directly structure the BCS owing to top-down control. Second, since all lakes are different regarding the local parameters, it is possible that overall the statistical analysis does not detect any tendency. Third, the small influence of deterministic factors on the bacterial community assembly could also originate from the dominance of generalist taxa in these ecosystems. However, the relatively high beta diversity tends to invalidate this hypothesis. Finally, Gilbert and Bennett (2010) demonstrated that variance-partitioning analysis tends to underestimate the relative importance of the deterministic factors.

Although only a small amount of total variance was explained by these factors, several variables were identified as having a significant effect on the BCS. Among the environmental factors, the trophic status was the only variable systematically considered significant over the 3 years of sampling. This parameter has already been characterized in the literature to shape the BCS (Lindström 2000; Yannarell *et al.* 2003; Yannarell and Triplett 2004). Among the spatial factors, a relative similar fraction of total BCS variations was weakly explained by the potential ability of bacteria to disperse (characterized by the type of linkage to the hydrographical network) and the spatial structure of the bacterial community at the regional scale (characterized by the PCNMs variables) (Supplementary Table S2). Furthermore, in this analysis, no significant direct influence of storm sewer discharges into the lakes was identified. Similarly, no influence of the dominant watershed land use (*i.e.* natural, agricultural, open or dense urban area) was detected, though previous literature suggested an impact of land use on BCS in rock biofilms or on bacterial population (Scopel, Harris and McLellan 2006; McLellan *et al.* 2007; Lear and Lewis 2009).

Bacterial assembly shape by neutral processes at the regional scale

The neutral community model was very powerful in explaining the T-RF proportion observed in our panel of contrasted lakes (mean R^2 = 76%). This result suggested that the bacterial community was shaped by a stochastic balance between loss and gain of taxa (Hubbell 2001; Sloan *et al.* 2006). Although numerous microcosm or environmental experiments found evidence that bacterial assembly was jointly shaped by local factors and neutral processes (Ayarza and Erijman 2011; Langenheder and Székely 2011; Lee *et al.* 2013; Pholchan *et al.* 2013), our results are in an agreement with the study performed by Drakare and Liess (2010) on 13 Swedish lakes. Indeed, these authors also reported the predominance of neutral processes compared with local factors. Our results could be explained by the strong adaptability and plasticity of the bacterial compartment to environmental perturbations and gradients, reducing the apparent sensitivity to local factors (Östman *et al.* 2010). Although our data well fitted the neutral community model, it is difficult to infer the absence of the influence of deterministic factors on BCS (Chisholm and Pacala 2010), especially as the fraction of variance unexplained by the VPA could not be attributed to stochastic processes (Vellend *et al.* 2014). Moreover, as discussed in Anderson *et al.* (2011), the neutral and the deterministic mechanisms might coincidentally mirror the same patterns. Therefore, isolating the importance of both components separately could be perilous without experimental approaches.

Regarding the immigration rate, large inter-annual discrepancies were observed with an increase from 2011 to 2013. According to the neutral theory (Hubbell 2001), this is consistent with the decrease in the beta diversity over the 3 years of sampling. The inter-annual variation of the immigration rate could also result from meteorological conditions that may have led to higher bacterial dispersion in 2013 (*e.g*., strong wind). However, these variations could also result from a difference in the metacommunity size and not from the immigration rate.

Furthermore in this study, we didn't identify any other important mechanism apart from the neutral processes that could influence the assembly of communities. Indeed, we were unable to clearly identify a mass effect in lakes linked to rivers as observed by Lindström *et al.* (2006) and Nelson, Sadro and Melack (2009). According to the literature, this absence of effect could be due to the trophic status of the 49 lakes (mainly ranging from eutrophic to hypereutrophic levels). Indeed, a mass effect appears to be more effective in oligotrophic lakes (van der Gucht *et al.* 2007), whereas more productive lakes seem to be dominated by stochastic processes compared with deterministic processes (Chase 2010).

This survey showed the importance of the neutral dynamics on the bacterioplankton metacommunity assembly at the regional scale. However, this sole approach did not allow identification of spatial and temporal mechanisms that govern the BCS at the scale of a single ecosystem though these processes could have a direct influence on the dynamics at the regional scale. These issues were explored at the local scale within the Créteil Lake.

Influence of deterministic and neutral processes at the local scale

In Créteil Lake, we expected a spatial structure to exist as from the storm sewer outlet to the lake outlet as it may represent an important source of exogenous taxa (Sercu *et al.* 2009). In addition, lakes that have a long water retention time (>180 days for Créteil Lake) are characterized by significant dissimilarities in BCS between the lake water column and the inlet (Lindström *et al.* 2006; Nelson, Sadro and Melack 2009). However, our 2-year monthly survey did not reveal any impact of the storm sewer effluent on the dominant bacterial taxa in the surface waters, even after strong rainfall events. The lack of intra-lake horizontal heterogeneity may originate from the moderate size of Créteil Lake and thus the small distance between the three sampling points. Indeed, important in-lake BCS spatial variation has generally been described in larger lakes (*e.g.*, Yannarell and Triplett 2004; de Wever *et al.* 2005; Jones *et al.* 2012) contrary to smaller lakes (Jones *et al.* 2012). In Créteil Lake, this result only applies to three surface locations and cannot be generalized to the whole lake or other compartments such as the sediments. In addition, the T-RFLP technique only assesses dominant taxa (Bent, Pierson and Forney 2007) and we cannot exclude an impact of storm sewer outlet on less abundant taxa such as pathogens or fecal indicators. Regarding the vertical profile, no BCS stratification was observed at the central station, even when the water column was stratified for several days (from 3 to 30 days of stratification). These two results showed a relative spatial homogeneity of the dominant bacterioplankton community on the different sampled points that could result from neutral processes. Overall, the neutral community model explained a large proportion of the bacterial assemblage variance per campaign (R^2_{mean} = 65%). However, no distinct seasonal pattern of neutral model parameters (R^2 and $N_T m$) was observed over the 2 years of survey, suggesting that the influence of stochastic processes was rather constant through the year.

Furthermore, the BCS within Créteil Lake displayed a marked temporal variation all along the sampling campaigns, essentially driven by the water temperature. Such importance of this deterministic parameter has already been reported in the literature in a wide range of aquatic habitats (*e.g.*, Muylaert *et al.* 2002; Stepanauskas *et al.* 2003; Jardillier *et al.* 2004). Moreover, this temporal analysis revealed that the taxa replacement was the main process shaping BCS in summer while random variation in T-RF richness occurred in winter. This pattern could result from the consequence of a trade-off between competitive abilities and resistance to predators over low and high productivity levels (Leibold 1999; Horner-Devine *et al.* 2003) or simply be due to purely random input of bacterial taxa, for example owing to the sediment resuspension or the sewer outlet input during rainy periods after a long period of dry weather. In addition to the importance of the water temperature on the BCS, some T-RFs displayed neutral time-series dynamics that explained up to 80% of their relative abundance variability. However, with our results, it is difficult to generalize the importance over time of the neutral processes on BCS since a large fraction of the T-RFs were not observed to be shaped by stochastic processes, indicating that bacterial groups were not all governed by the same processes (*e.g.*, Barberán and Casamayor 2010; Székely and Langenheder 2014), probably due to specific ecological traits (Philippot *et al.* 2010).

This study showed that the bacterial assembly in the 49 lakes located in the Paris area appeared to be strongly governed by stochastic processes. Coupled to this approach, the finer monitoring performed on Créteil Lake revealed that neutral dynamics spatially structured the bacterial community. BCS was temporally governed by a deterministic factor, *i.e.* the water temperature, and also by stochastic processes for some T-RFs.

Further studies on artificial shallow lakes, performed on a balanced group in terms of trophic status and type of linkage to the hydrographical network, should be carried out in order to develop a more robust overview of the mechanisms shaping the bacterial community.

Acknowledgments

The authors thank Pierre Legendre for his helpful advices in the statistical analysis, Sophie Penisson for her support in decrypting the Sloan's neutral model equation and the two anonymous reviewers whose comments improved the manuscript. We warmly thank Ross Holland from Aqua Cytometry (UK) for his precious contribution in the bacterial count. We are grateful to the stakeholders and owners for their authorization to collect samples in each lake. We also thank the leisure base of Créteil Lake for lending us their boats over the 2 years of sampling. Finally we warmly thank all the colleagues who helped in sampling and quantifying the ecological parameters in the 49 lakes and Créteil Lake during these 3 years.

Funding

This work was supported by the French National Research Agency through the PULSE (Peri-Urban Lakes, Society and Environment) research project (ANR-10-CEPL-010).

Conflict of interest. None declared

References

Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Freestone AL, Sanders NJ, Cornell H V, Comita LS, Davies KF, Harrison SP, Kraft NJB, Stegen JC and Swenson NG (2011) Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. Ecol Lett 14:19–28.

Ayarza JM and Erijman L (2011) Balance of neutral and deterministic components in the dynamics of activated sludge floc assembly. Microb Ecol 61:486–495.

Azeria ET, Ibarzabal J, Boucher J and Hérbert C (2011) Towards a better understanding of beta diversity: deconstructing composition patterns of saproxylic beetles breeding in recently burnt boreal forests. Res Biodivers Model Appl 75–94.

Barberán A and Casamayor EO (2010) Global phylogenetic community structure and β- diversity patterns in surface bacterioplankton metacommunities. Aquat Microb Ecol 59:1–10.

Baselga A (2010) Partitioning the turnover and nestedness components of beta diversity. Glob Ecol Biogrography 19:134–143.

Baselga A, Orme D, Villeger S, de Bortoli J and Leprieur F (2013) Package "betapart". Partitioning beta diversity into turnover and nestedness components, version 1.3.

Beisner BE, Peres-Neto PR, Lindström ES, Barnett A and Longhi ML (2006) The role of environmental and spatial processes in structuring lake communities from bacteria to fish. Ecology 87:2985–2991.

Bent SJ, Pierson JD and Forney LJ (2007) Measuring species richness based on microbial community fingerprints: the emperor has no clothes. Appl Environ Microbiol 73:2399–2401.

Borcard D, Gillet F and Legendre P (2011) Numerical ecology with R. Springer Science and Business Media.

Borcard D and Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol Modell 153:51–68.

Borcard D, Legendre P and Drapeau P (1992) Partialling out the spatial component of ecological variation. Ecology 73:1045–1055.

Castro-Carrera T, Toral PG, McEwan NR, Hervas G, Abecia L, Pinloche E, Girdwood SE and Belenguer A (2014) Rumen bacterial community evaluated by 454 pyrosequencing and terminal restriction fragment length polymorphism analyses in dairy sheep fed marine algae. J Dairy Sci 97:1661–1669.

Catherine A, Mouillot D, Escoffier N, Bernard C and Troussellier M (2010) Cost effective prediction of the eutrophication status of lakes and reservoirs. Freshw Biol 55:2425– 2435.

Catherine A, Troussellier M and Bernard C (2008) Design and application of a stratified sampling strategy to study the regional distribution of cyanobacteria (Ile-de-France, France). Water Res 42:4989–5001.

Chase JM (2010) Stochastic community assembly causes higher biodiversity in more productive environments. Science 328:1388–1391.

Chisholm RA and Pacala SW (2010) Niche and neutral models predict asymptotically equivalent species abundance distributions in high-diversity ecological communities. Proc Natl Acad Sci USA 107:15821–

15825.

Culman SW, Bukowski R, Gauch HG, Cadillo-Quiroz H and Buckley DH (2009) T-REX: software for the processing and analysis of T-RFLP data. BMC Bioinformatics 10:171.

de Wever A, Muylaert K, van der Gucht K, Pirlot S, Cocquyt C, Descy JP, Plisnier PD and Vyverman W (2005) Bacterial community composition in Lake Tanganyika: vertical and horizontal heterogeneity. Appl Environ Microbiol 71:5029–5037.

Drakare S and 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.

Elsayed OF, Maillard E, Vuilleumier S and Imfeld G (2014) Bacterial communities in batch and continuous-flow wetlands treating the herbicide S-metolachlor. Sci Total Environ 499, 327–335.

Gilbert B and Bennett JR (2010) Partitioning variation in ecological communities: do the numbers add up? J Appl Ecol 47:1071–1082.

Horner-Devine MC, Leibold MA, Smith VH and Bohannan BJM (2003) Bacterial diversity patterns along a gradient of primary productivity. Ecol Lett 6:613–622.

Hubbell SP (2001) The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton.

INSEE (Institut national de la statistique et des études économiques) (2012) Occupation physique du territoire en 2012.

INSEE (Institut national de la statistique et des études économiques) (2013) La France en Bref - France in figures.

Jardillier L, Basset M, Domaizon I, Belan A, Amblard C, Richardot M and Debroas D (2004) Bottom-up and top-down control of bacterial community composition in the euphotic zone of a reservoir. Aquat Microb Ecol 35:259–273.

Jones SE, Cadkin TA, Newton RJ and McMahon KD (2012) Spatial and temporal scales of aquatic bacterial beta diversity. Front Microbiol 318:1–10.

Jones SE, Newton RJ and McMahon KD (2009) Evidence for structuring of bacterial community composition by organic carbon source in temperate lakes. Appl Environ Microbiol 11:2463–2472.

Kent AD, Yannarell AC, Rusak JA, Triplett EW and McMahon KD (2007) Synchrony in aquatic microbial community dynamics. ISME J 1:38–47.

Langenheder S and Ragnarsson H (2007) The role of environmental and spatial factors for the composition of aquatic bacterial communities. Ecology 88:2154–2161.

Langenheder S and Székely AJ (2011) Species sorting and neutral processes are both important during the initial assembly of bacterial communities. ISME J 5:1086–1094.

Lear G and Lewis GD (2009) Impact of catchment land use on bacterial communities within stream biofilms. Ecol Indic 9:848–855.

Lee JE, Buckley HL, Etienne RS and Lear G (2013) Both species and neutral processes drive assembly of

bacterial communities in aquatic microcosms. FEMS Microbiol Ecol 86:288–302.

Legendre P, Borcard D and de Cáceres M (2012) Package 'STI'. Space-time ANOVA models without replications, version 1.0.2.

Legendre P, De Cáceres M and Borcard D (2010) Community surveys through space and time: testing the space–time interaction in the absence of replication. Ecology 91:262–272.

Legendre P and Gauthier O (2014) Statistical methods for temporal and space−time analysis of community composition data. Proc R Soc B 281:20132728.

Leibold MA (1999) Biodiversity and nutrient enrichment in pond plankton communities. Evol Ecol Res 1:73–95.

Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, Holt RD, Shurin JB, Law R, Tilman DG, Loreau M and Gonzalez A (2004) The metacommunity concept: a framework for multi-scale community ecology. Ecol Lett 7:601–613.

Lindström ES (2000) Bacterioplankton community composition in five lakes differing in trophic status and humic content. Microb Ecol 40:104–113.

Lindström ES, Forslund M, Algesten G and Bergstöm AK (2006) External control of bacterial community structure in lakes. Limnol Oceanogr 51:339–342.

Lindström ES, Kamst-van Agterveld MP and 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.

Lindström ES and Langenheder S (2012) Local and regional factors influencing bacterial community assembly. Environ Microbiol Rep 4:1–9.

Logares R, Lindström ES, Langenheder S, Logue JB, Paterson H, Laybourn-Parry J, Rengefors K, Tranvik L and Bertilsson S (2013) Biogeography of bacterial communities exposed to progressive long-term environmental change. ISME J 7:937– 948.

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

Marie D, Partensky F, Jacquet D and Vaulot D (1997) Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. Appl Environ Microbiol 63:186–193.

Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Ovreås L, Reysenbach AL, Smith VH and Staley JT (2006) Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol 4:102–112.

McLellan SL, Hollis EJ, Depas MM, van Dyke M, Harris J and Scopel CO (2007) Distribution and fate of Escherichia coli in lake Michigan following contamination with urban stormwater and combined sewer overflows. J Great Lakes Res 33:566–580.

Muylaert K, van der Gucht K, Vloemans N, de Meester L, Gillis M and Vyverman W (2002) Relationship

between bacterial community composition and bottom-up versus top- down variables in four eutrophic shallow lakes. Appl Environ Microbiol 68:4740– 4750.

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

OECD (Organization for Economic Cooperation and Development) (1982) Eutrophication of waters: monitoring, assessment and control, OECD, Paris.

Ofiţeru ID, Lunn M, Curtis TP, Wells GF, Criddle CS, Francis CA and Sloan WT (2010) Combined niche and neutral effects in a microbial wastewater treatment community. Proc Natl Acad Sci USA 107:15345–15350.

Oksanen J, Blanchet G, Kindt R, Legendre P, Minchin P, O'Hara R, Simpson G, Solymos P, Stevens M and Wagner H (2015) Package "vegan". Community ecology package, version, 2(9).

Osborn AM, Moore ER and Timmis KN (2000) An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. Environ Microbiol 2:39–50.

Östman Ö, Drakare S, Kritzberg ES, Langenheder S, Logue JB and Lindström ES (2010) Regional invariance among microbial communities. Ecol Lett 13:118–127.

Pedrós-Alió C (2006) Marine microbial diversity: can it be determined? Trends Microbiol 14:257–263.

Peres-Neto PR, Legendre P, Dray S and Borcard D (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. Ecology 87:2614–2625.

Pholchan MK, Baptista JDC, Davenport RJ, Sloan WT and Curtis TP (2013) Microbial community assembly, theory and rare functions. Front Microbiol 4:1–9.

Philippot L, Andersson SGE, Battin TJ, Prosser JI, Schimel JP, Whitman WB and Hallin S (2010) The ecological coherence of high bacterial taxonomic ranks. Nat Rev Microbiol 8:523–529.

R Development Core Team (2014) R: a language and environment for statistical computing. Vienna, Autria: R Foundation for Statistical Computing.

Rogora M, Minella M, Orrù A and Tartari GA (2006) A comparison between high- temperature catalytic oxidation and persulphate oxidation for the determination of total nitrogen in freshwater. Int J Environ Anal Chem 86:1065–1078.

Scheffer M (2004) Ecology of shallow lakes. Berlin: Springer.

Scopel CO, Harris J and McLellan SL (2006) Influence of nearshore water dynamics and pollution sources on beach monitoring outcomes at two adjacent Lake Michigan beaches. J Great Lakes Res 32:543–552.

Sercu B, van der Werfhorst LC, Murray J and Holden PA (2009) Storm drains are sources of human fecal pollution during dry weather in three urban southern california watersheds. Environ Sci Technol 43:293–298.

Shade A, Kent AD, Jones SE, Newton RJ, Triplett EW and McMahon KD (2007) Interannual dynamics

and phenology of bacterial communities in a eutrophic lake. Limnol Oceanogr 52:487–494.

Sloan WT, Lunn M, Woodcock S, Head IM, Nee S and Curtis TP (2006) Quantifying the roles of immigration and chance in shaping prokaryote community structure. Environ Microbiol 8:732–740.

Soininen J, Korhonen JJ, Karhu J and Vetterli A (2011) Disentangling the spatial patterns in community composition of prokaryotic and eukaryotic lake plankton. Limnol Oceanogr 56:508–520.

Stepanauskas R, Moran MA, Bergamaschi BA and Hollibaugh JT (2003) Covariance of bacterioplankton composition and environmental variables in a temperate delta system. Aquat Microb Ecol 31:85–98.

Székely AJ and Langenheder S (2014) The importance of species sorting differs between habitat generalists and specialists in bacterial communities. FEMS Microbiol Ecol 87:102–112.

Thomson BC, Tisserant E, Plassart P, Uroz S, Griffiths RI, Hannula SE, Buée M, Mougel C, Ranjard L, van Veen JA, Martin F, Bailey MJ and Lemanceau P (2015) Soil conditions and land use intensification effects on soil microbial communities across a range of European field sites. Soil Biol Biochem 88:403–413.

van der Gucht K, Cottenie K, Muylaert K, Vloemans N, Cousin S, Declerck S, Jeppesen E, Conde-Porcuna JM, Schwenk K, Zwart G, Degans H, Vyverman W and de Meester L (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.

van Dorst J, Bissett A, Palmer AS, Brown M, Snape I, Stark JS, Raymond B, McKinlay J, Ji M, Winsley T and Ferrari BC (2014) Community fingerprinting in a sequencing world. FEMS Microbiol Ecol 89:316–330.

Vellend M, Srivastava DS, Anderson KM, Brown CD, Jankowski JE, Kleynhans EJ, Kraft NJB, Letaw AD, Macdonald AAM, Maclean JE, Myers-Smith IH, Norris AR and Xue X (2014) Assessing the relative importance of stochasticity in ecological communities. Oikos 123:1420–1430.

Woodcock S, van der Gast CJ, Bell T, Lunn M, Curtis TP, Head IM and Sloan WT (2007) Neutral assembly of bacterial communities. FEMS Microbiol Ecol 62:171–180.

Yannarell AC, Kent AD, Lauster GH, Kratz TK and Triplett EW (2003) Temporal patterns in bacterial communities in three temperate lakes of different trophic status. Microb Ecol 46:391–405.

Yannarell AC and Triplett EW (2004) Within- and between-lake variability in the composition of bacterioplankton communities: investigations using multiple spatial scales. Appl Environ Microbiol 70:214–223.

Yannarell AC and Triplett EW (2005) Geographic and environmental sources of variation in lake bacterial community composition geographic and environmental sources of variation in lake bacterial community composition. Appl Environ Microbiol 71:227– 239.

Zubkov MV and Burkill PH (2006) Syringe pumped high speed flow cytometry of oceanic phytoplankton. Cytometry Part A 69:1010–1019.